Microsoft Research 微软亚洲研究院

Information Visualization: Connecting the Abstract and Visual Worlds

Shixia Liu (刘世霞)

Internet Graphics Group,

Microsoft Research Asia

Outline

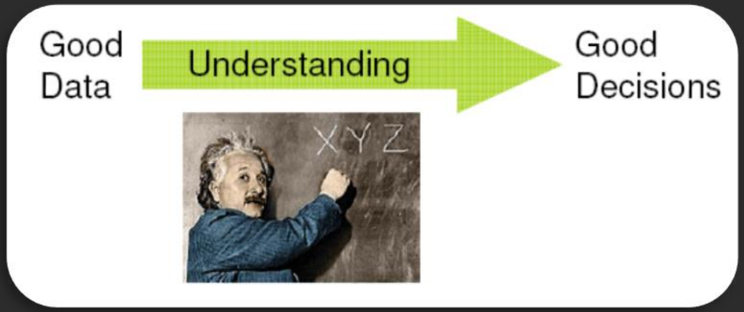
- Definition
- Criteria
- History
- InfoVis 1.0
- InfoVis 2.0
- InfoVis 3.0
- Opportunities and challenges

Outline

- **Definition**
- Criteria
- History
- InfoVis 1.0
- InfoVis 2.0
- InfoVis 3.0
- Opportunities and challenges

What is Information Visualization

- Use graphical representations of abstract information to facilitate
  - Explanation
  - Information understanding
  - Decision making
  - …



- Translation from the abstract language to the visual/intuitive language

Microsoft Research 微软亚洲研究院

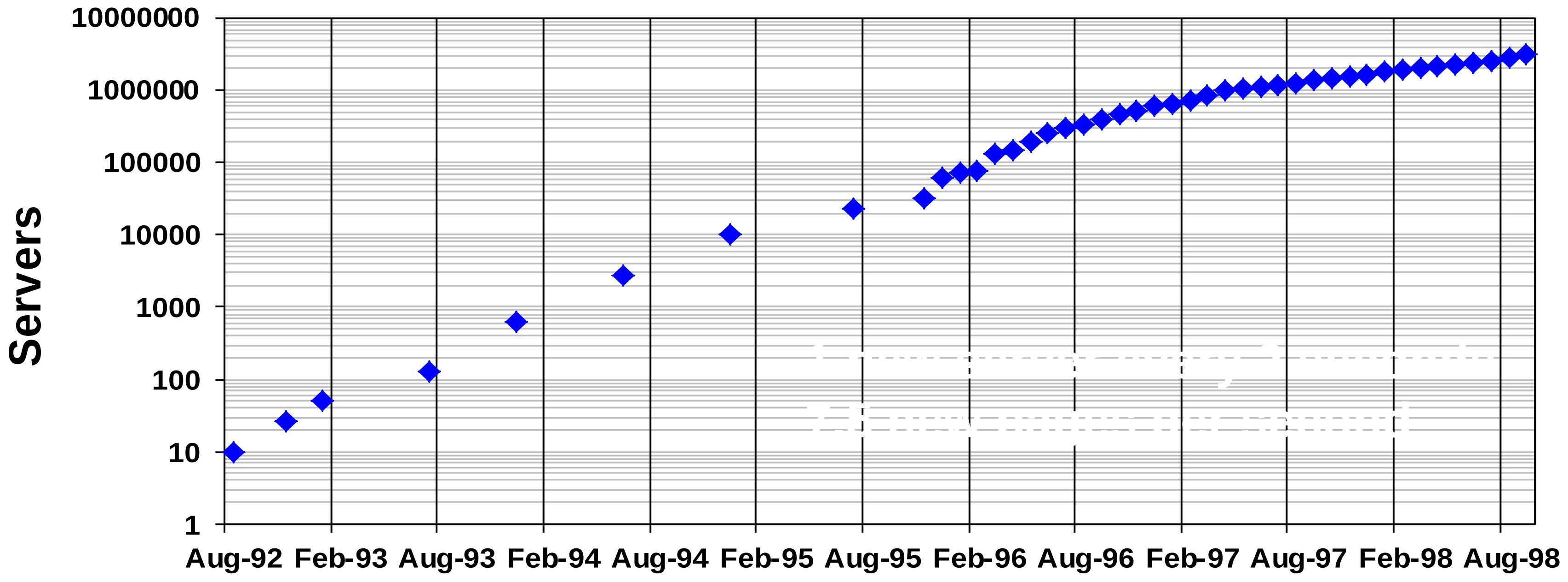
Why InfoVis – Basic Problem

We live in a new ecology: how do you **make sense of it**

People are 5 times faster with the visual aid

(slide "borrowed" from PARC User Interface Research Group)

Why InfoVis - Web Ecologies



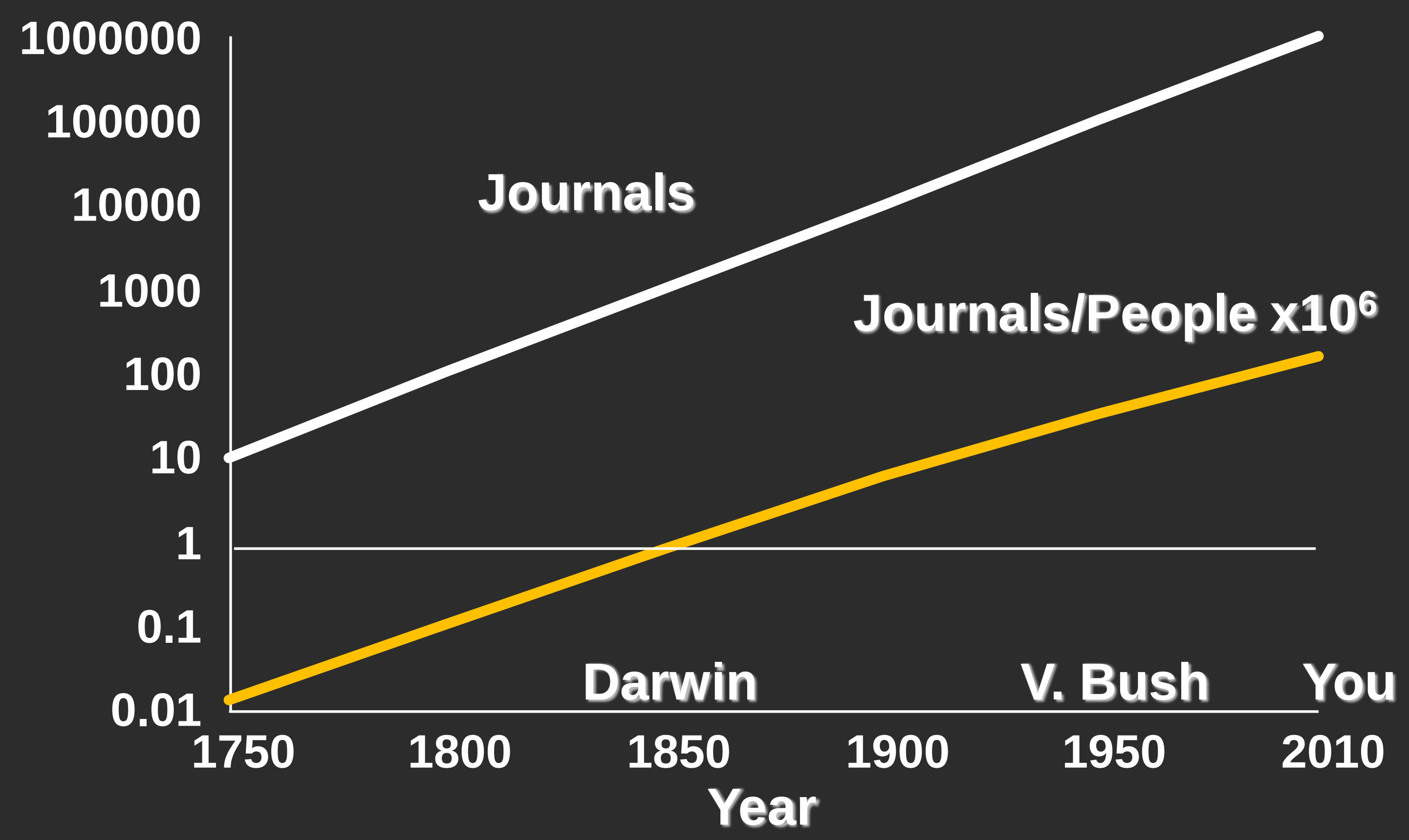
Source: World Wide Web Consortium, Mark Gray, Netcraft Server Survey

(slide “borrowed” from PARC User Interface Research Group)

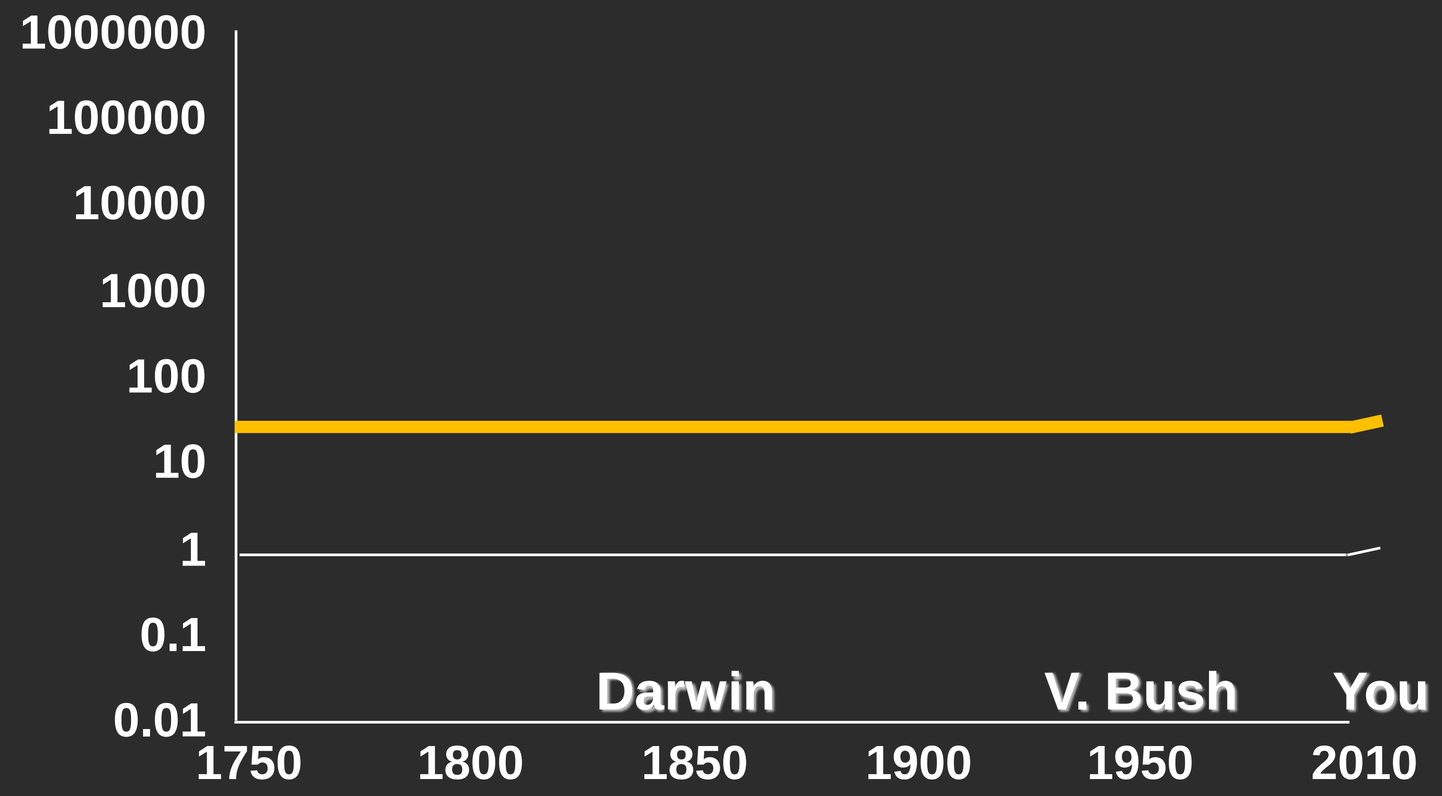
Microsoft Research 微软亚洲研究院

Why InfoVis - Scientific Journals

Journals/person increases 10X every 50 years

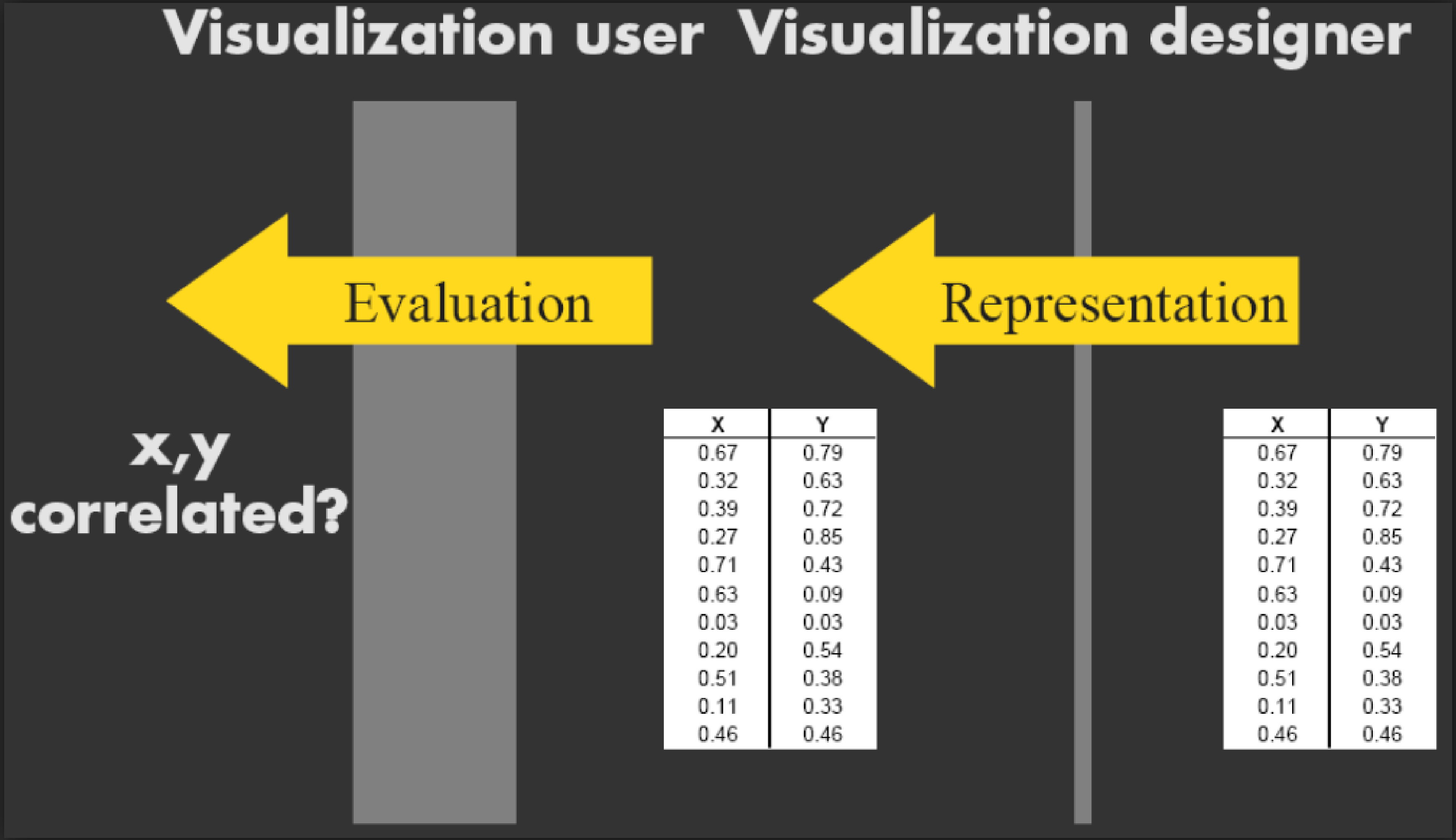


Why InfoVis - Innate Human Capacity



Microsoft Research 微软亚洲研究院

Bad Visualization?

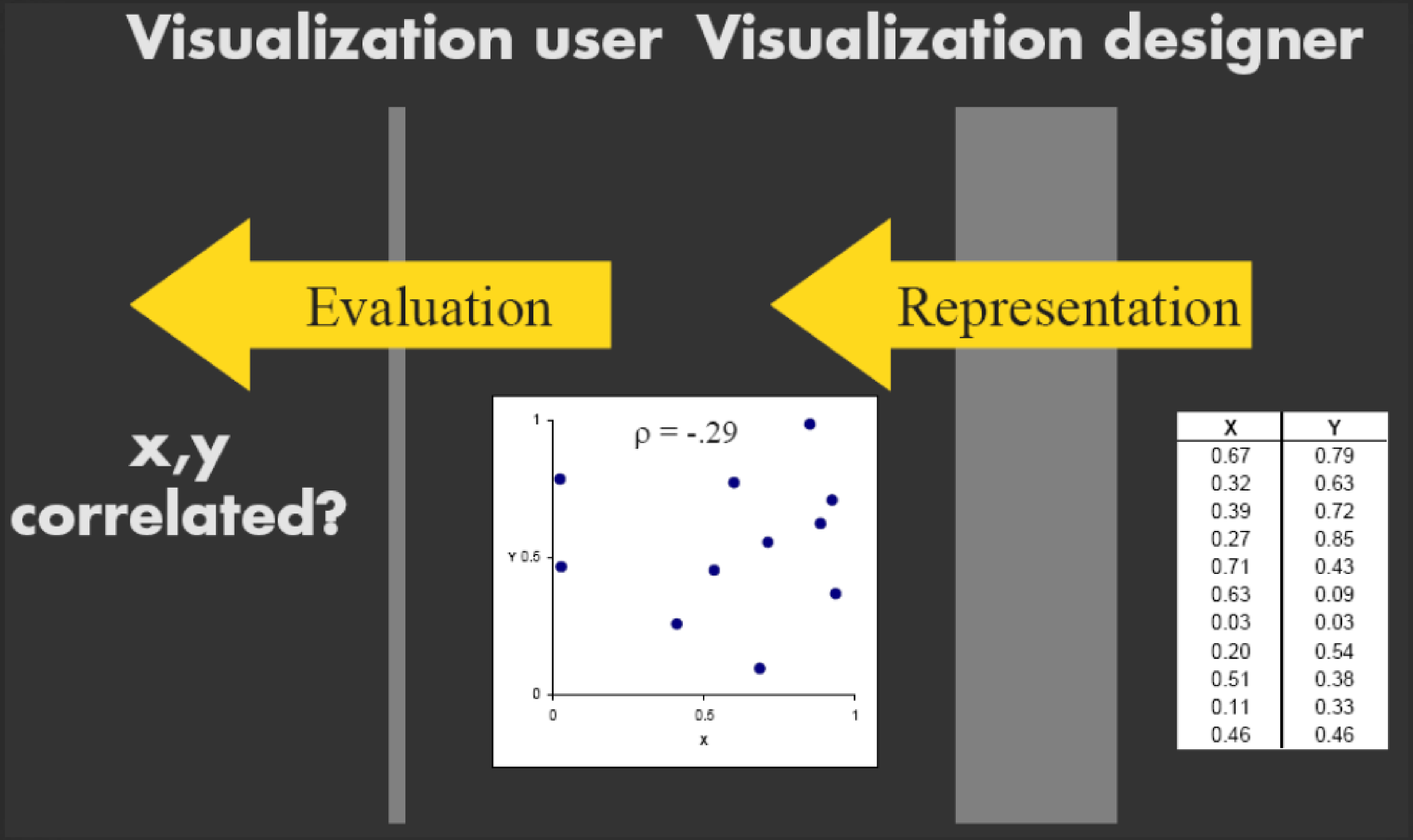


X	Y
0.67	0.79
0.32	0.63
0.39	0.72
0.27	0.85
0.71	0.43
0.63	0.09
0.03	0.03
0.20	0.54
0.51	0.38
0.11	0.33
0.46	0.46

X	Y
0.67	0.79
0.32	0.63
0.39	0.72
0.27	0.85
0.71	0.43
0.63	0.09
0.03	0.03
0.20	0.54
0.51	0.38
0.11	0.33
0.46	0.46

Microsoft Research 微软亚洲研究院

Good Visualization?

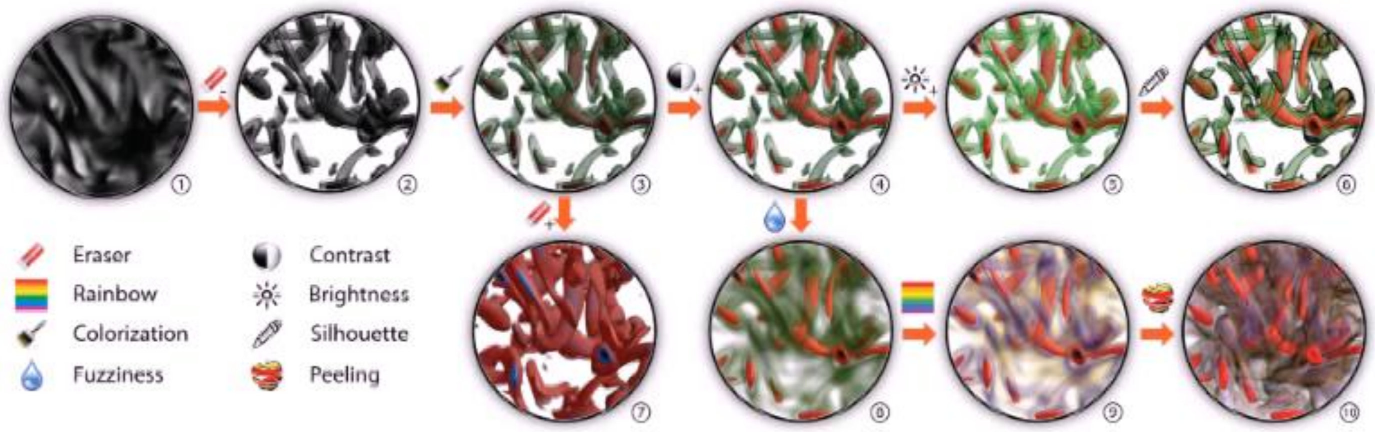


X	Y
0.67	0.79
0.32	0.63
0.39	0.72
0.27	0.85
0.71	0.43
0.63	0.09
0.03	0.03
0.20	0.54
0.51	0.38
0.11	0.33
0.46	0.46

Outline

- Definition
- **Criteria**
- History
- InfoVis 1.0
- InfoVis 2.0
- InfoVis 3.0
- Opportunities and challenges

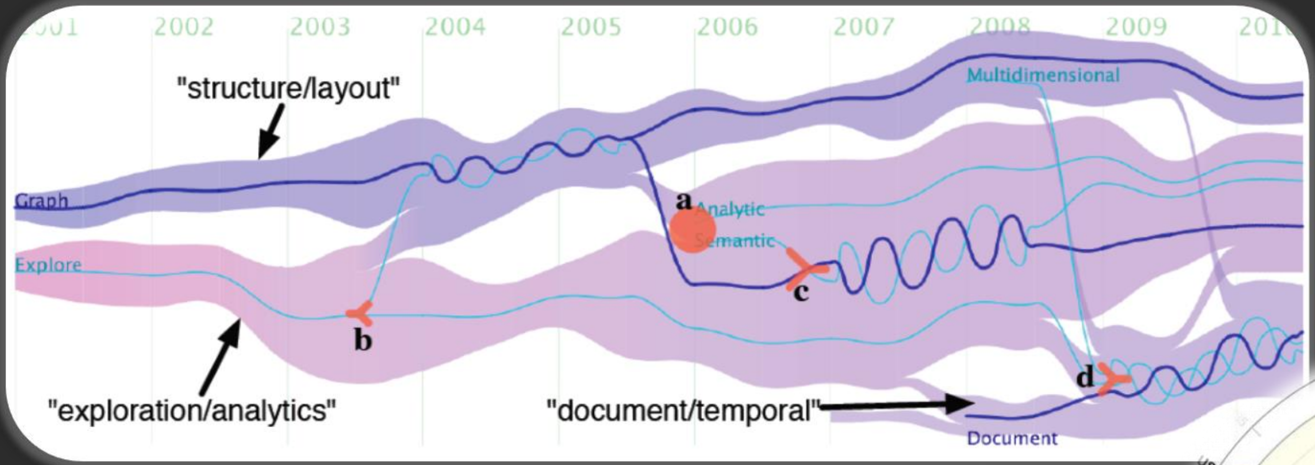
Criteria: Trustfulness (信)



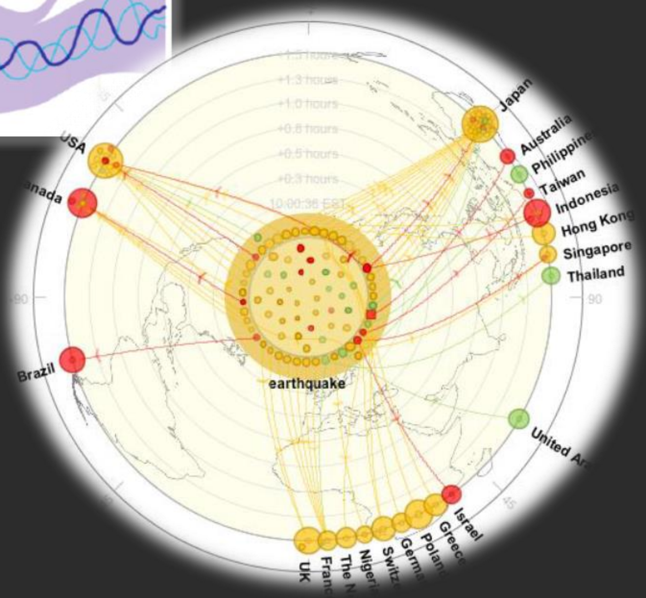
Criteria: Effectiveness (达)

Criteria: Elegance (雅)

TextFlow



Whisper



Outline

- Definition
- Objective
- **History**
- InfoVis 1.0
- InfoVis 2.0
- InfoVis 3.0
- Opportunities and challenges

Microsoft Research 微软亚洲研究院

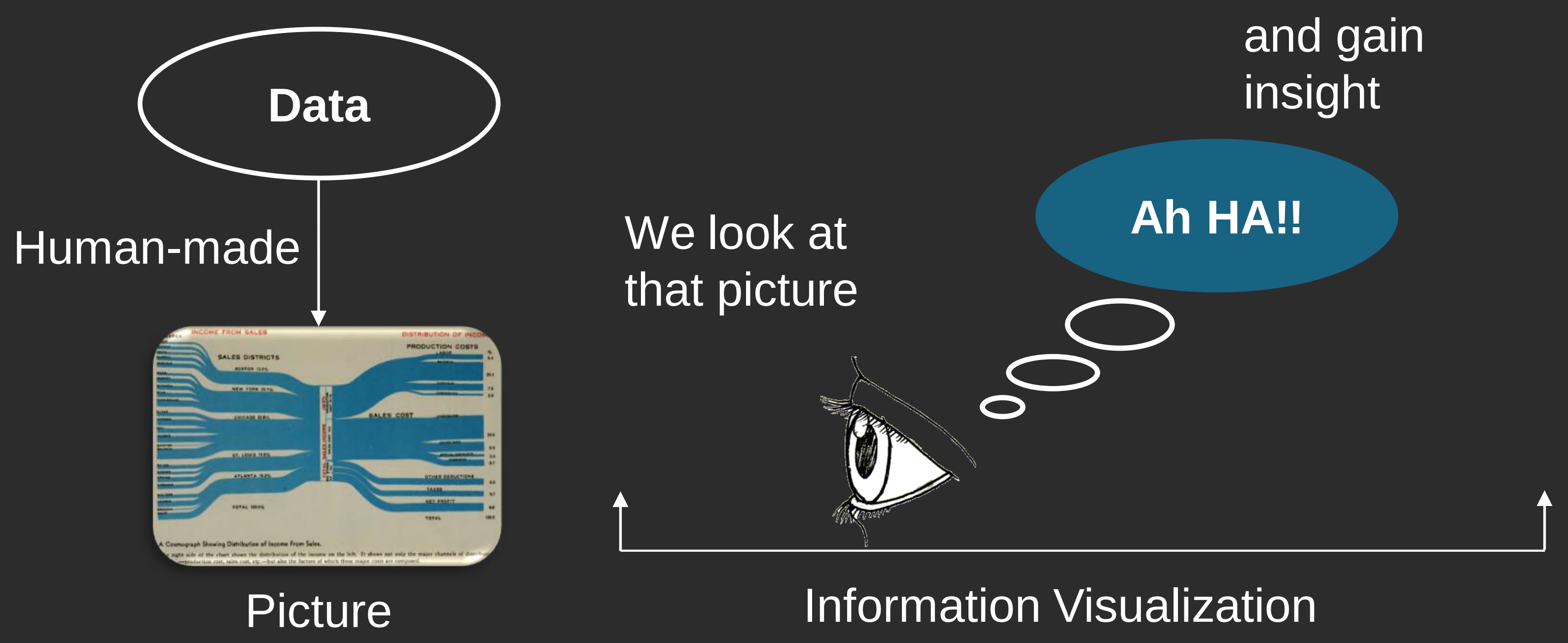
History of Information Visualization

- InfoVis1.0 (Bronze Age： 青铜时代)
  – Hand-made drawing (Info Graphics)
- InfoVis 2.0 (Silver Age： 白银时代)
  – Information display
- InfoVis 3.0 (Golden Age： 黄金时代)
  – Visual analytics

Outline

- Definition
- Objective
- History
- **InfoVis 1.0**
- InfoVis 2.0
- InfoVis 3.0
- Opportunities and challenges

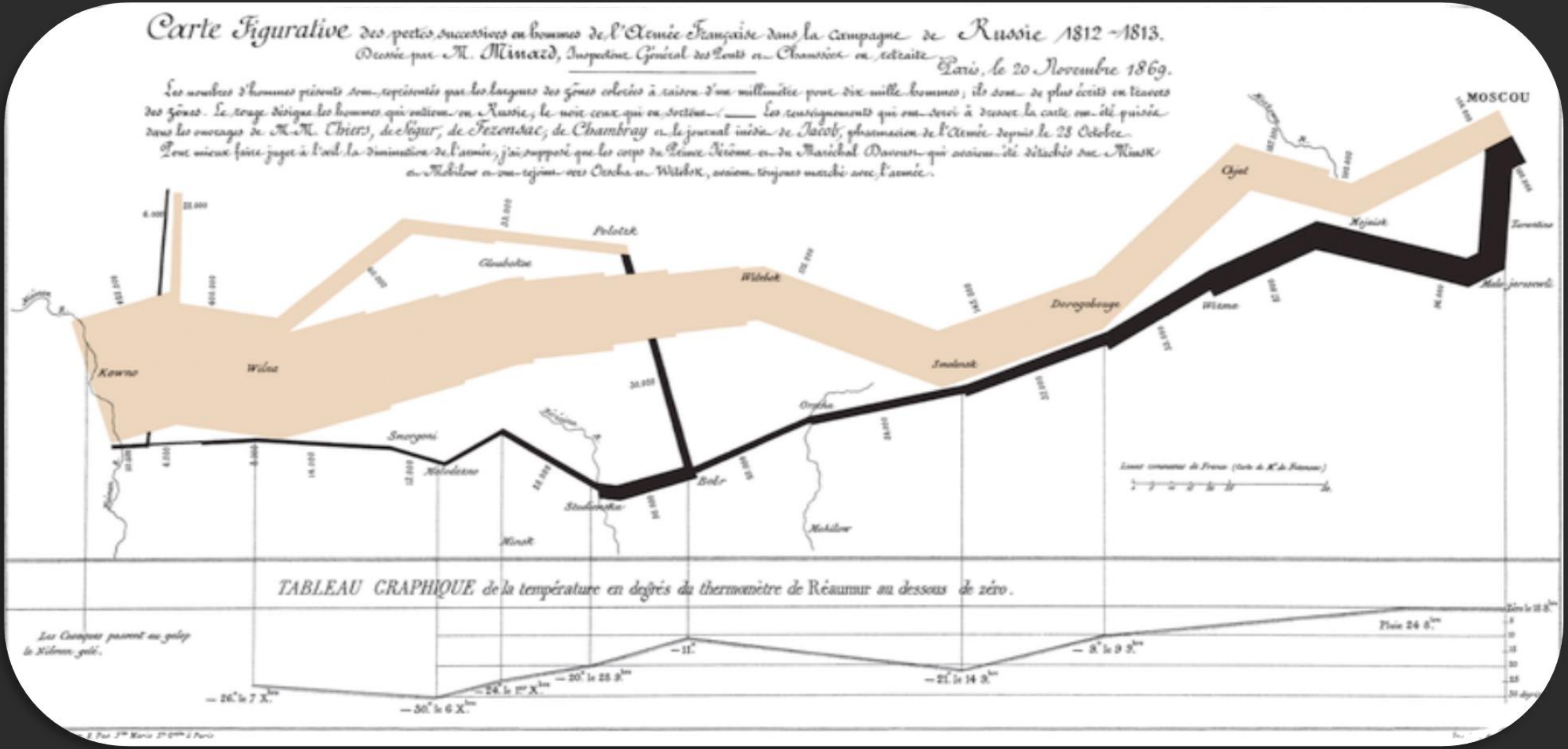
InfoVis 1.0: Overview



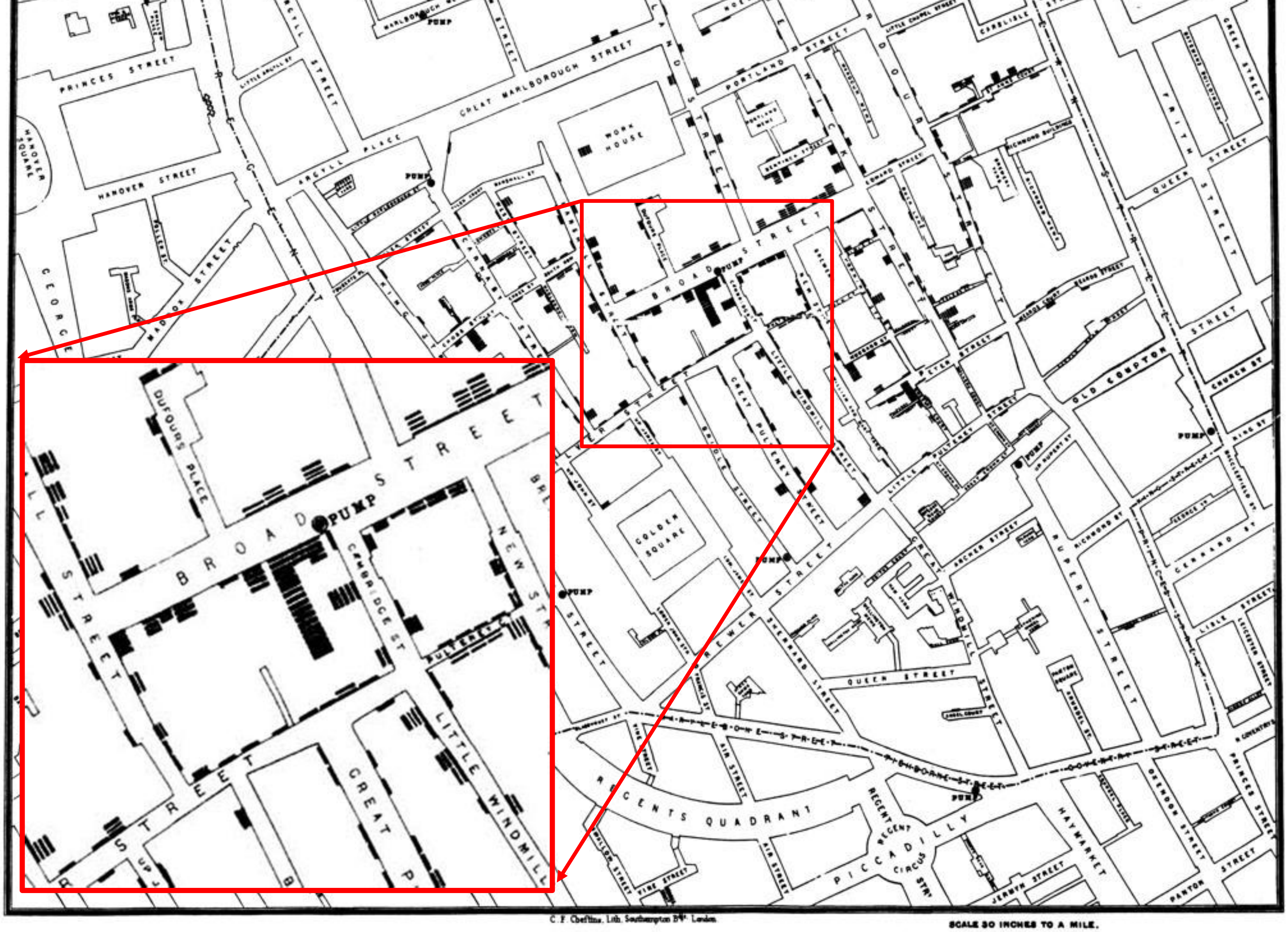
Goals

- Making invisible visible
- Communication
- Understanding
- Gaining insight

Example



Carte Figurative des pertes successives en hommes de l'Armée Française dans la Campagne de Russie 1812-1813.
Paris, le 20 Novembre 1869.
MOSCOU
TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.



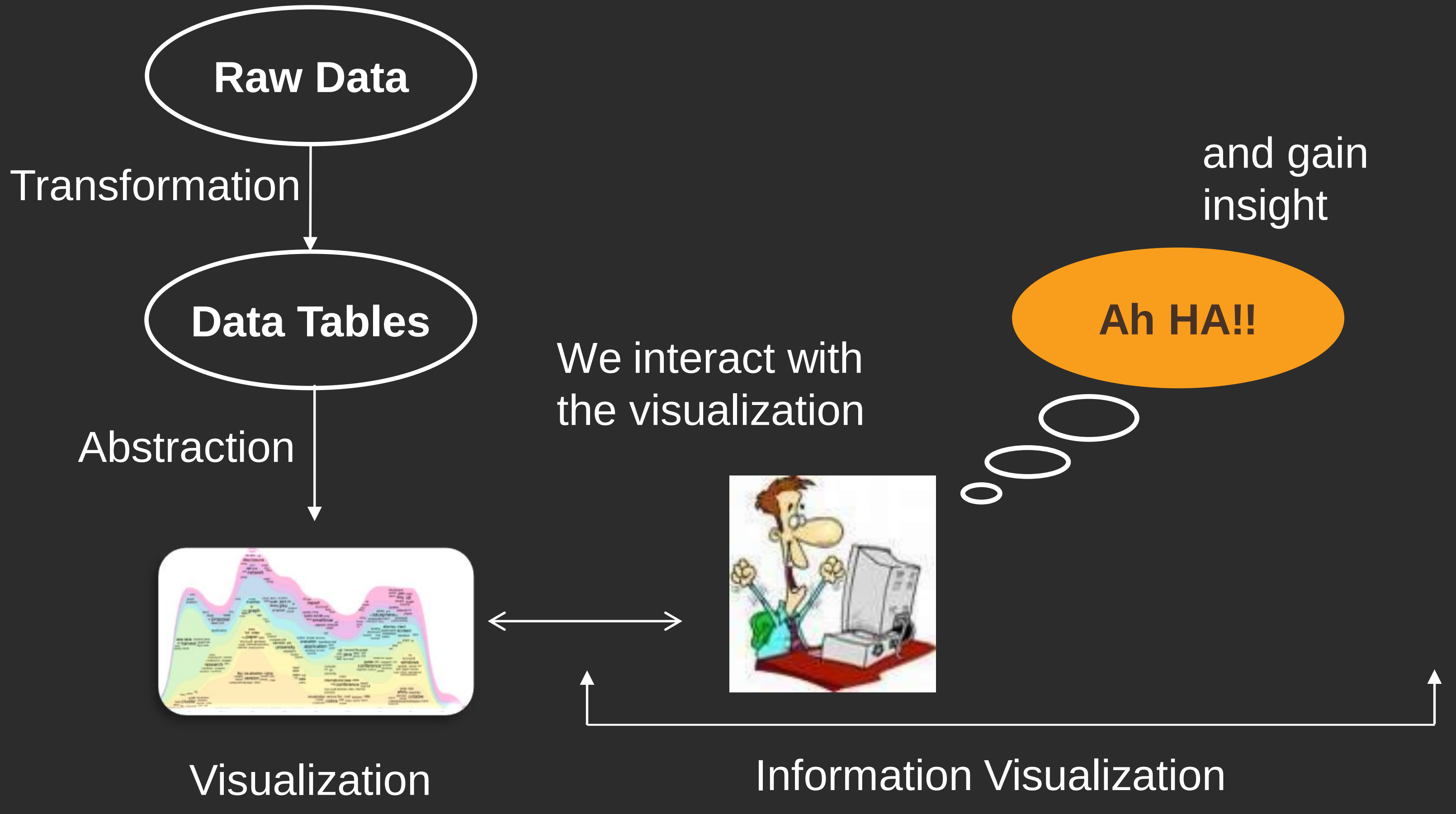
PRINCES STREET
GREAT MARLBOROUGH STREET
HANOVER SQUARE
HANOVER STREET
ARGYLL PLACE
WORK HOUSE
PORTLAND STREET
PUMP
BROAD STREET
DUFOURS PLACE
CAMBRIDGE ST
NEW ST
PULTENEY
LITTLE WINDMILL
GREAT P
GOLDEN SQUARE
OLD COMPTON
REGENTS QUADRANT
PICCADILLY
REGENT CIRCUS
QUEEN STREET
HAYMARKET
RUPERT STREET
BREWER
C. F. Cheffins, Lith. Southampton Bldgs. London.
SCALE 30 INCHES TO A MILE.

arch
研究院

Outline

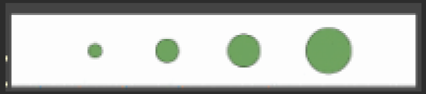
- Definition
- Objective
- History
- InfoVis 1.0
- **InfoVis 2.0**
- InfoVis 3.0
- Opportunities and challenges

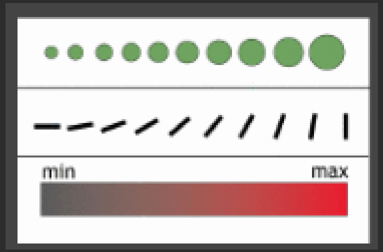
InfoVis 2.0: Overview



Basic Types of Visual Encodings

- "Retinal" properties
  - spatial position (e.g., x-y axes)
  - size
  - shape
  - color
  - orientation
  - texture
- "Gestalt" properties
  - connectivity
  - grouping (e.g., enclosure)
- Animation
  - view transitions
  - animated elements





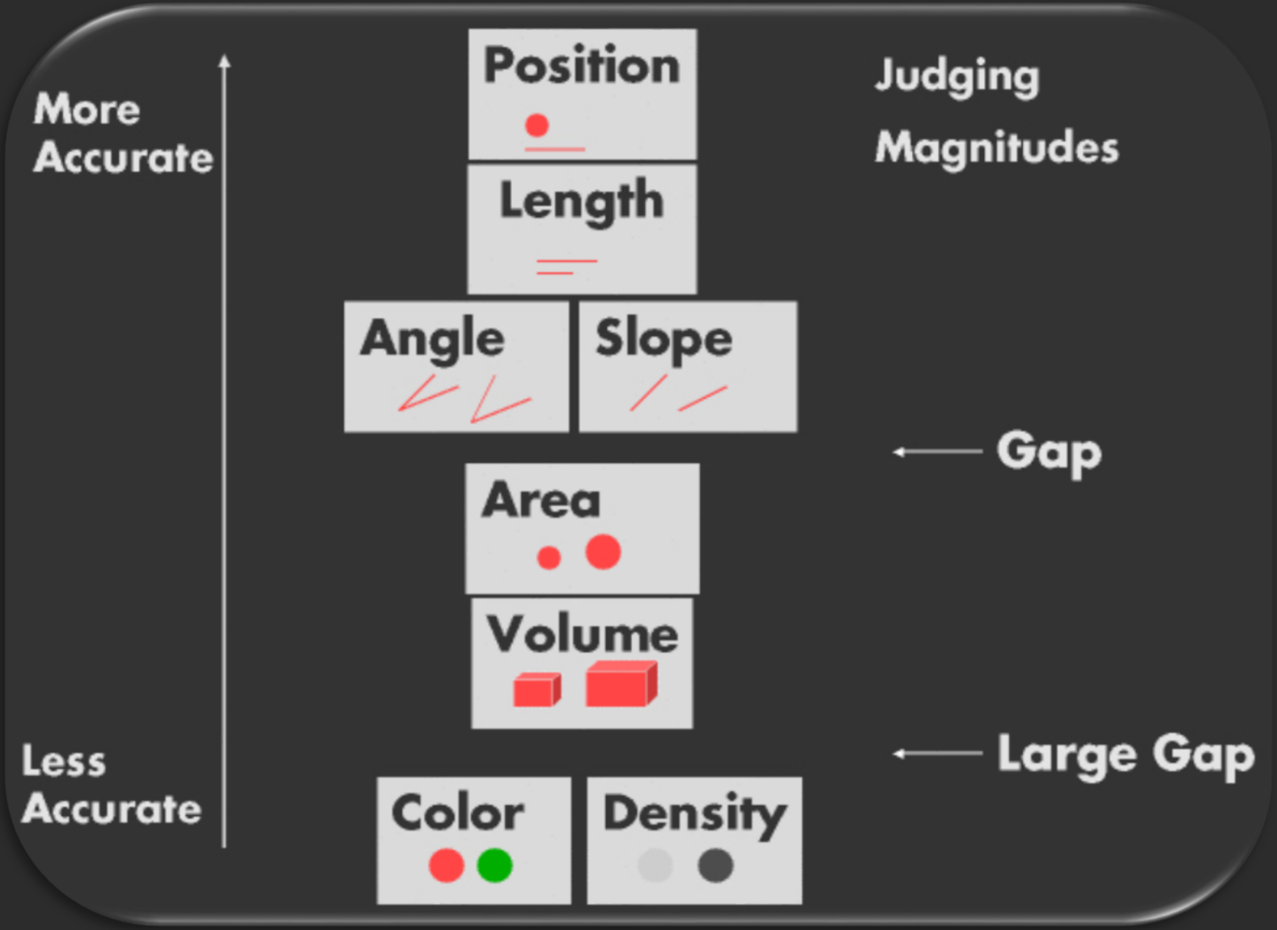
Later in the semester we'll examine guidelines on when to apply these mappings…

Pre-attentive Processing

- < 200 - 250ms qualifies as pre-attentive
  - eye movements take at least 200ms
  - yet certain processing can be done very quickly, implying low-level processing in parallel

- If a decision takes a fixed amount of time regardless of the number of distractors, it is considered to be preattentive.

Accuracy Ranking of Quantitative Perceptual Tasks Estimated; only pairwise comparisons have been validated (Mackinlay 88 from Cleveland & McGill)



Microsoft Research 微软亚洲研究院

Major Topics

- List
- Tree
- Graph
- Multidimensional

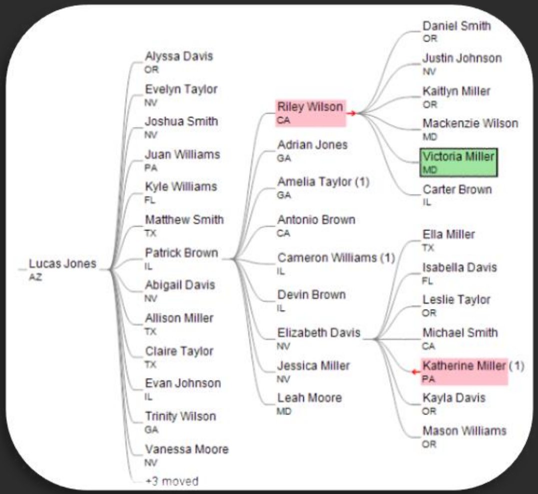
- Chart, timeline
- Tree visualization
- Graph visualization
- Multidimensional data viz

Microsoft Research 微软亚洲研究院

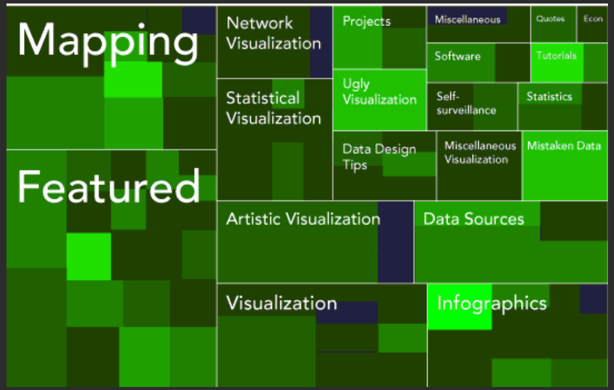
List Visualization

Tree Visualization

- Node-link diagram



- Space-filling



Tree Visualization: Hyperbolic Tree

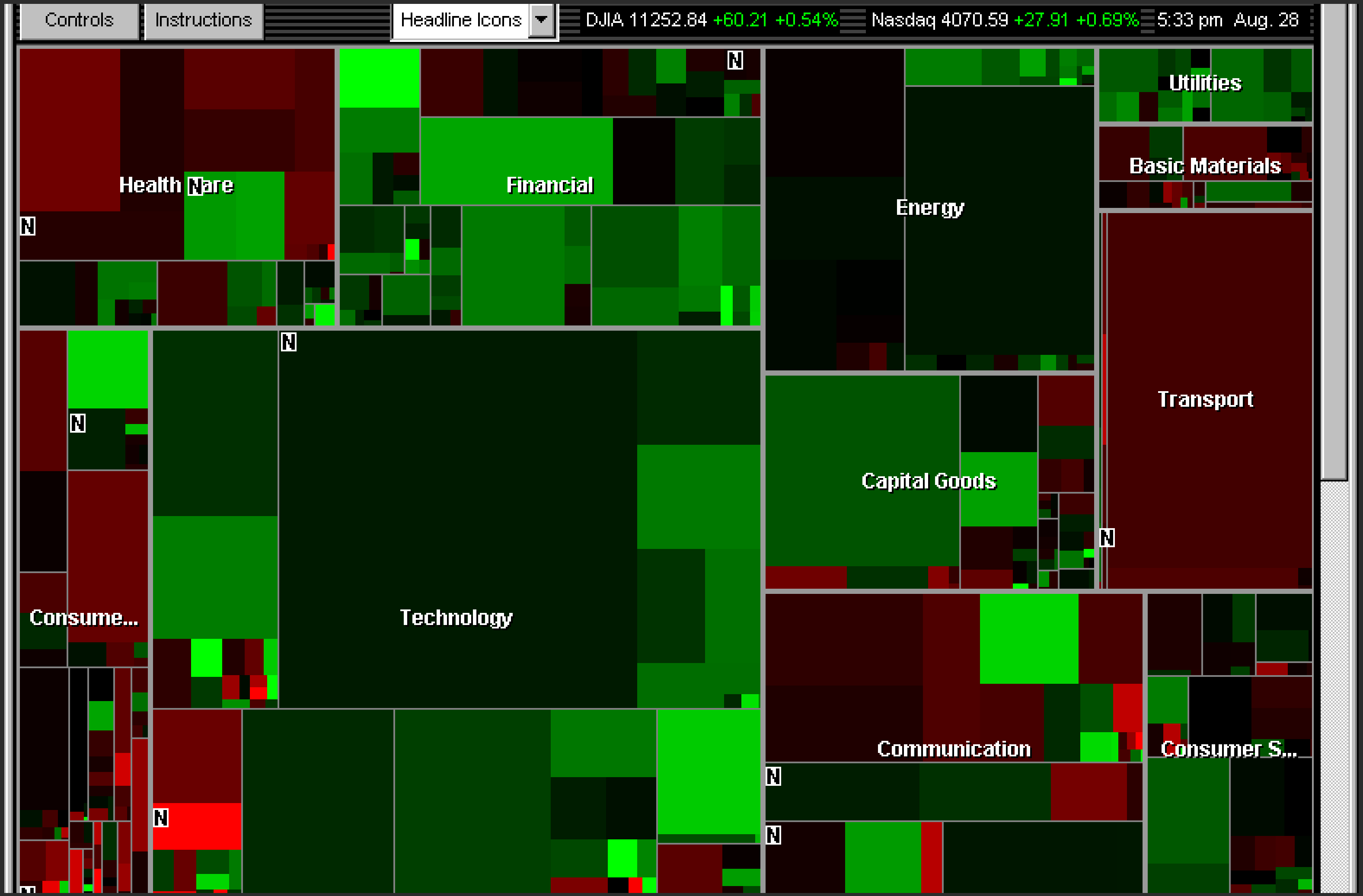
- Problems
  - Large hierarchies are everywhere
  - Scalability
- Basic idea
  - Focus + context (fisheye)
  - Lay out the hierarchy on a hyperbolic plane and map this plane onto a circular display region

Microsoft Research 微软亚洲研究院

Hyperbolic Tree: Demo

VisualBluePage.exe

Tree Visualization: Treemap

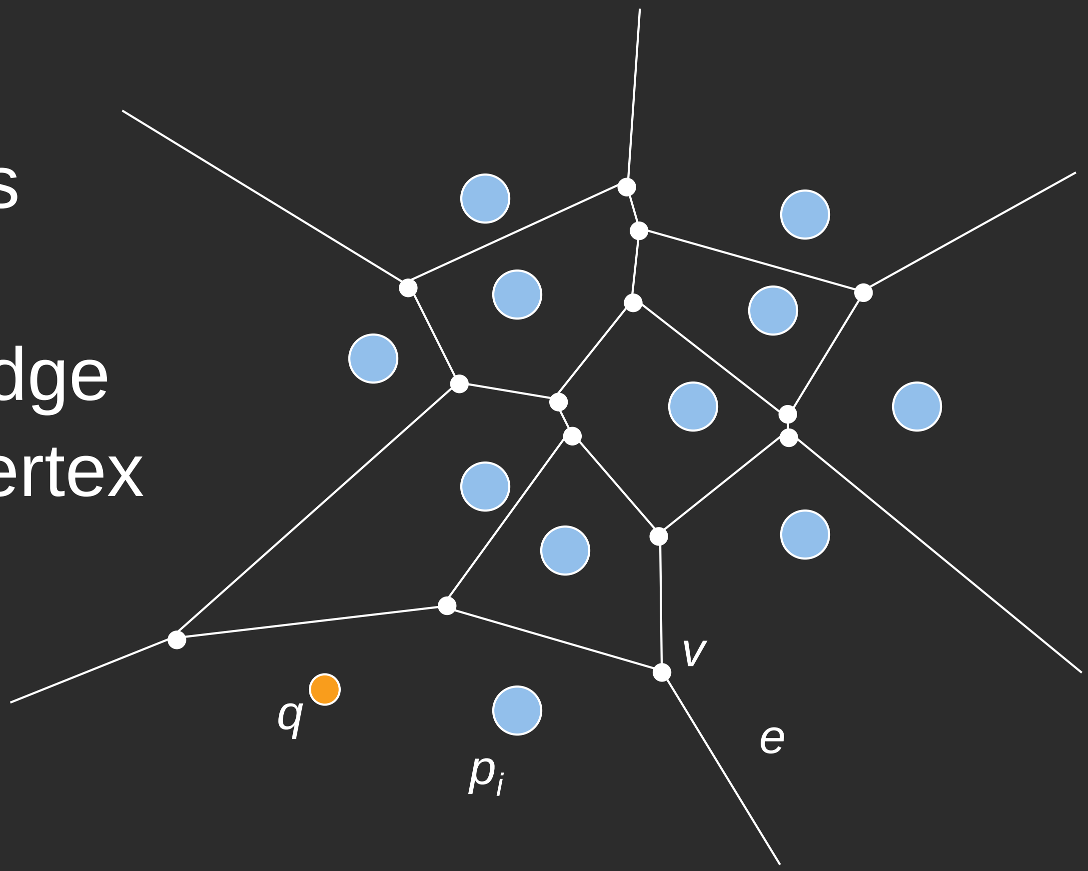


Voronoi Treemap

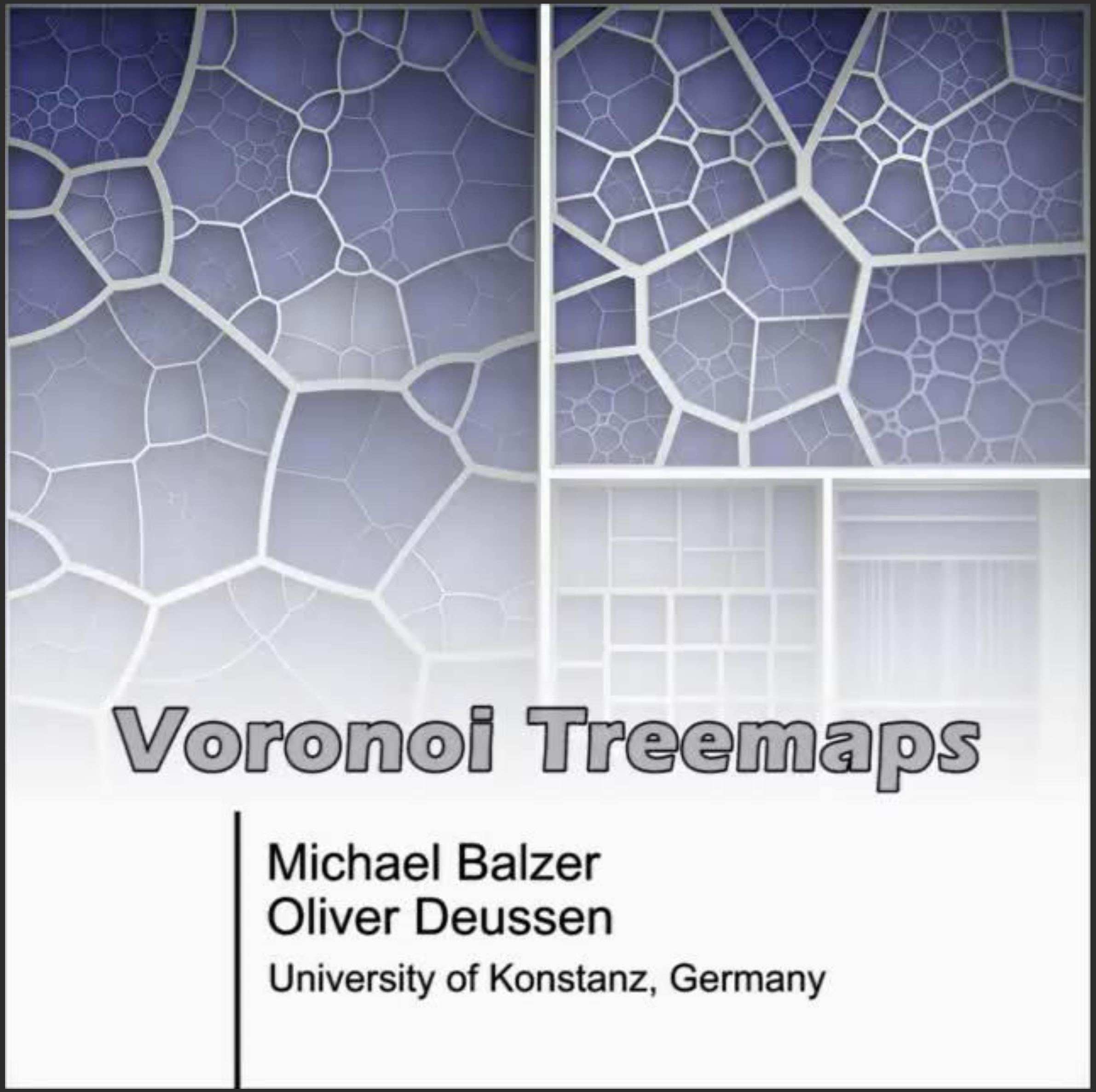
- Treemap Problems
  - Aspect ratios uncontrolled leads to lots of skinny boxes that clutter
  - Hard to understand the hierarchical structure
- Solution
  - Voronoi treemap

Post Office: What is the area of service?

p_i : site points
q : free point
e : Voronoi edge
v : Voronoi vertex

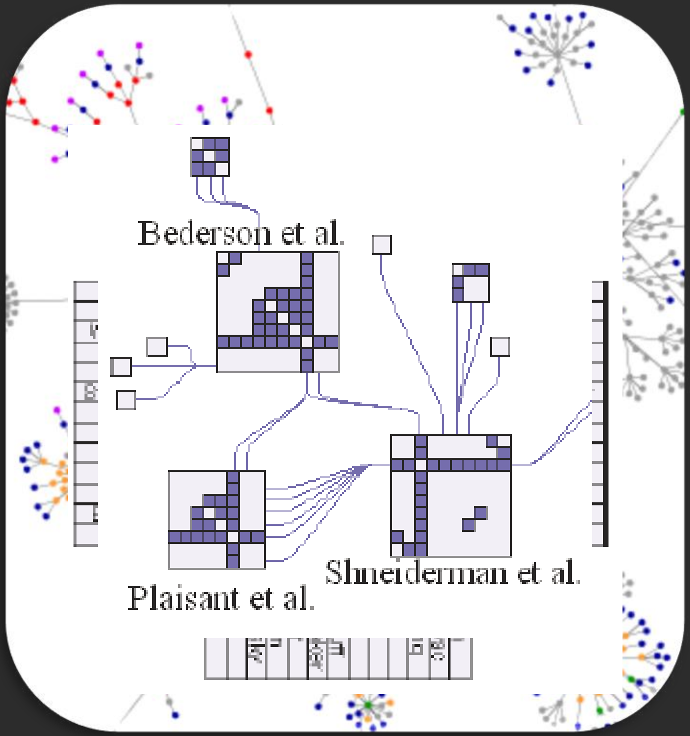


Demo


Voronoi Treemaps
Michael Balzer
Oliver Deussen
University of Konstanz, Germany

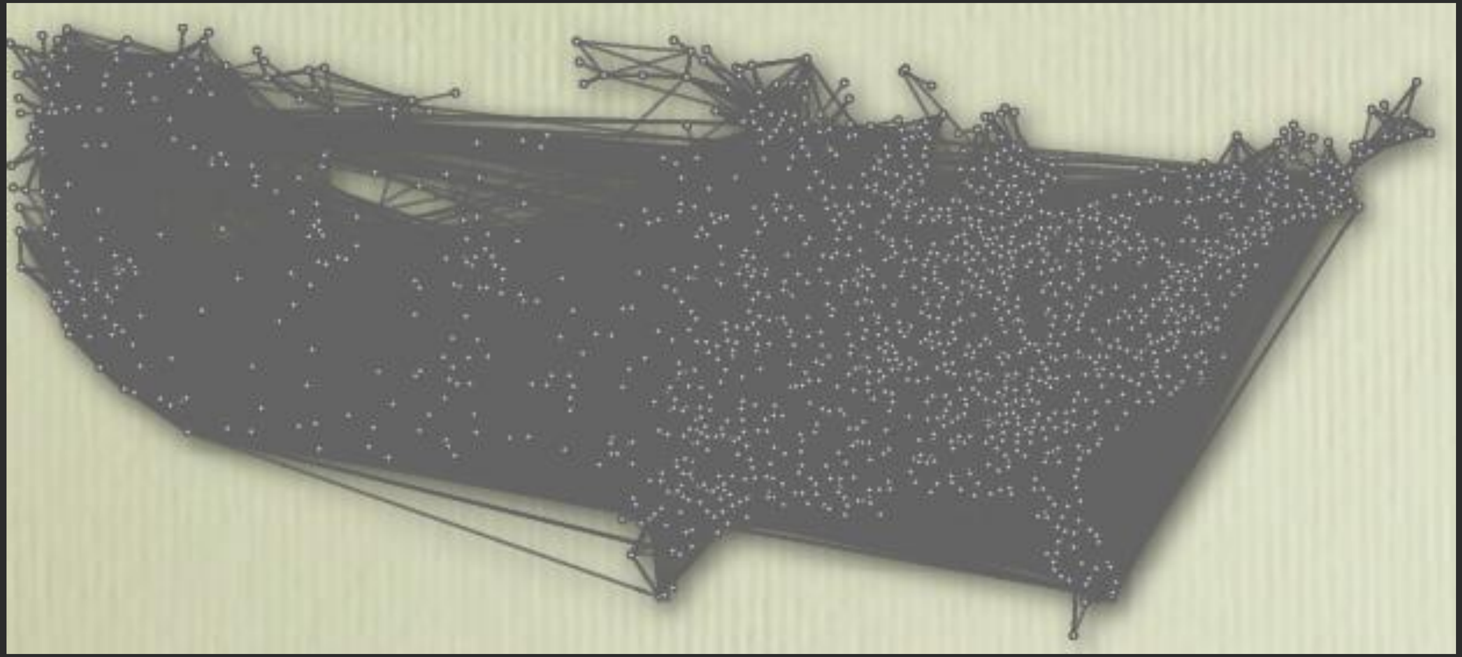
Graph Visualization

- Node-link diagram
- Matrix
- Hybrid



Clutter Reduction in Node-Link Diagram

- U.S. immigration graph with 1790 nodes and 9798 edges



Clutter Reduction Techniques

	Clutter reduction techniques
Appearance	sampling
	filtering
	change point size
	change opacity
	clustering
Spatial distortion	point/line displacement
	topological distortion
	space-filling
	pixel-plotting
	dimensional reordering
Temporal	animation

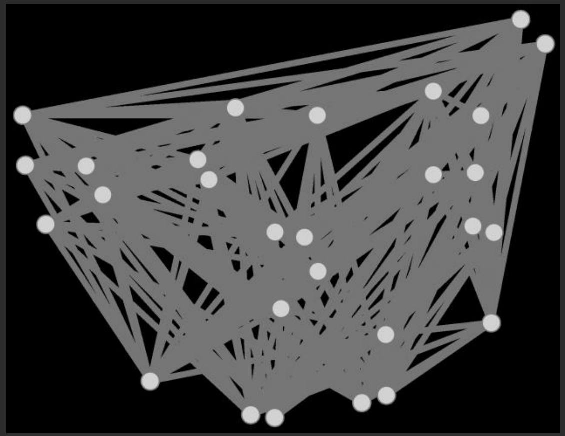
Clustering – Edge Bundling

- Visual clutter is a serious problem in huge graph visualization
- Solution
  - Geometry-based edge bundling

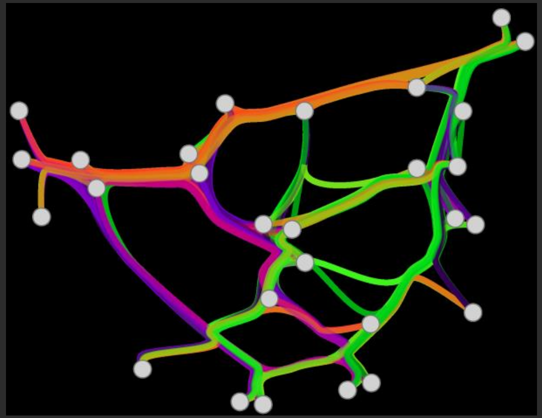
Microsoft Research 微软亚洲研究院

Inspiration

- Road-map style graphs
  - Cut straight lines into segments
  - Provide different levels of details
  - Familiar to everyone

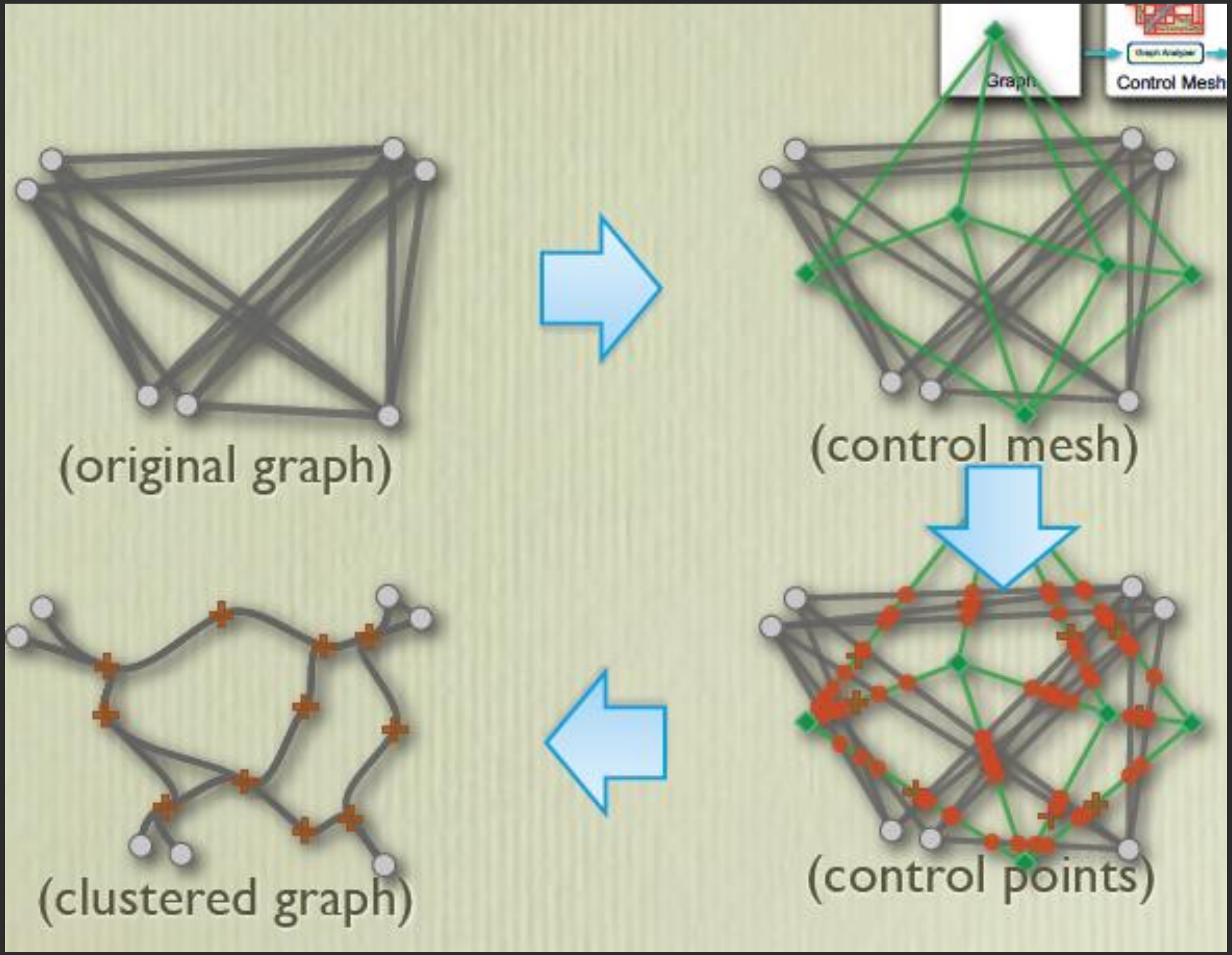


(straight line graph)



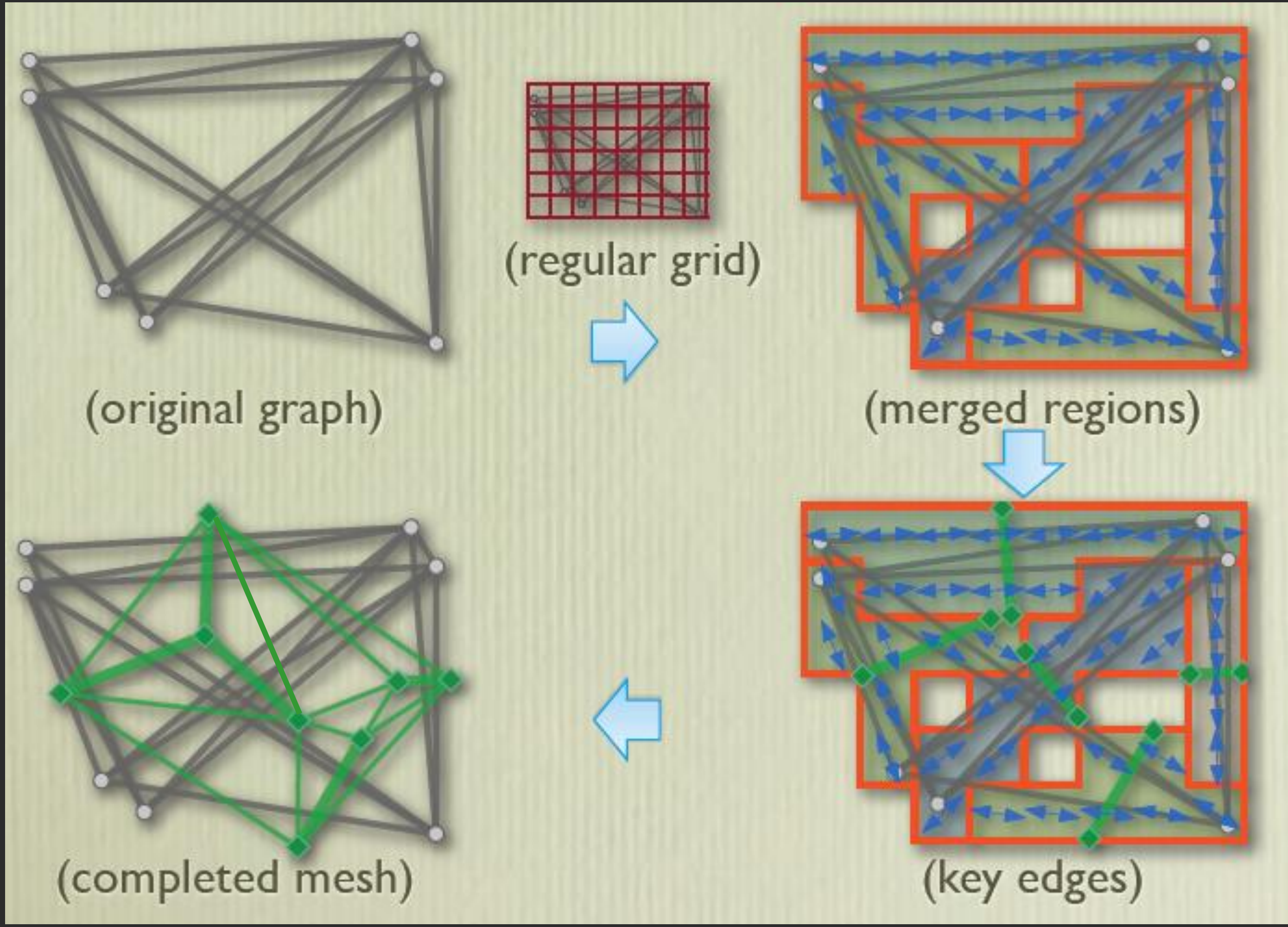
(road-map style graph)

Method



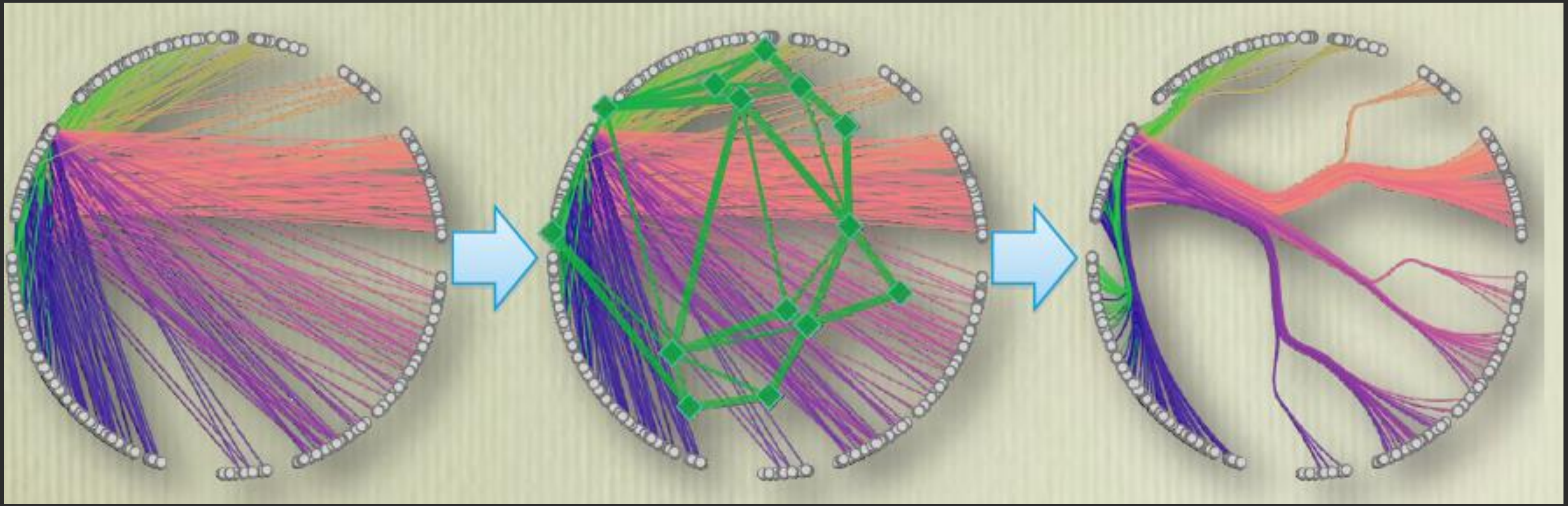
Graph
Control Mesh
(original graph)
(control mesh)
(clustered graph)
(control points)

Control Mesh



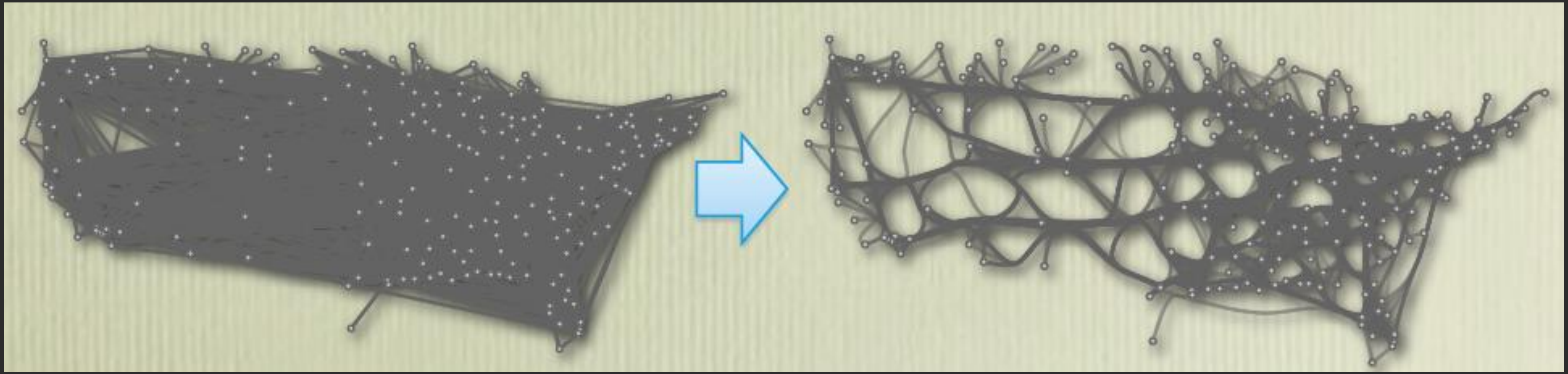
Example: Artificial Data

128 links, arranged in a circle, roughly from a source

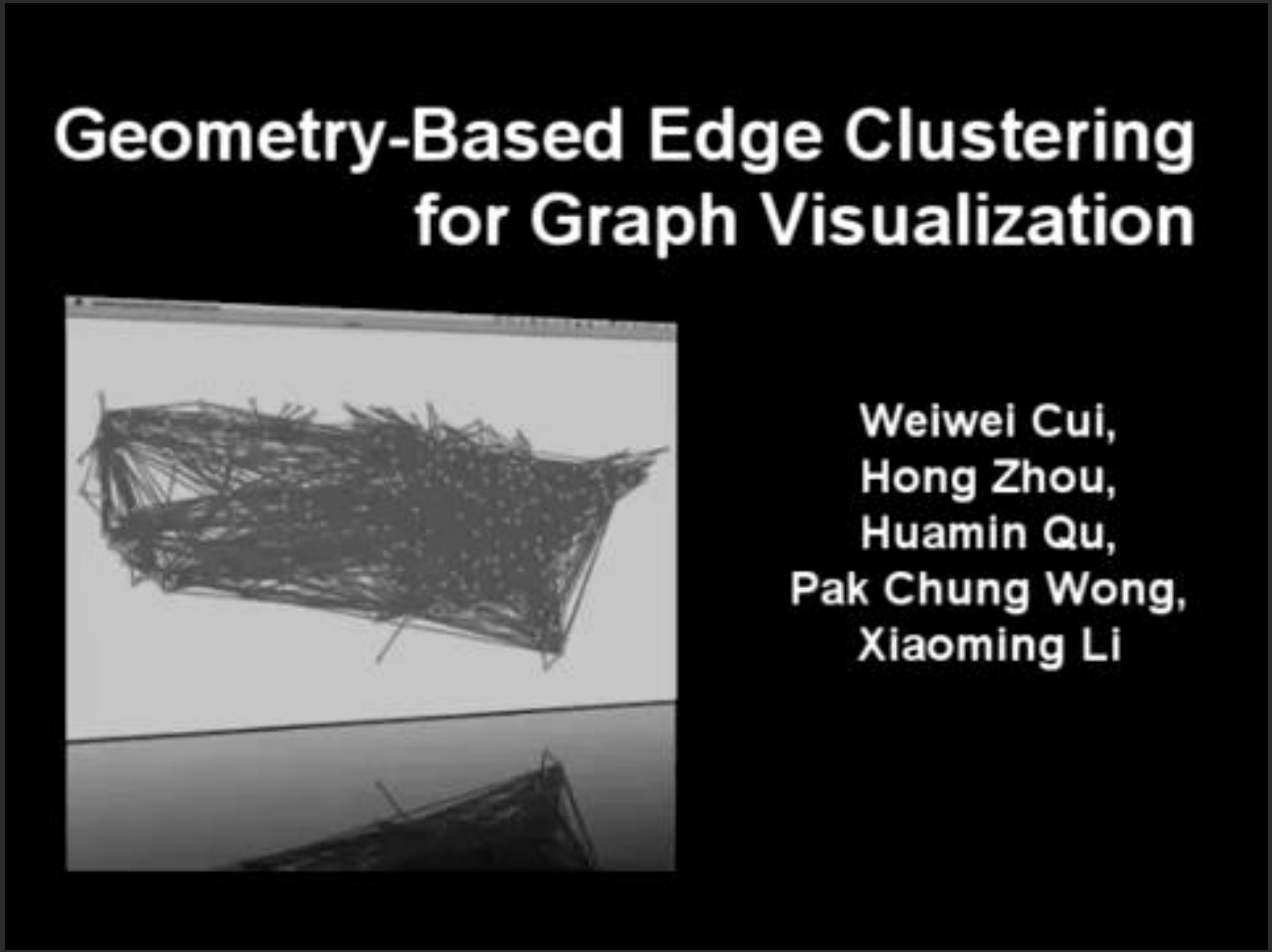


Example: Airlines

Northwest Airlines in the U.S.
2101 links, 235 nodes



Video


Geometry-Based Edge Clustering
for Graph Visualization
Weiwei Cui,
Hong Zhou,
Huamin Qu,
Pak Chung Wong,
Xiaoming Li

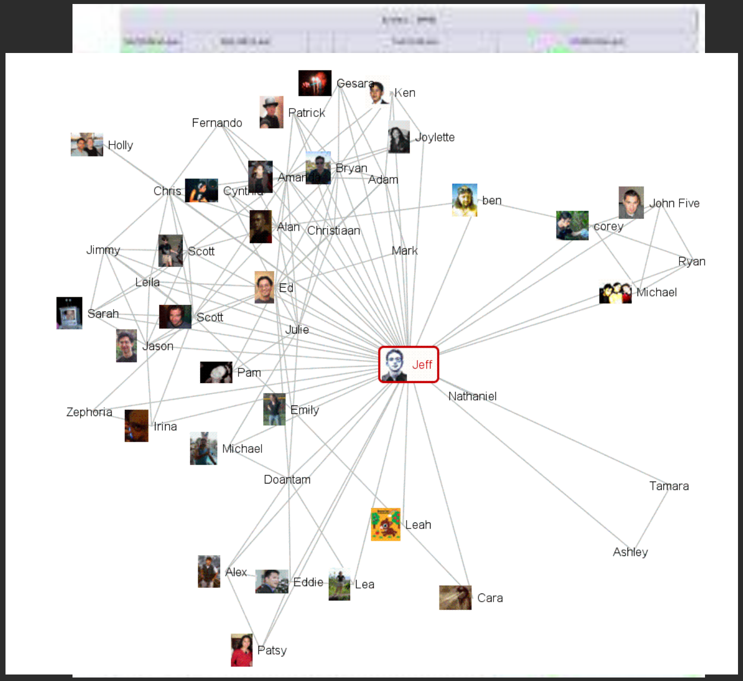
Hybrid Graph Visualization

- Problem
  - Huge amounts of rich context social media
  - These data involve multiple social relations and content context
- Challenges
  - Illustrate both global and local graph structures
  - Offer the multiscale and cross-scale exploration
  - Scalability

Hybrid Graph Visualization

Microsoft Research 微软亚洲研究院

- Traditional solutions
  - Force-directed layout
  - Matrix visualization
- New solution
  - Hybrid method

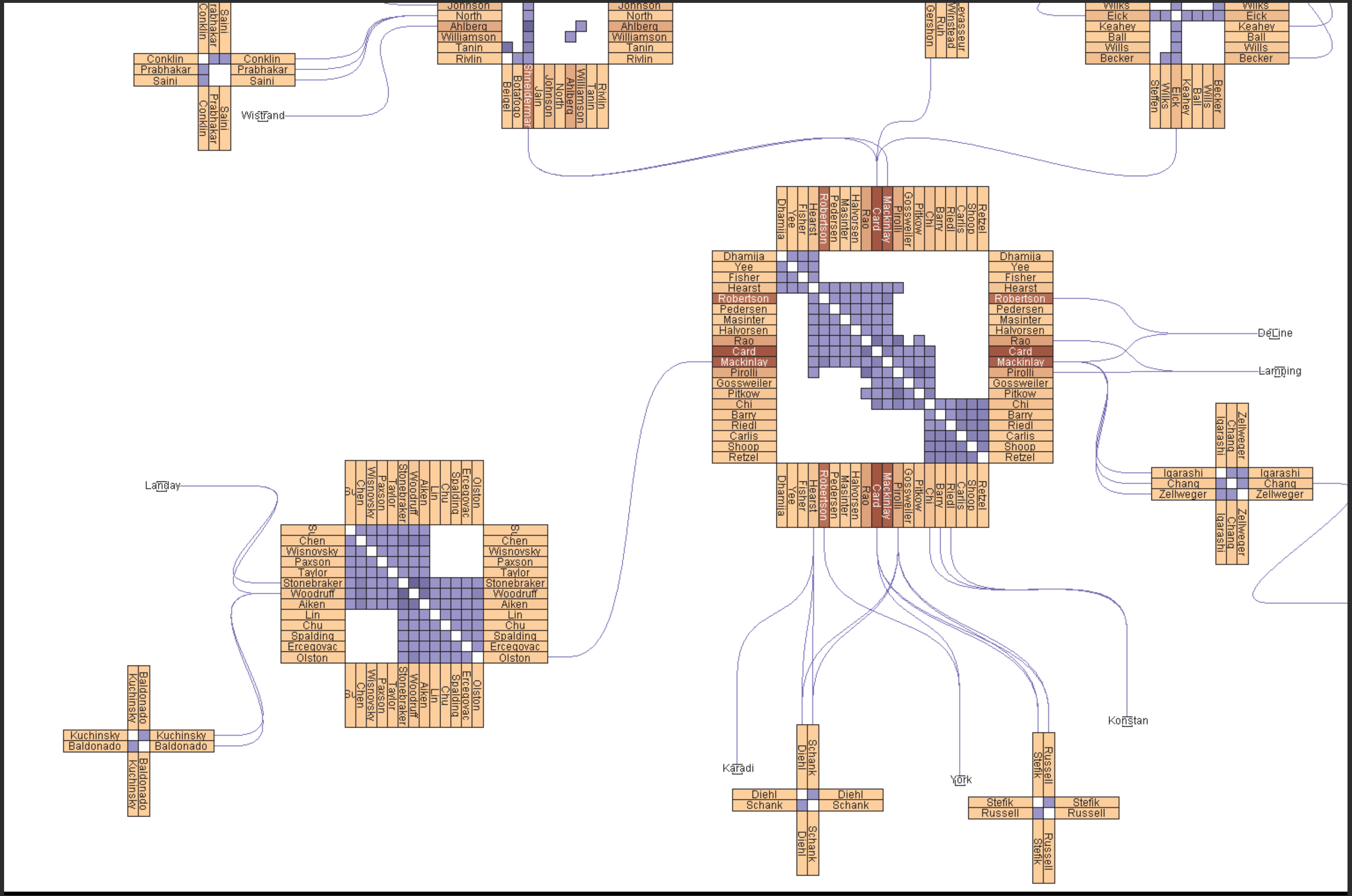


Hybrid Method: NodeTrix

- A hybrid visualization for analyzing social networks
  - Node-link diagrams: the global structure of a network
  - Adjacency matrices: better support the analysis of communities

Microsoft Research 微软亚洲研究院

NodeTrix



Video

Outline

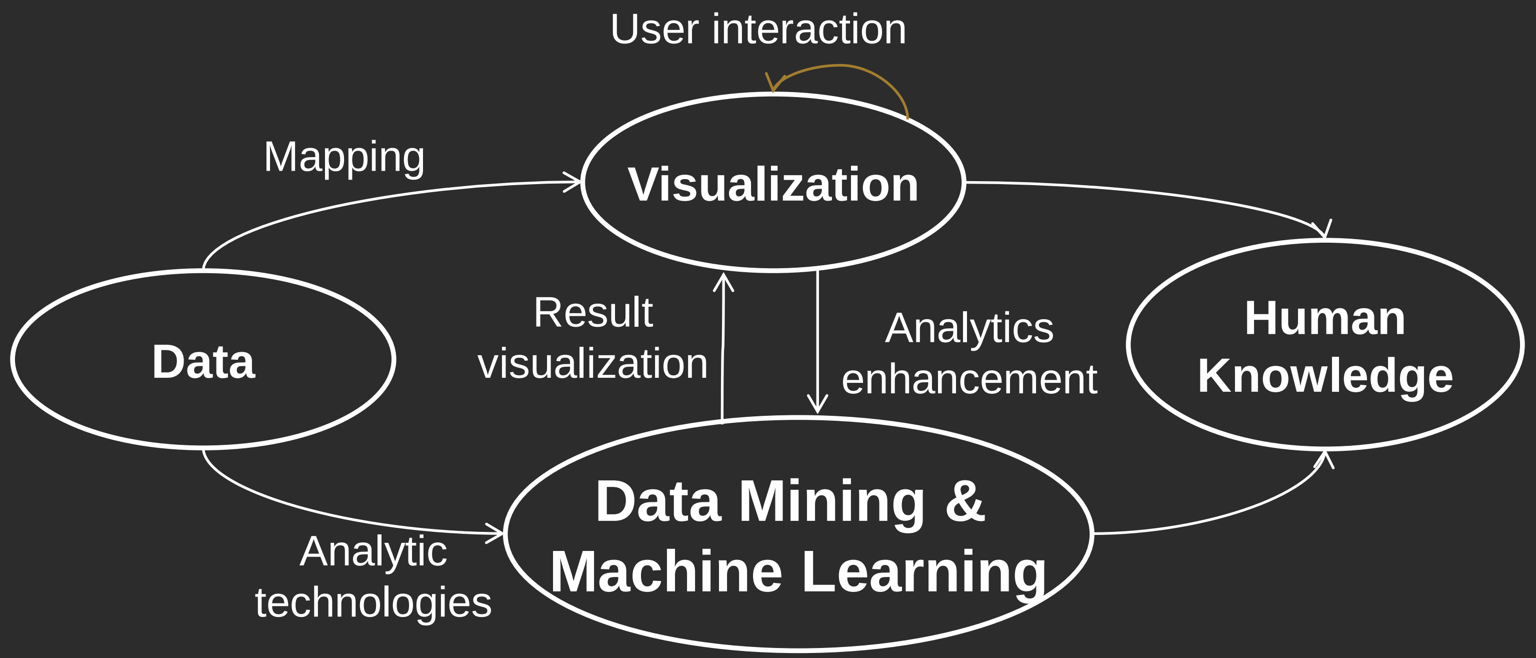
- Definition
- Objective
- History
- InfoVis 1.0
- InfoVis 2.0
- **InfoVis 3.0**
- Opportunities and challenges

InfoVis 3.0: Motivation

- Data/Information is everywhere
  - Structured + unstructured
- Data mining is important, however, it is far from perfect

Visual Analytics in a Nutshell

Research and develop visual analytics methods and systems to facilitate individual or collective human decision making

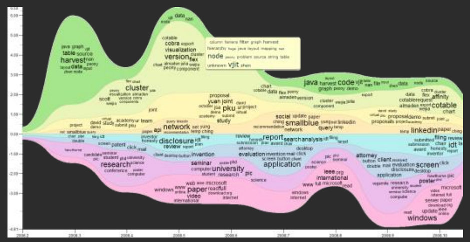


Visual Analytics @MSRA

- Visual text analytics
- Visual social analytics
- Visual search log analytics

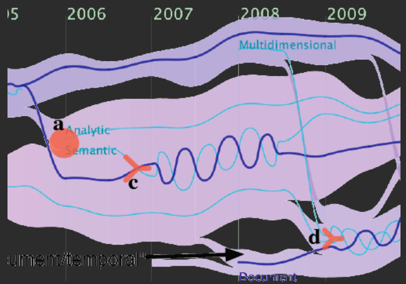
Topic-based Text Visualization

TIARA

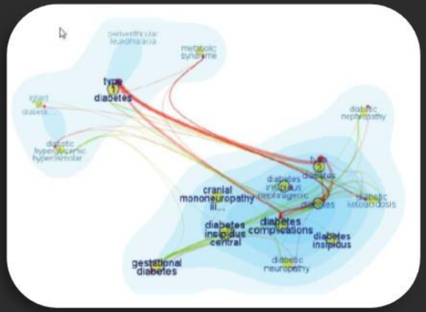


ACM TIST ' 12, KDD'10, TKDE' 12

TextFlow

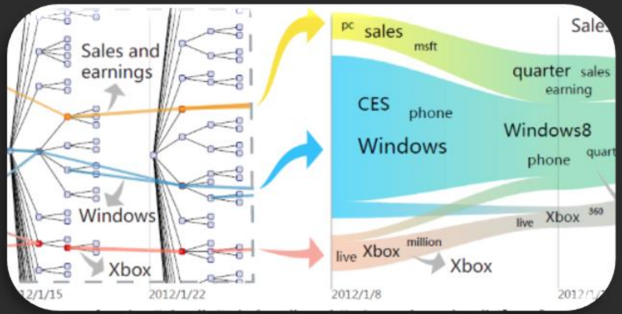


InfoVis' 11



FacetAltas
InfoVis' 10

Hierarchical topic evolution

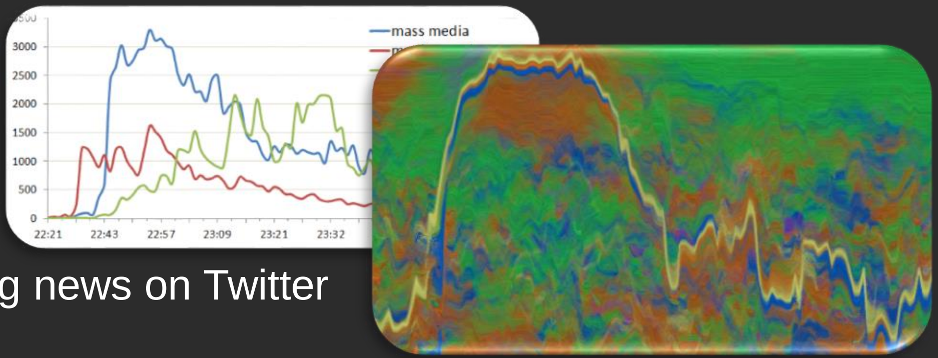


KDD'13

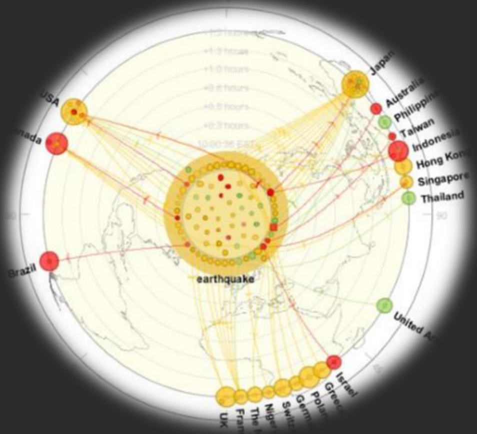
ViSizer
TVCG'13

Microsoft Research 微软亚洲研究院

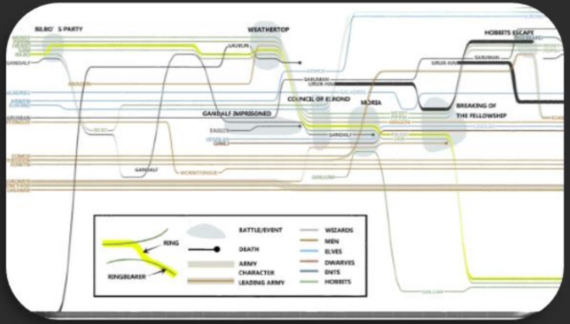
Visual User Behavior Analytics



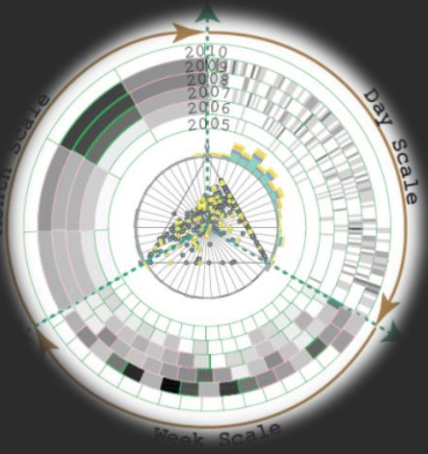
Breaking news on Twitter
CHI'12



Whisper
InfoVis'12

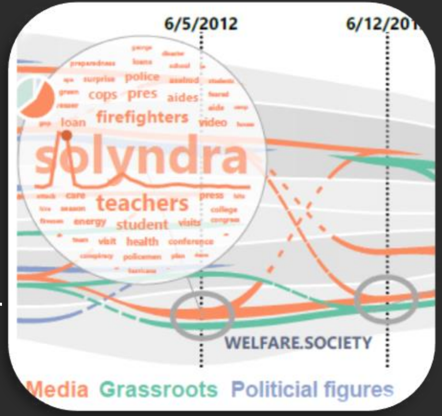


StoryFlow
Infovis'13



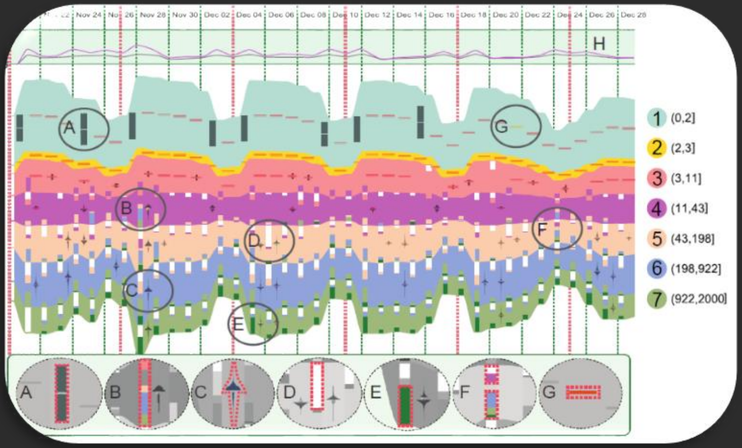
OpinionSeer
InfoVis' 10

Issue competitor
VAST'13



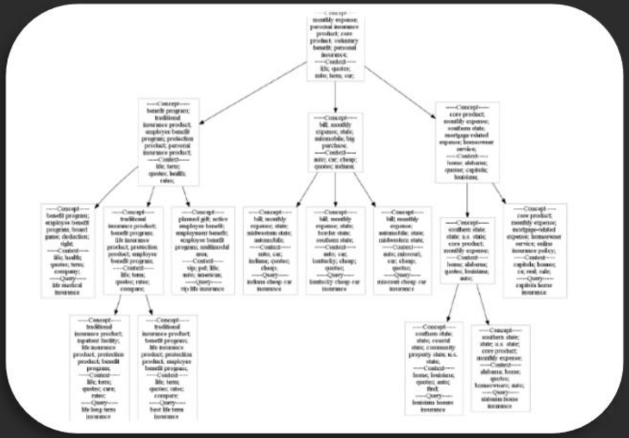
Search Log Analytics

RankExplorer



InfoVis’12

Taxonomy Construction



KDD’12

Microsoft Research 微软亚洲研究院

Given a Large Document Collection

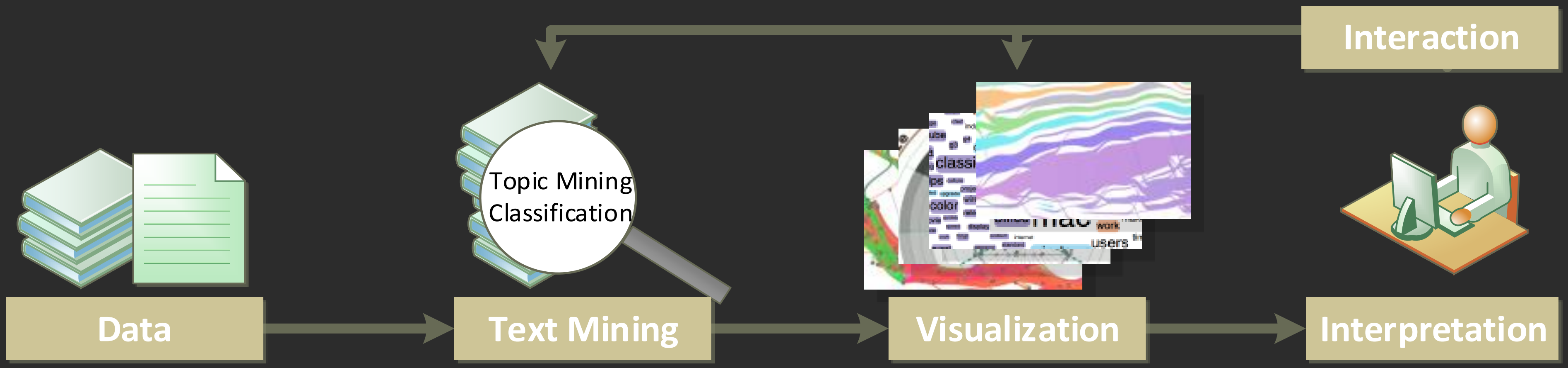
Big Question

- What can information visualization provide to help users in understanding and gathering information from text and document collections?

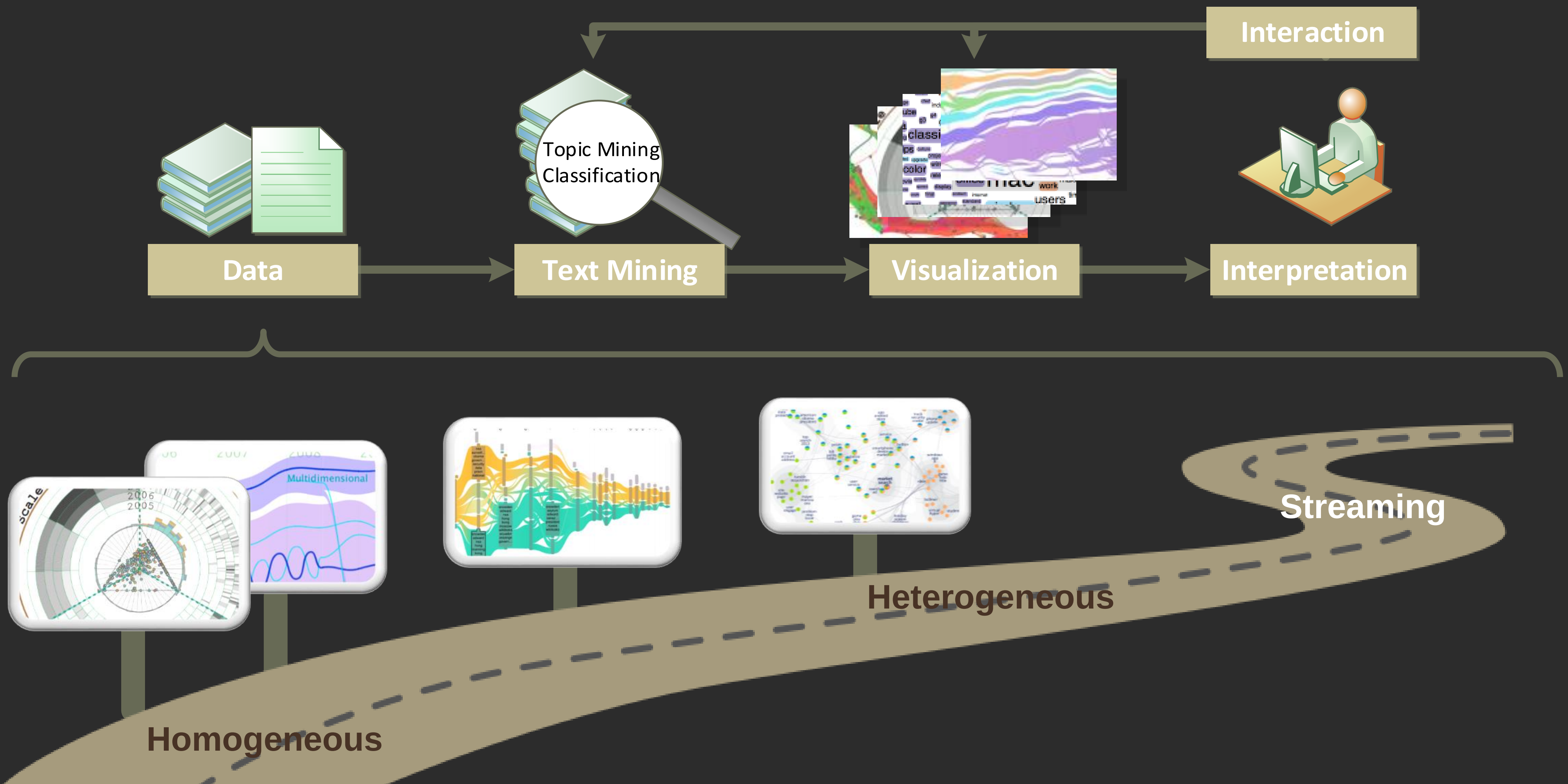
Microsoft Research 微软亚洲研究院

Visual Text Analytics: Pipeline

Combine advanced text analytics and interactive visualization to help end users in the decision making process

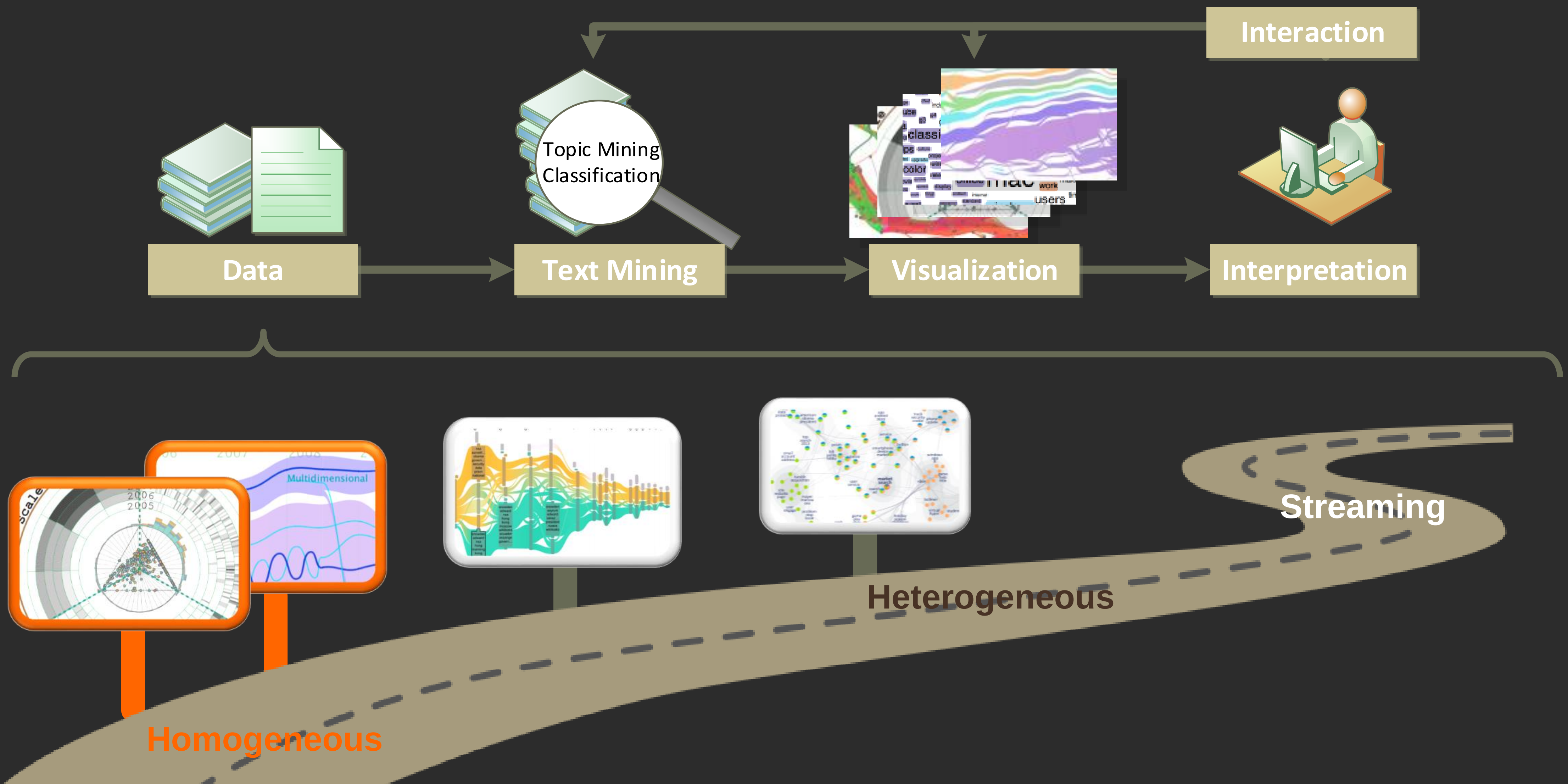


Our Focus: Data



Microsoft Research 微软亚洲研究院

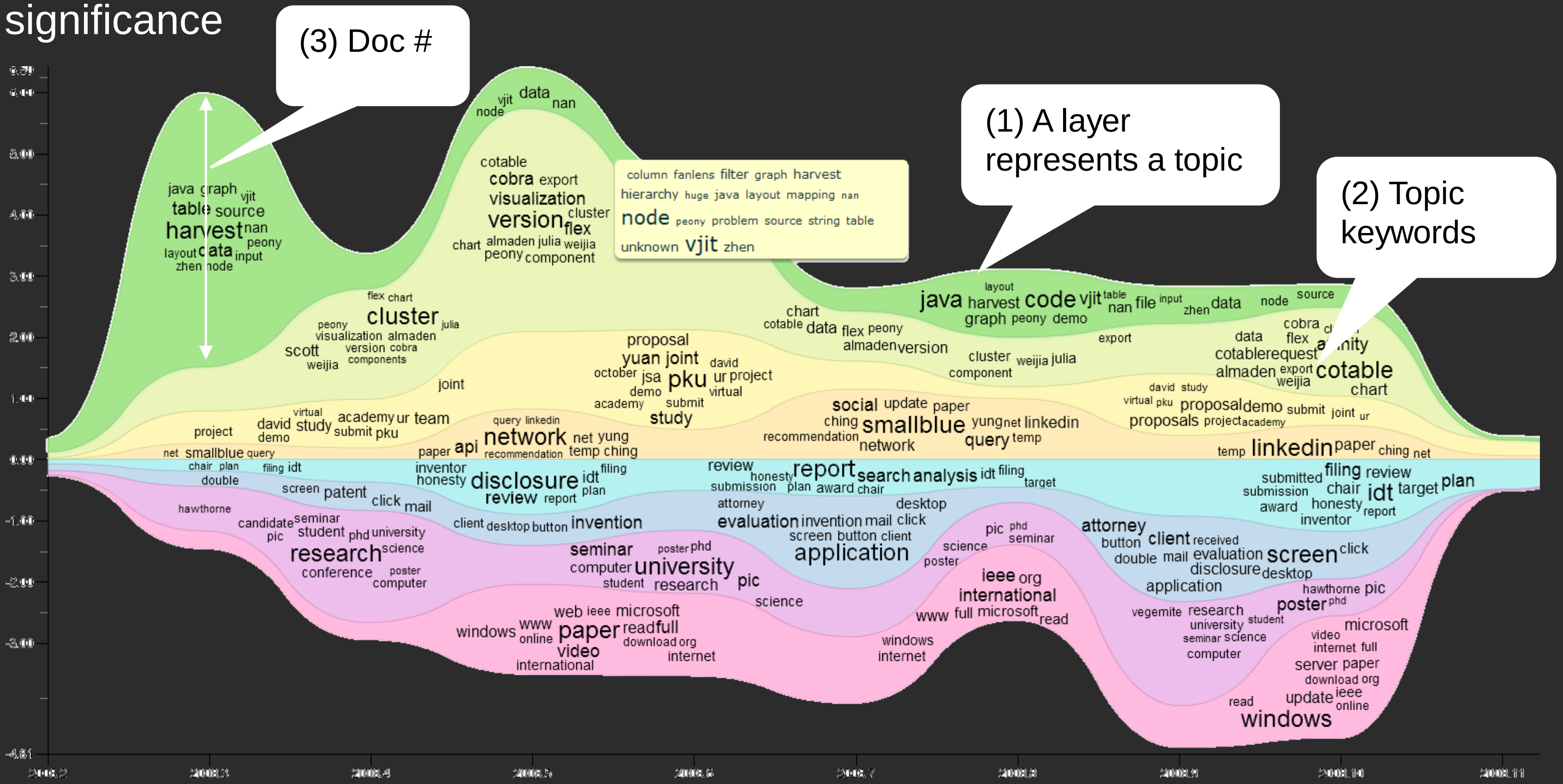
Our Focus: Data



Microsoft Research 微软亚洲研究院

Demo

Y axis encodes topic significance



X axis encodes time

~10,000 emails in 2008

Interactive, Time-based Visual Email Summarization

Shixia Liu, Michelle X Zhou, Shimei Pan,
Weihong Qian, Weijia Cai, Xiaoxiao Lian

IBM Research

Key Challenges

- Summarize text corpora
  - Huge amounts of complex information
  - Time-varying
- Visually explain summarization results
  - Consistent visualization
- Provide feedback or articulate their needs
  - Imperfect summarization results or varied user needs

Microsoft Research 微软亚洲研究院

TIARA Overview

Data collection

Text Preprocessing

Text content + meta data

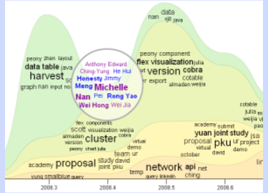
Text Summarization

Summarization results

Data Transformation

Transformed results

Visualization



User Interaction

Microsoft Research 微软亚洲研究院

TIARA Overview

Data collection

Text Preprocessing

Text content + meta data

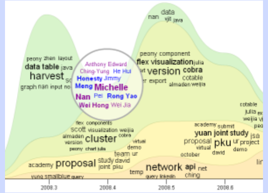
Text Summarization

Summarization results

Data Transformation

Transformed results

Visualization



User Interaction

Text Summarization

- Latent Dirichlet Allocation (LDA) model [Blei et al. 03]
  - High portability
  - High compaction rate for scalability
  - A finer grained model

*k*th document in the collection

$\{\ldots, P(T_i \mid D_k), \ldots\}$

A set of topic *probabilities*

$\{T_1, \ldots, T_i, \ldots T_N\}$

A set of topics

A set of keywords

$\{W_1, \ldots, W_j, \ldots, W_M\}$

A set of word *probabilities*

$\{\ldots, P(W_j \mid T_i), \ldots\}$

LDA Data Transformation

A set of topics

kth document in the collection

$\{\ldots, P(T_i \mid D_k), \ldots\}$

A set of topic *probabilities*

$\{T_1, \ldots T_i, \ldots T_N\}$

A set of keywords

$\{W_1, \ldots, W_j, \ldots, W_M\}$

A set of word *probabilities*

$\{\ldots, P(W_j \mid T_i), \ldots\}$

Rank the topics to present most interesting ones first

Select keyword sub-sets for time-based summary

$\{\ldots\}_{t-1}, \{\ldots, W_j, \ldots\}_t, \{\ldots\}_{t+1},$

[Song et al. CIKM'09]

Topic Ranking by User Interests

- Rank topics by "strength"

$$rank(T_k) = f\left(\mu(T_k), \sigma(T_k), \alpha(T_k)\right)$$

Domain-dependent activeness metric

topic coverage

$$\mu(T_k) = \frac{\sum_{m=1}^{M} N_m \hat{\theta}_{m,k}}{\sum_{m=1}^{M} N_m}$$

topic variance

$$\sigma(T_k) = \sqrt{\frac{\sum_{m=1}^{M} N_m \left(\hat{\theta}_{m,k} - \mu(T_k)\right)}{\sum_{m=1}^{M} N_m}}$$

- Rank topics by "distinctiveness"

$$rank(T_k) = l(T_k) = \frac{\tilde{v}_k^T L \tilde{v}_k}{\tilde{v}_k^T D \tilde{v}_k}$$

graph Laplacian

graph degree matrix

doc-topic distribution

for each T_k, $v_k = \left(\hat{\theta}_{1,k}, \hat{\theta}_{2,k}, \ldots, \hat{\theta}_{M,k}\right)^T$

$\tilde{v}_k$ is normalized v_k

Microsoft Research 微软亚洲研究院

Experiments

- Goal
  - Measure which metric produces more “important” topics
- Data sets
  - Email: 8326 email messages
  - News: 34,690 documents
- Method
  - Users indicate the importance of top-K ranked topics
    - Very important, somewhat important, Unimportant

Results

- Email data (by F1 measure)

Retrieved	Top 5	Top 10
Strength	0.800 ± 0.000	0.620 ± 0.028
Distinctiveness	$\mathbf{1.000 \pm 0.000}$	$\mathbf{0.780 \pm 0.028}$
M.I.	0.760 ± 0.106	0.740 ± 0.035
T.S.	0.440 ± 0.057	0.480 ± 0.028

- News data (by F1 measure)

Retrieved	Top 5	Top 10
Strength	0.640 ± 0.057	0.68 ± 0.028
Distinctiveness	**0.760 ± 0.057**	**0.76 ± 0.035**
M.I.	**0.760 ± 0.057**	0.74 ± 0.035
T.S.	0.720 ± 0.069	0.70 ± 0.045

Microsoft Research 微软亚洲研究院

TIARA Overview

Data collection

Text Preprocessing

Text content + meta data

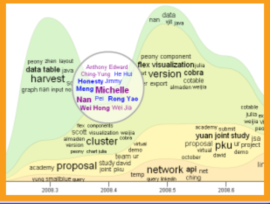
Text Summarization

Summarization results

Data Transformation

Transformed results

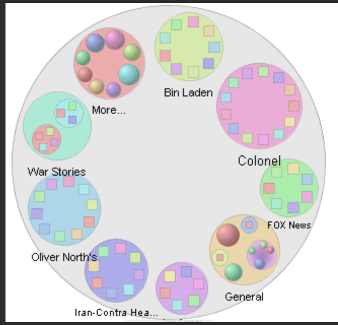
Visualization

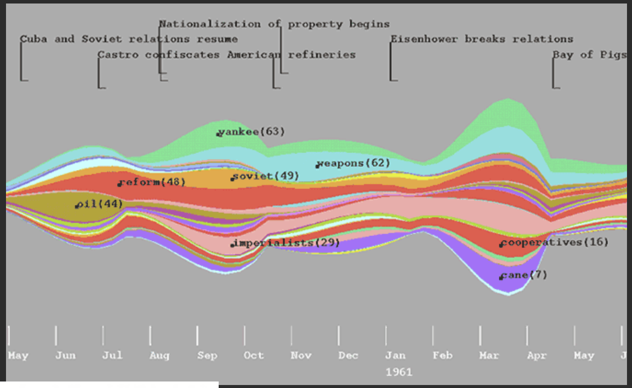


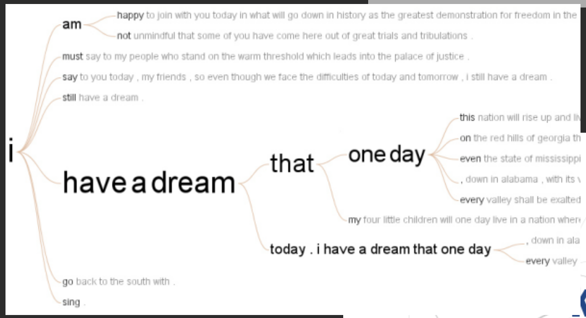
User Interaction

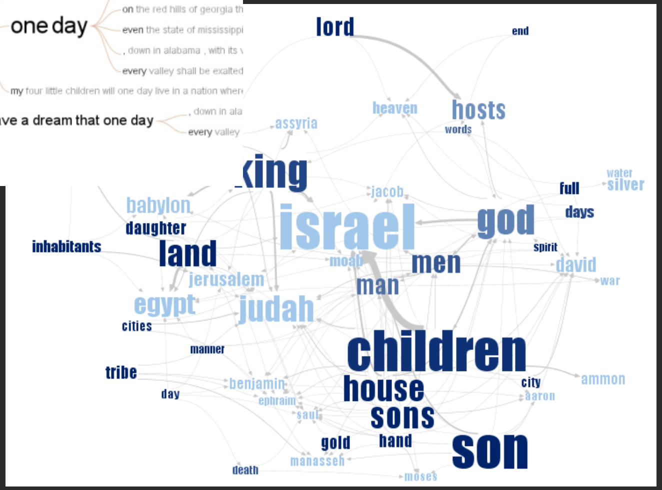
Visualizing Text: Existing Work

- Visualize text at a high level
- Visualize text at a low level
- Few on explaining advanced text analysis results









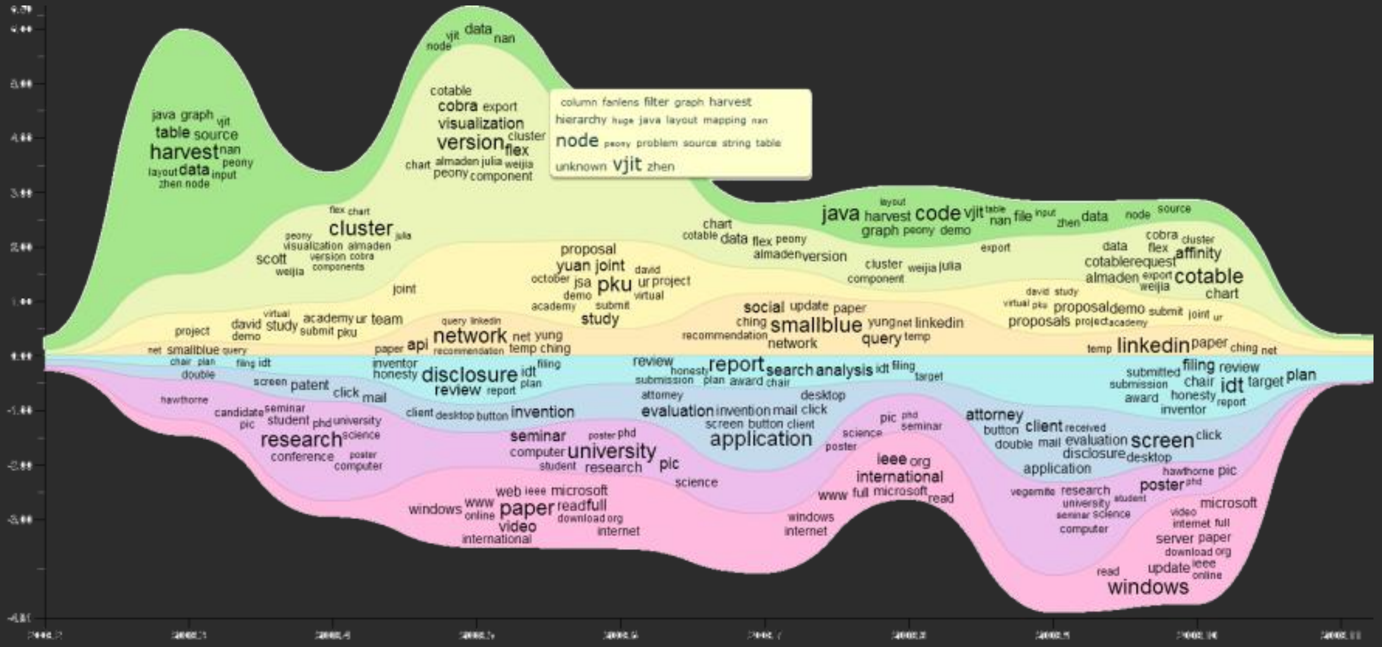
Microsoft Research 微软亚洲研究院

Visual Text Summary Metaphor

Data to be visualized:

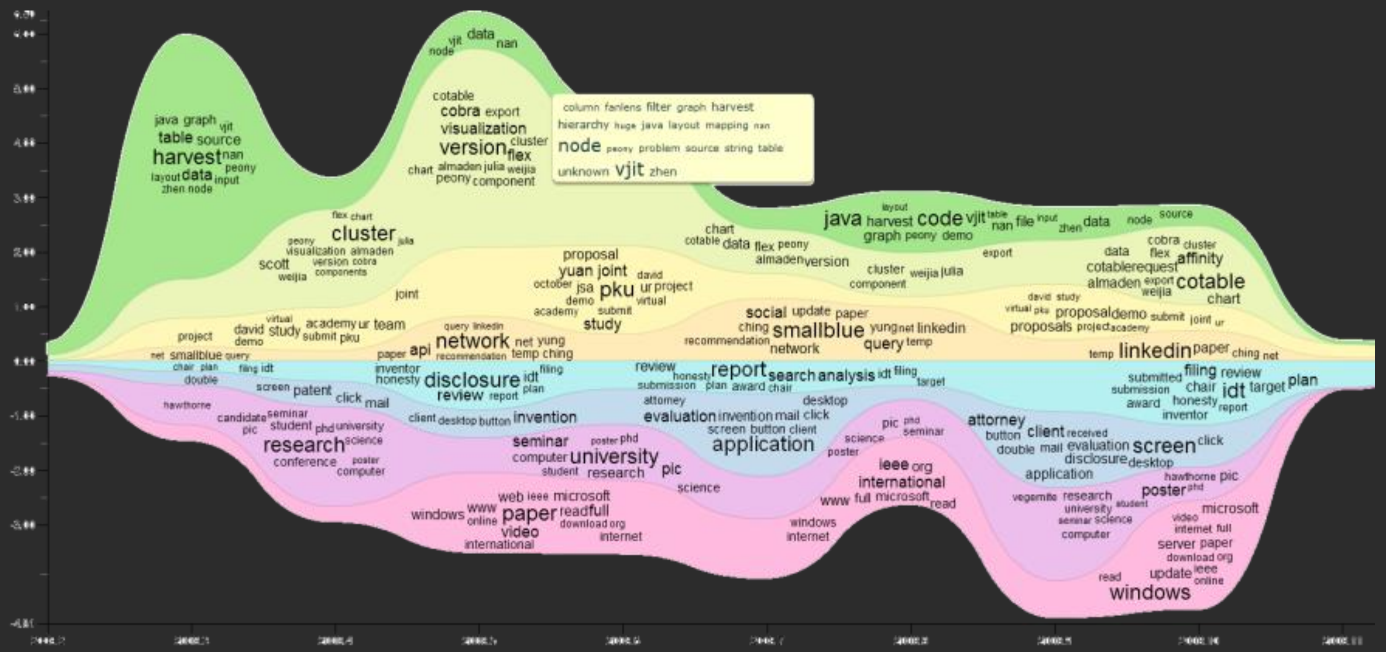
1. Topics: {T_1, …T_i,… T_N} and their probabilities
2. For each T_i ,Topic keywords by time: … {…,w^i_k, …, }$_t$, … and their probabilities over time
3. For each T_i , Topic strength: {..., $S^i(t)$, …} over time

Visual encoding: Augmented stacked graph



Enhanced Stacked Graph: Key Steps

- Computing geometry of layers
- Layer coloring
- Layer ordering
- Layer labeling

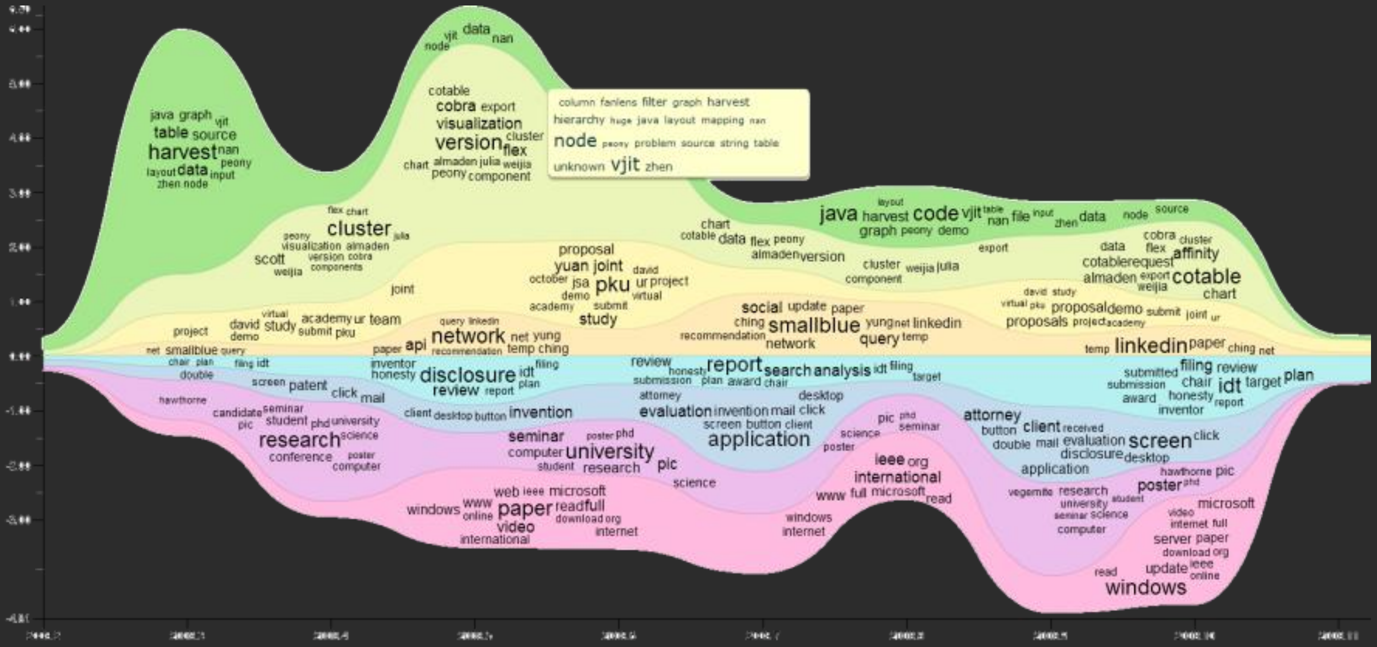


Microsoft Research 微软亚洲研究院

Enhanced Stacked Graph: Key Steps

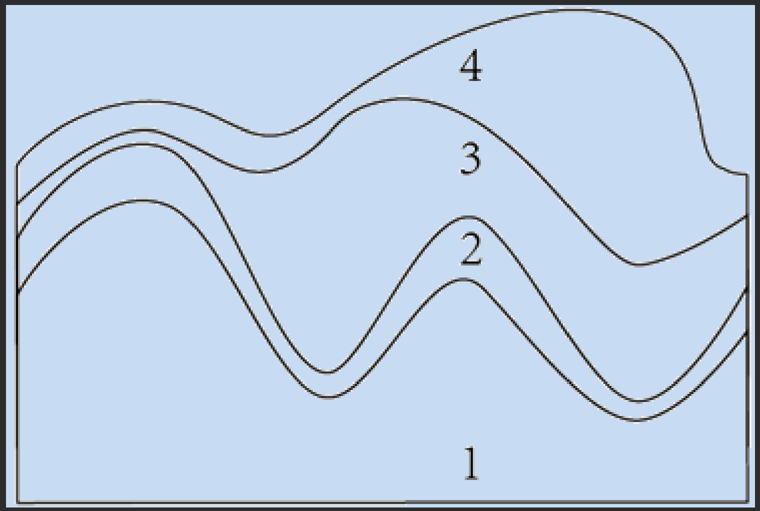
- Computing geometry of layers
- Layer coloring
- Layer ordering
- Layer labeling

Byron_Infovis08

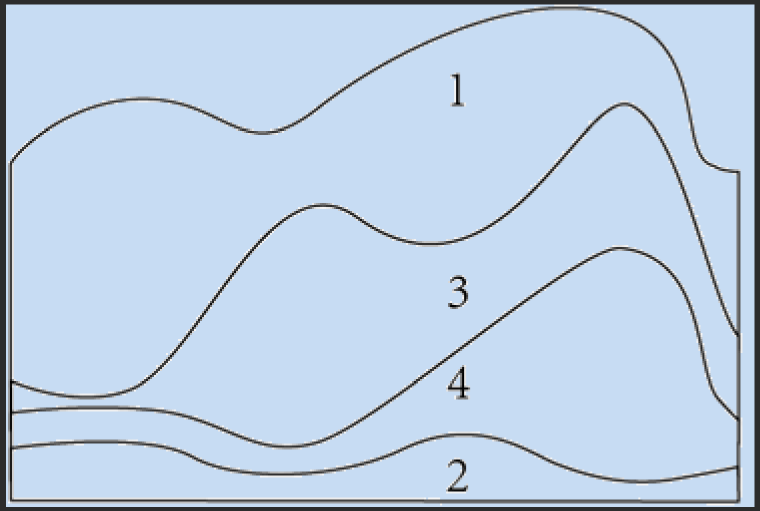


Layer Ordering

- Goals
  - Minimize distortion
  - Maximize usable space
  - Ensure semantic coherence



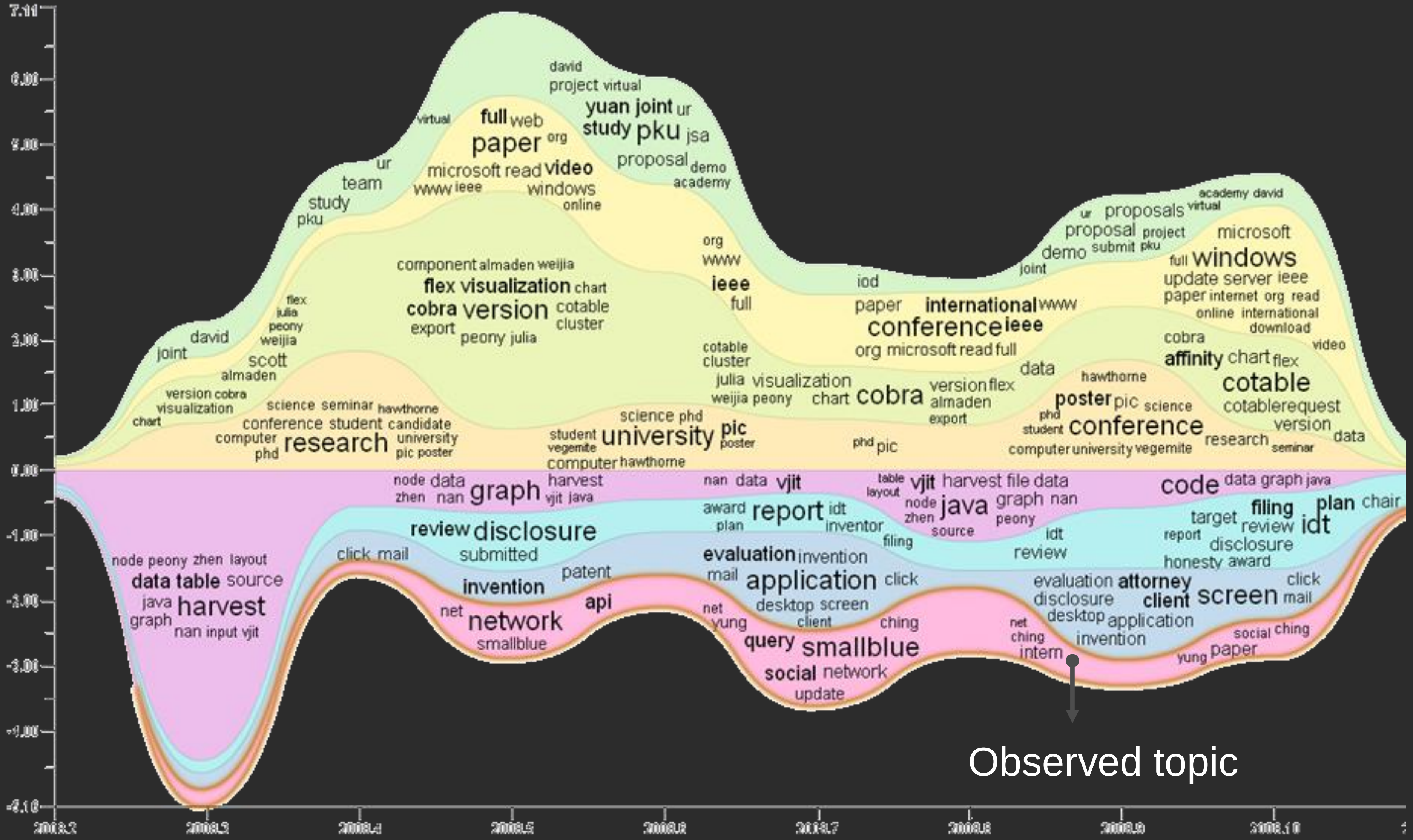
unordered



ordered

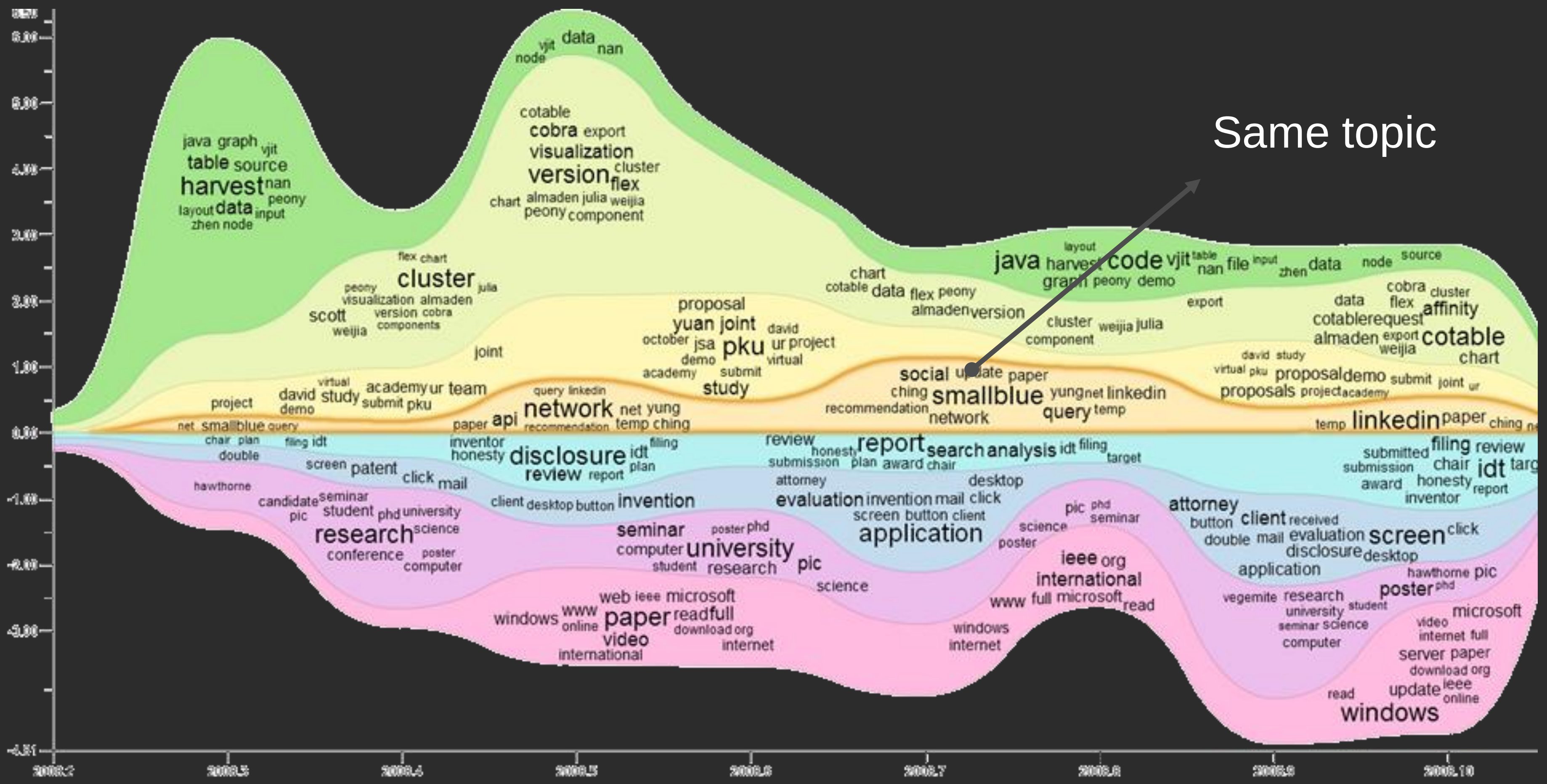
Microsoft Research 微软亚洲研究院

Layer Ordering - Comparison



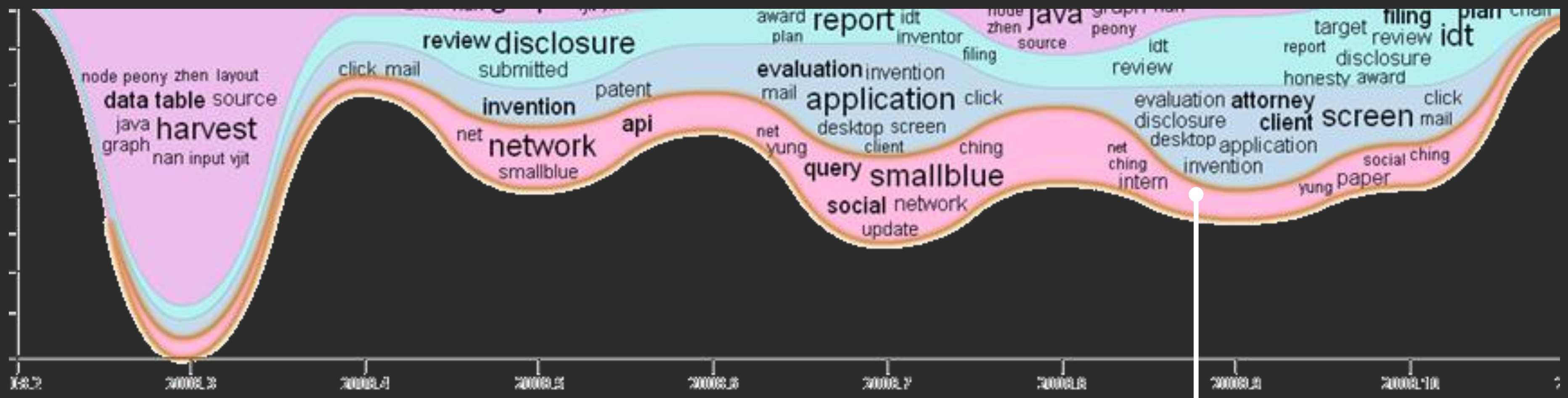
Microsoft Research
微软亚洲研究院

Layer Ordering - Comparison

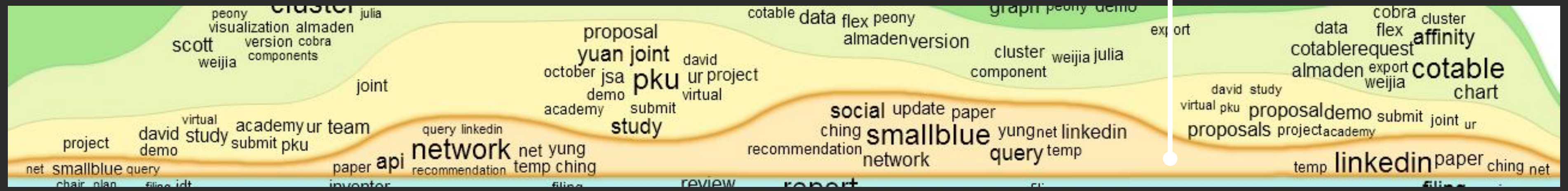


Same topic

Layer Ordering - Comparison



Same topic

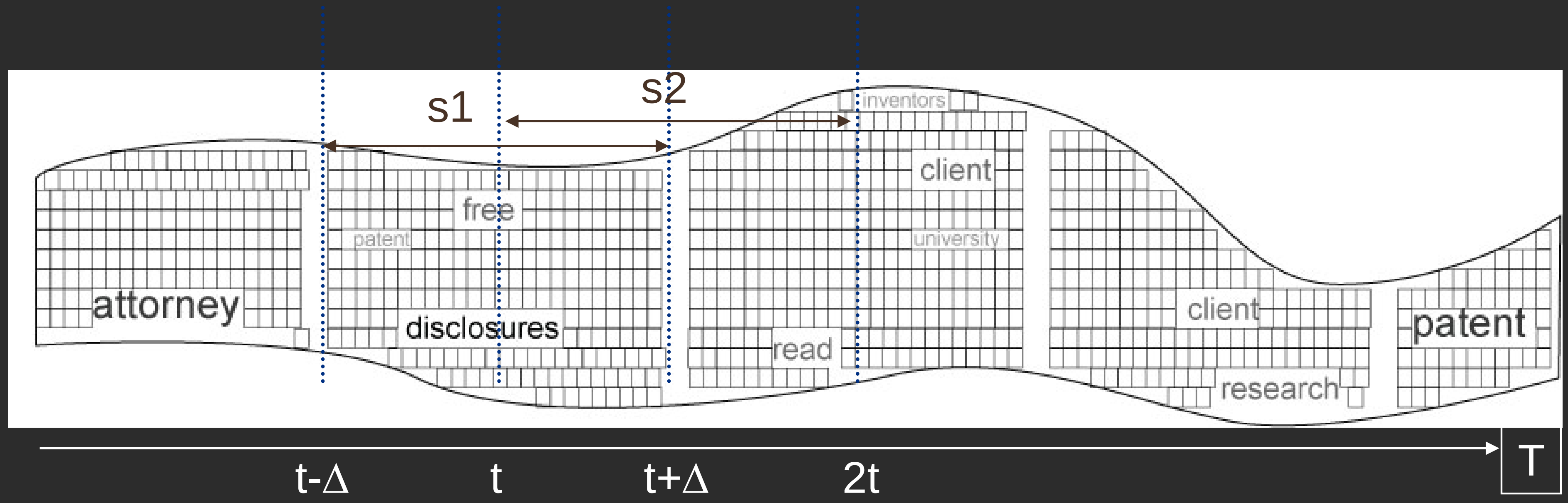


Enhanced Stacked Graph: Layer Labeling

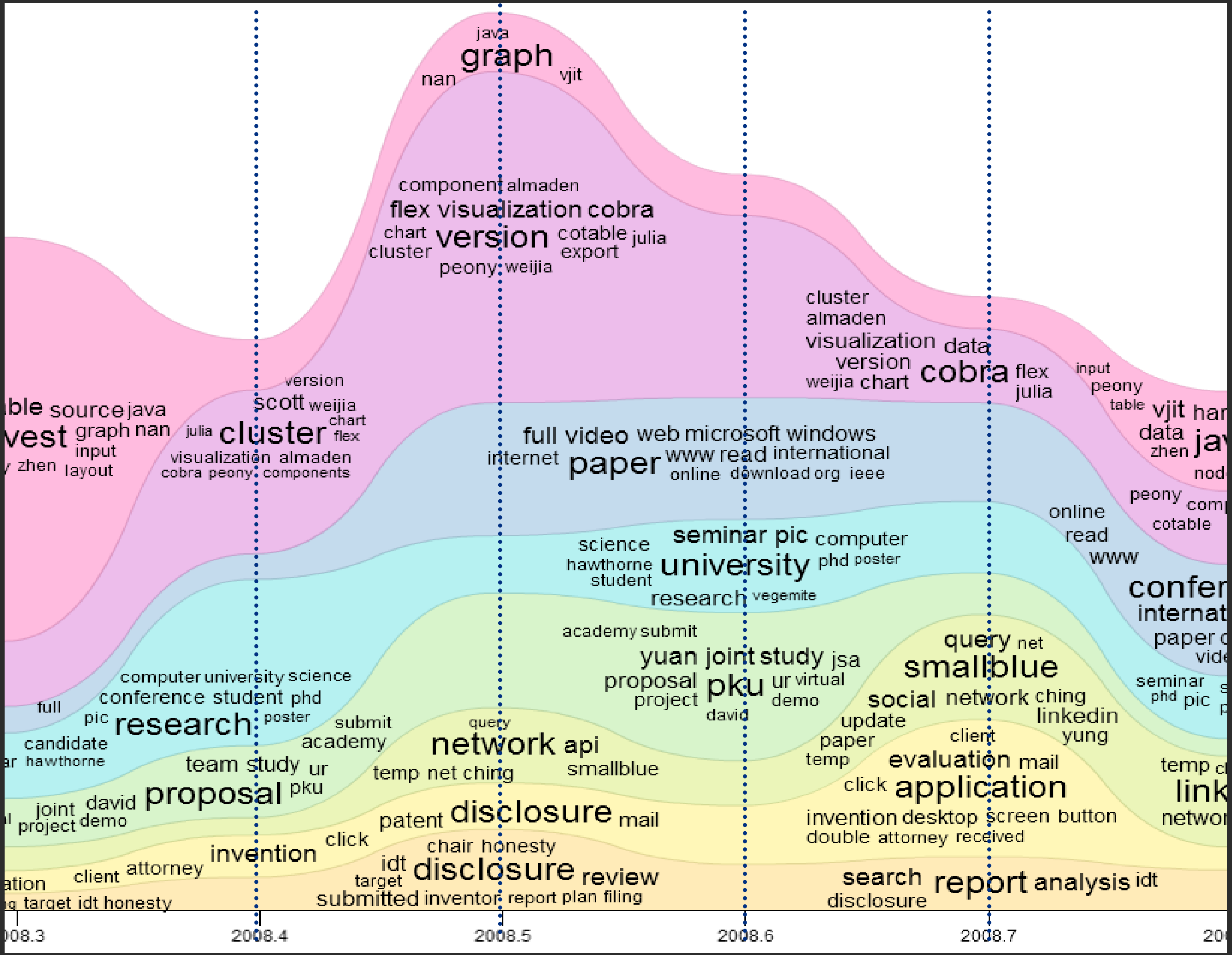
- Goals
  - Temporal proximity
  - Informativeness

Enhanced Stacked Graph: Layer Labeling (cont'd)

- Our approach [Liu et al. CIKM09]
  - Constraint-based space allocation
  - Particle-based layout [Luboschik et al. 08] + wordle

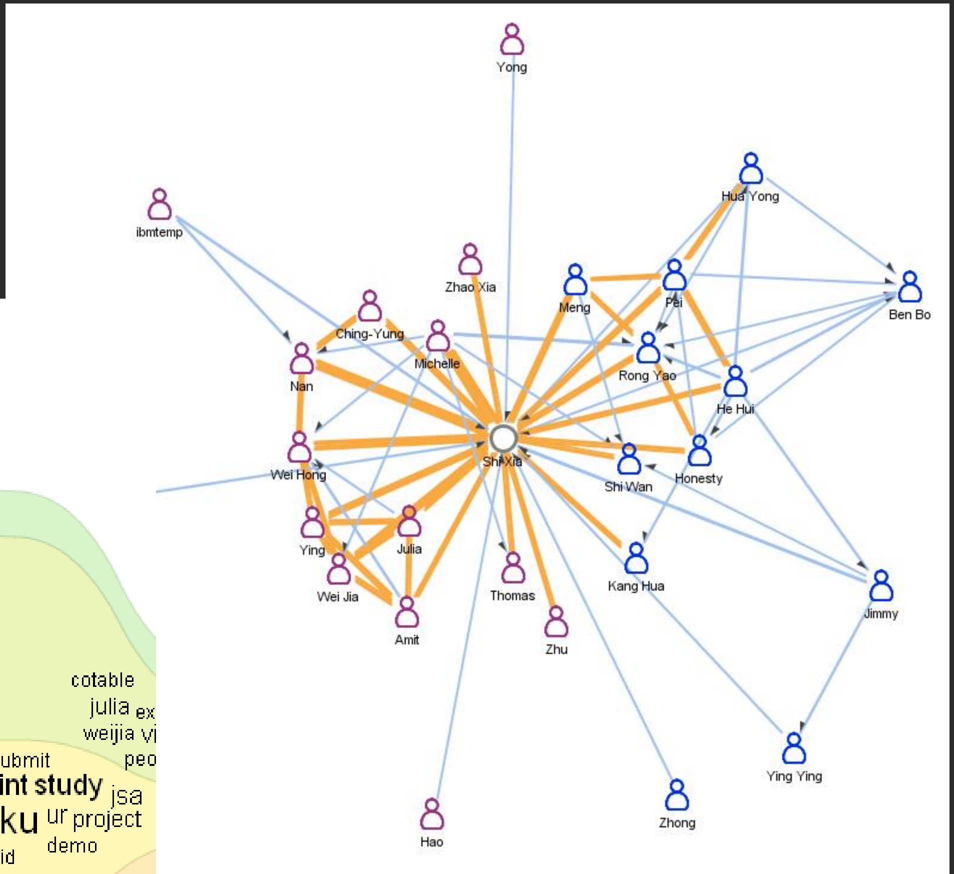


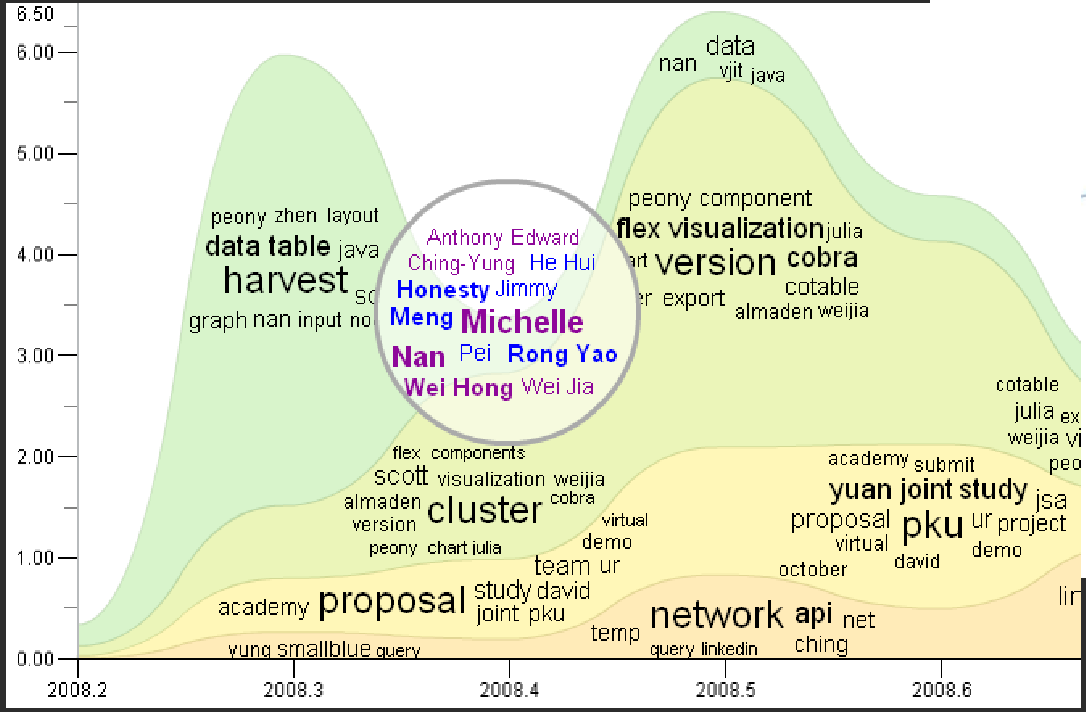
Enhanced Stacked Graph: Layer Labeling (cont'd)



Interacting with Visual Summary

“people” involved and their relationships

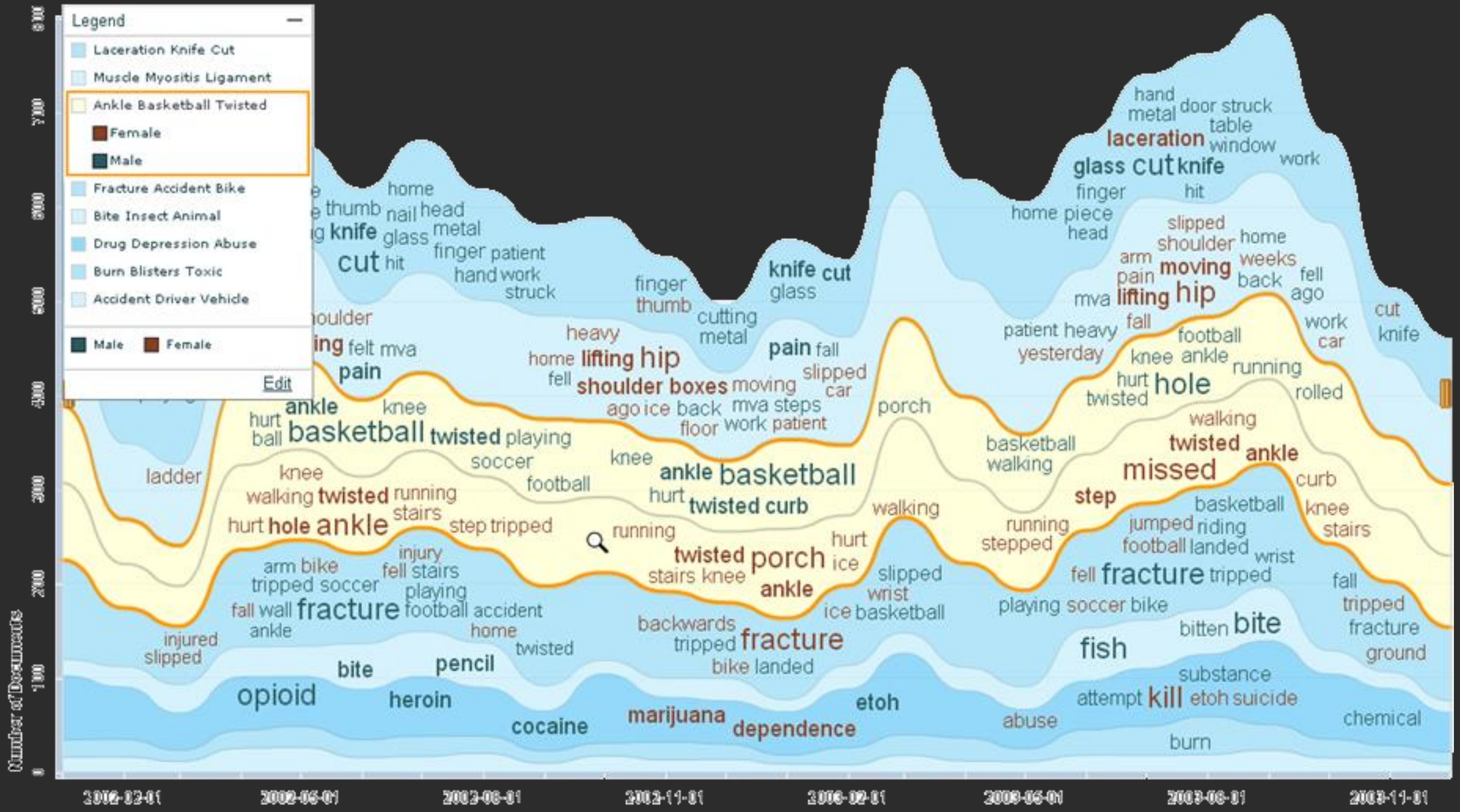




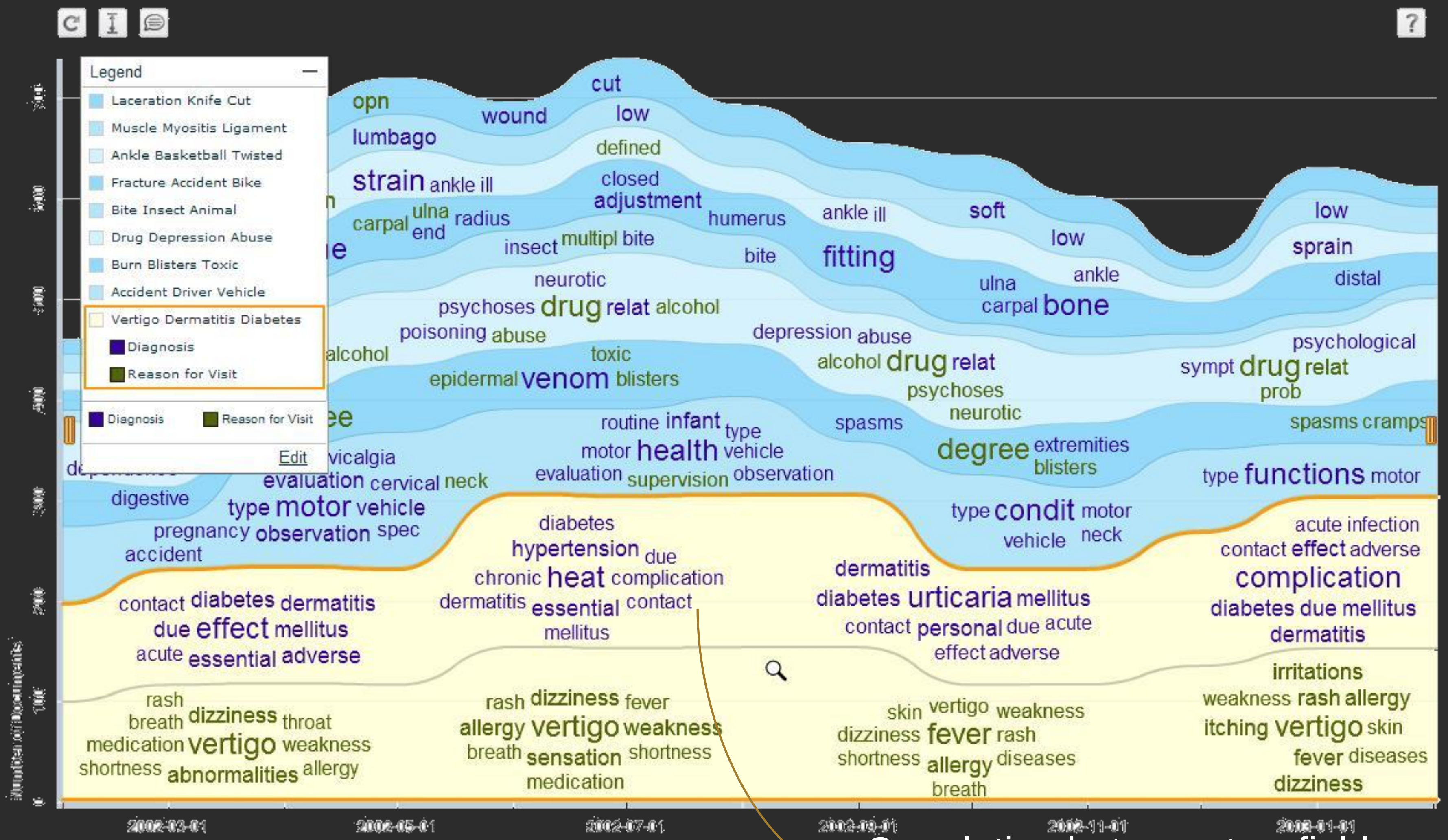
Application Example: Healthcare

- Visualize text to facilitate analysis
  - Cause of injury
  - Reason for visit
  - Diagnosis
- Multiple fields of text data and their correlation
- Leverage structured data to help better illustrate text information
  - Gender + Cause of injury

Correlation between Structured and Text Fields



Correlation between Text Fields



Correlation between two fields, *diagnosis* and *reason for visit*

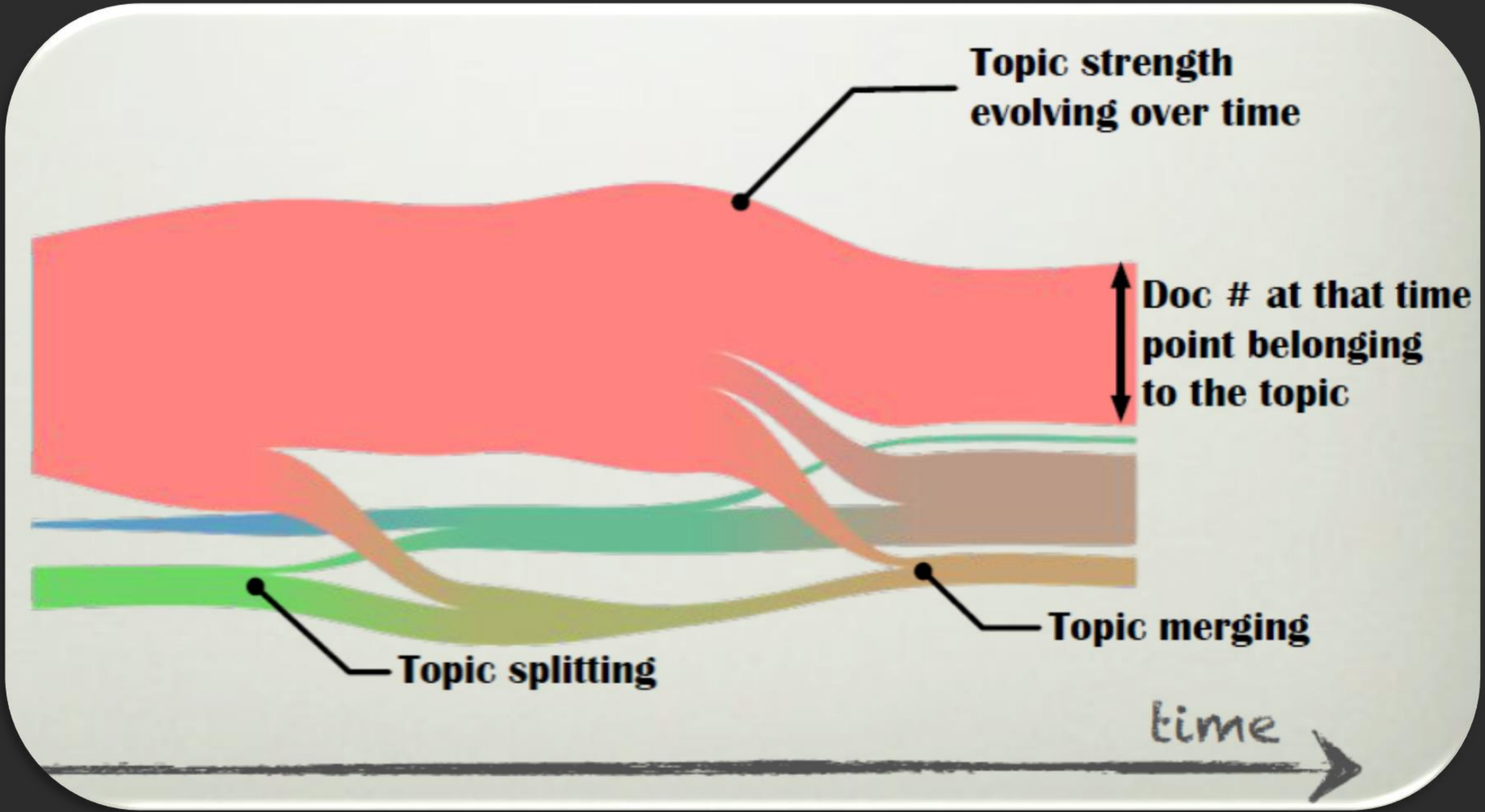
TextFlow - Problems

Understanding topic evolution in large text collections is important

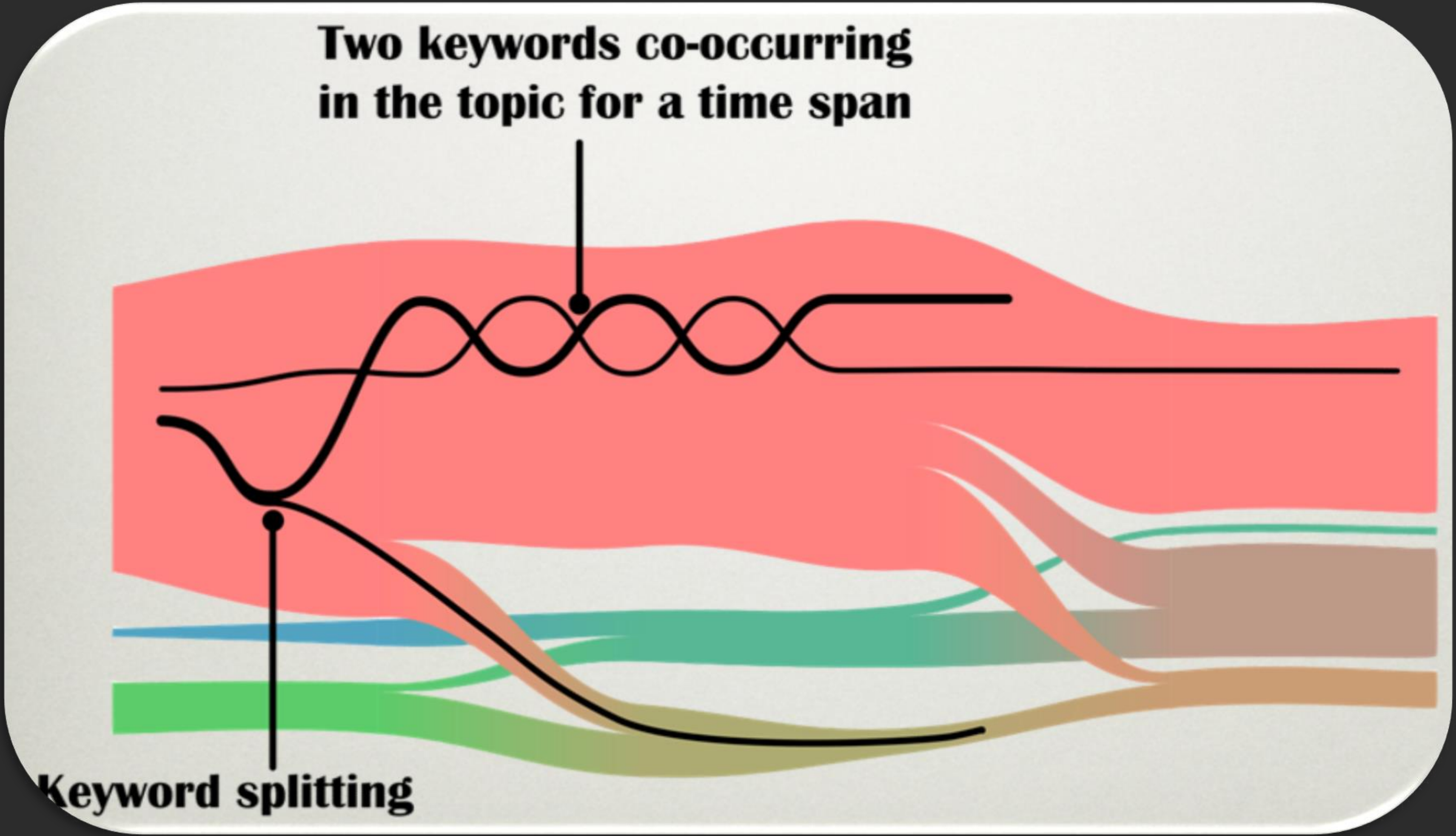
– **Keep abreast** of hot, new, and intertwining topics

– **Gain insight** into the latent topics

Application Example



Application Example

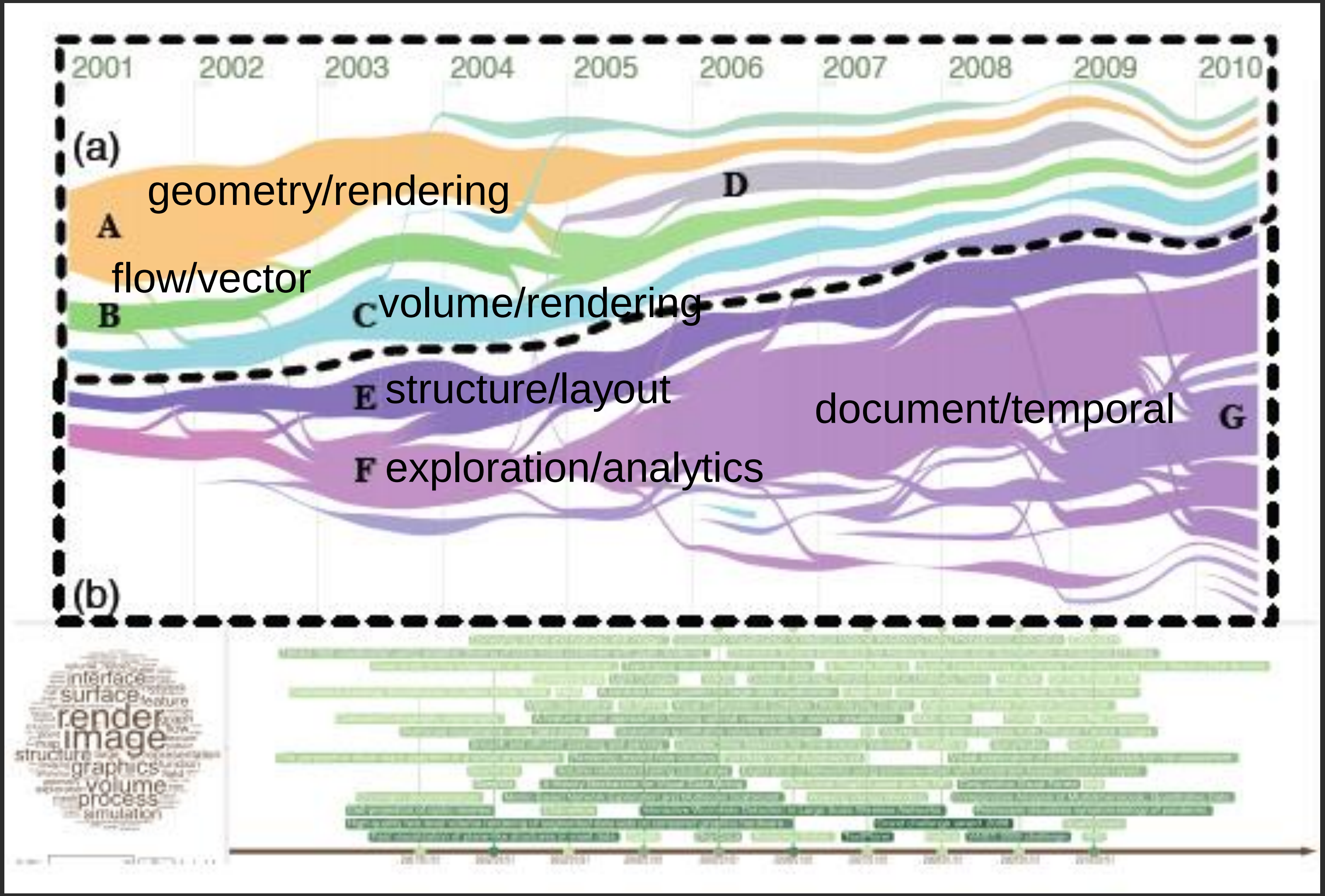


Microsoft Research 微软亚洲研究院

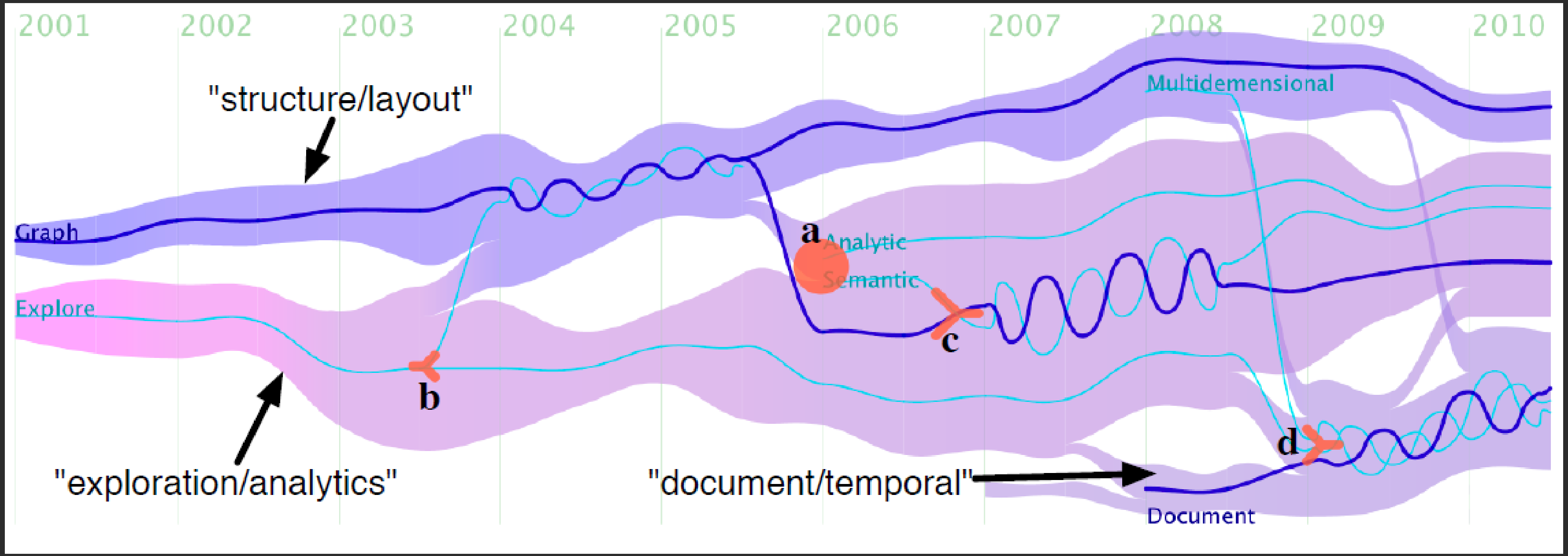
Demo

TextFlow: Towards Better Understanding of Evolving Topics in Text

Example – VisWeek Publications



Example – VisWeek Publications



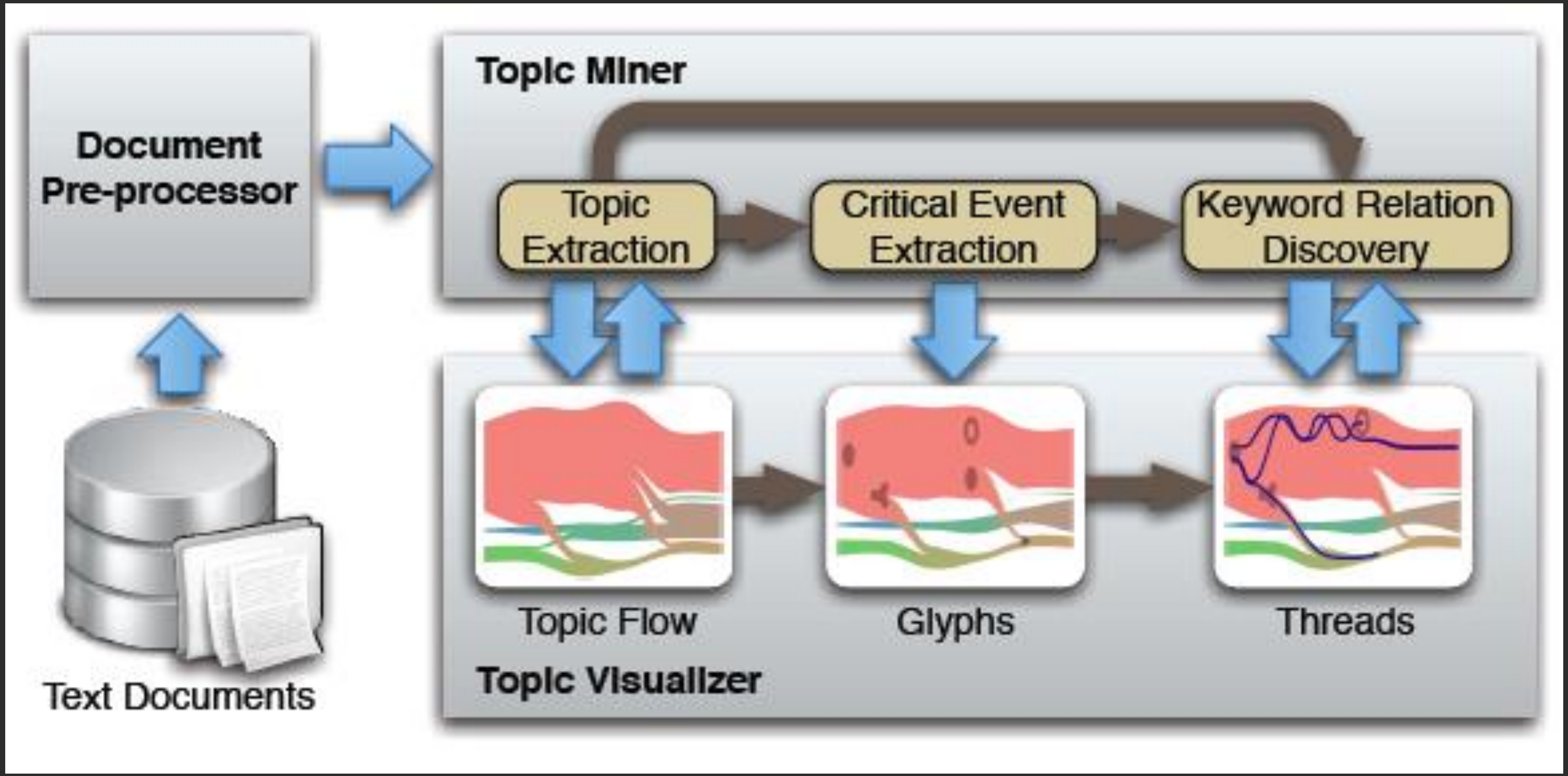
Challenges

- **Model**
  - Topic merging/splitting patterns
- **Visually convey**
  - Topic merging/splitting patterns in an intuitive way
- **Facilitate**
  - Analytical reasoning

Our Solution

- Leverage *hierarchical Dirichlet processes*
  - To model topic merging/splitting
- Augment familiar visual metaphors (rivers)
  - To convey the complex analytic results
- Support interactions at different levels
  - From global structure to local salient features

TextFlow Overview



Topic Data and Relationship Extraction

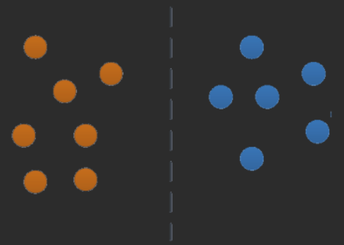
- **Incremental Hierarchical Dirichlet Processes**
  - Online learning of the topics in text
  - Automatically detect the topic numbers
  - Extract the merge/splitting relationships
    - Based on document topic change
    - Online compute the merging/splitting probabilities

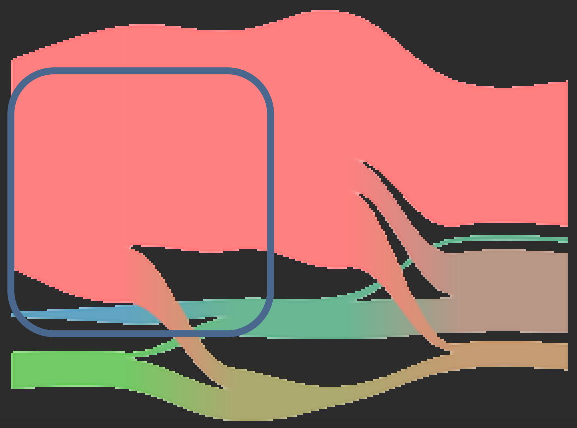
Splitting Relationship

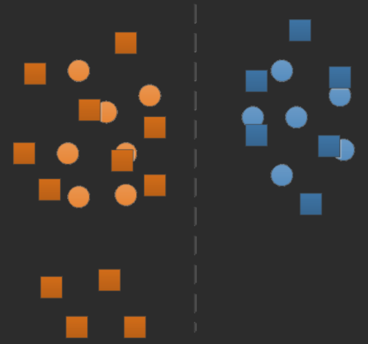
Time t

Time t+1

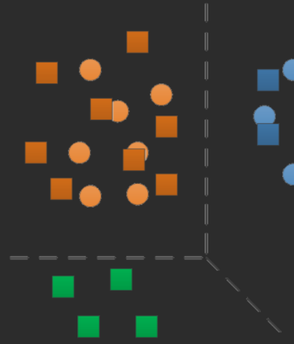
Predict







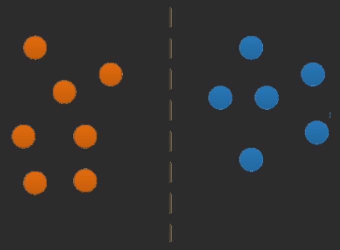
Refine



$$P^{out}_{t-1}(s \to r) \triangleq \frac{\sum_{\tau=t-T_{win}+1}^{t} \sum_{j,i} I(z_{ji}^{\tau,old} = s \ \& \ z_{ji}^{\tau,new} = r)}{\sum_{\tau=t-T_{win}+1}^{t} \sum_{j,i} I(z_{ji}^{\tau,old} = s)}$$

Merging Relationship

Time t → Time t+1



Predict

Refine

$$P_t^{in}(s \rightarrow r) \triangleq \frac{\sum_{\tau=t-T_{win}+1}^{t} \sum_{i,j} I(z_{ji}^{t,old} = s \,\&\, z_{ji}^{t,new} = r)}{\sum_{\tau=t-T_{win}+1}^{t} \sum_{i=1}^{n^t} I(z_{ji}^{t,new} = r)}$$

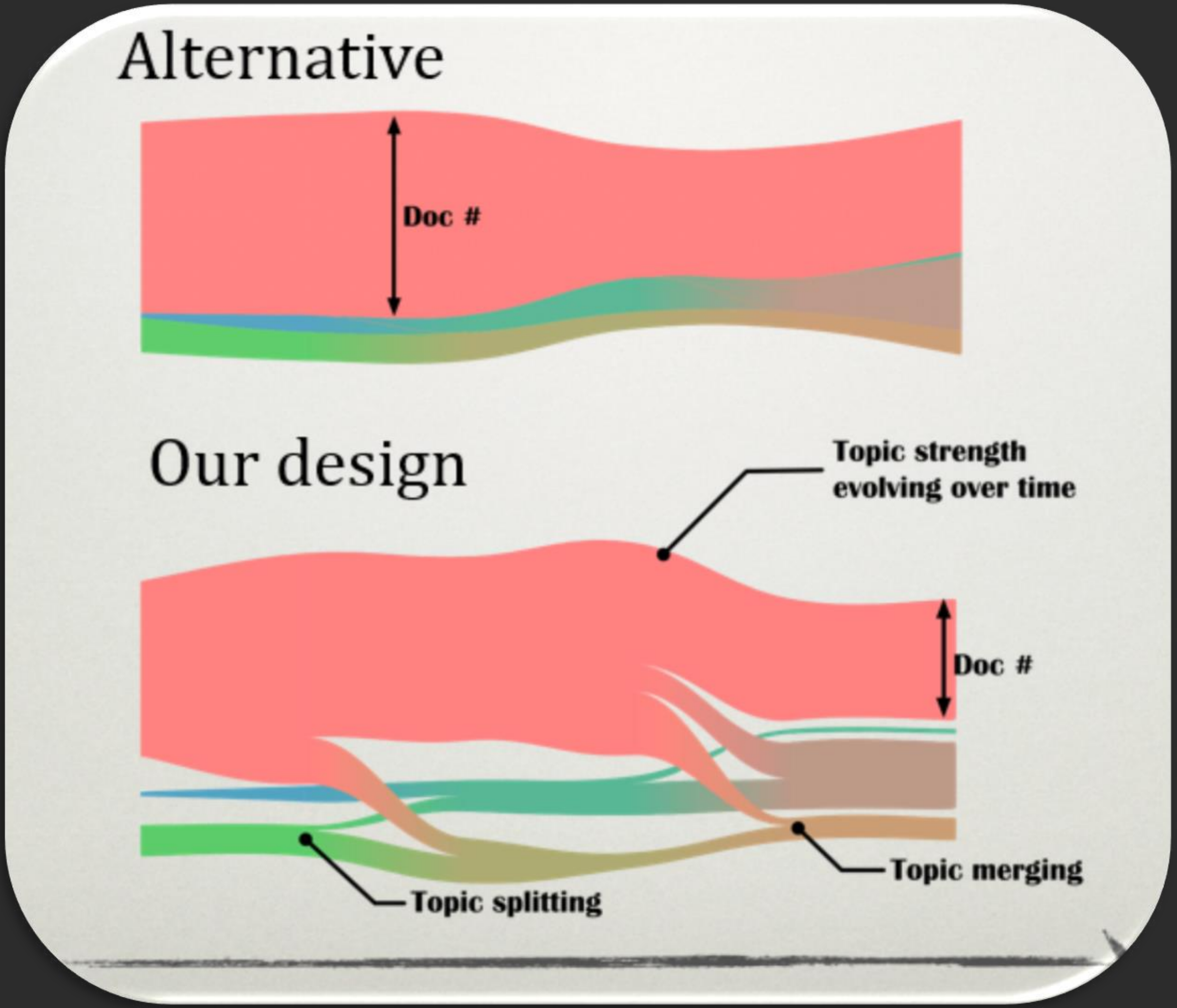
Critical Event Extraction

- **Types of critical events**
  - Birth, death, merge, and split
- **Scoring the merging/splitting event**
  - Number of the branches
  - Entropy of the branching probabilities

Keyword Correlation Discovery

- **Extract**
  - noun phrases, verb phrases, and named entities in each document
- **Count**
  - Co-occurrences among them
- **Be used to illustrate “why”**

Topic Evolution as Flow

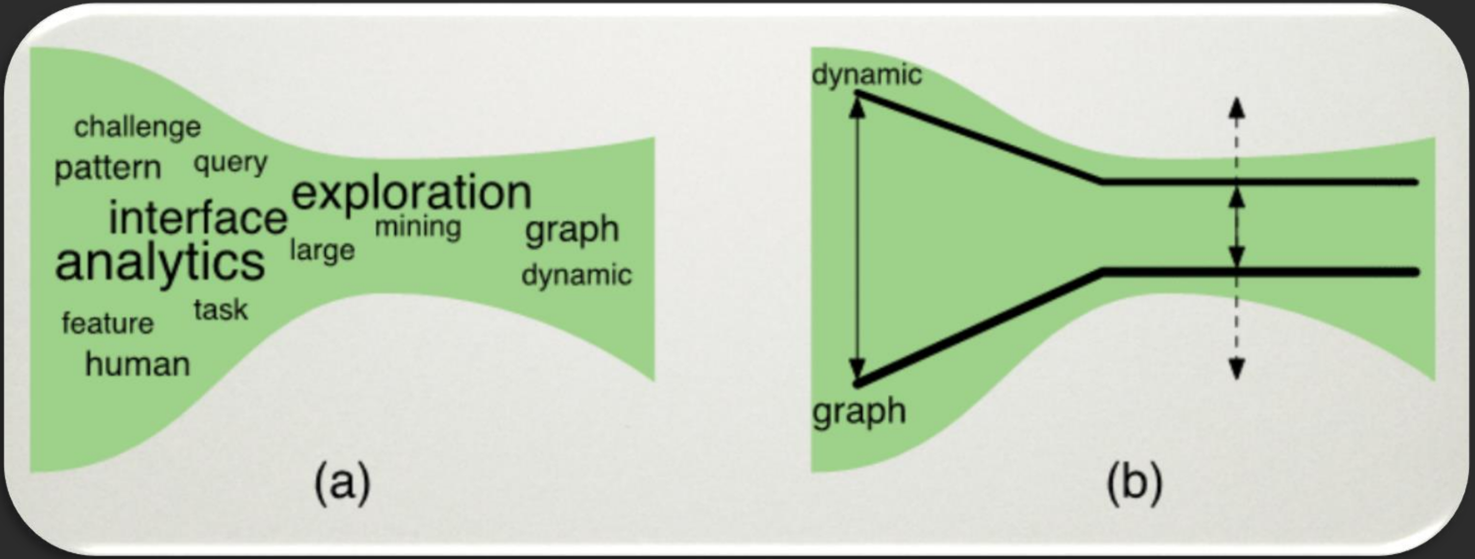


Critical Event as Glyph

- Emerge, dissolve, split, and merge

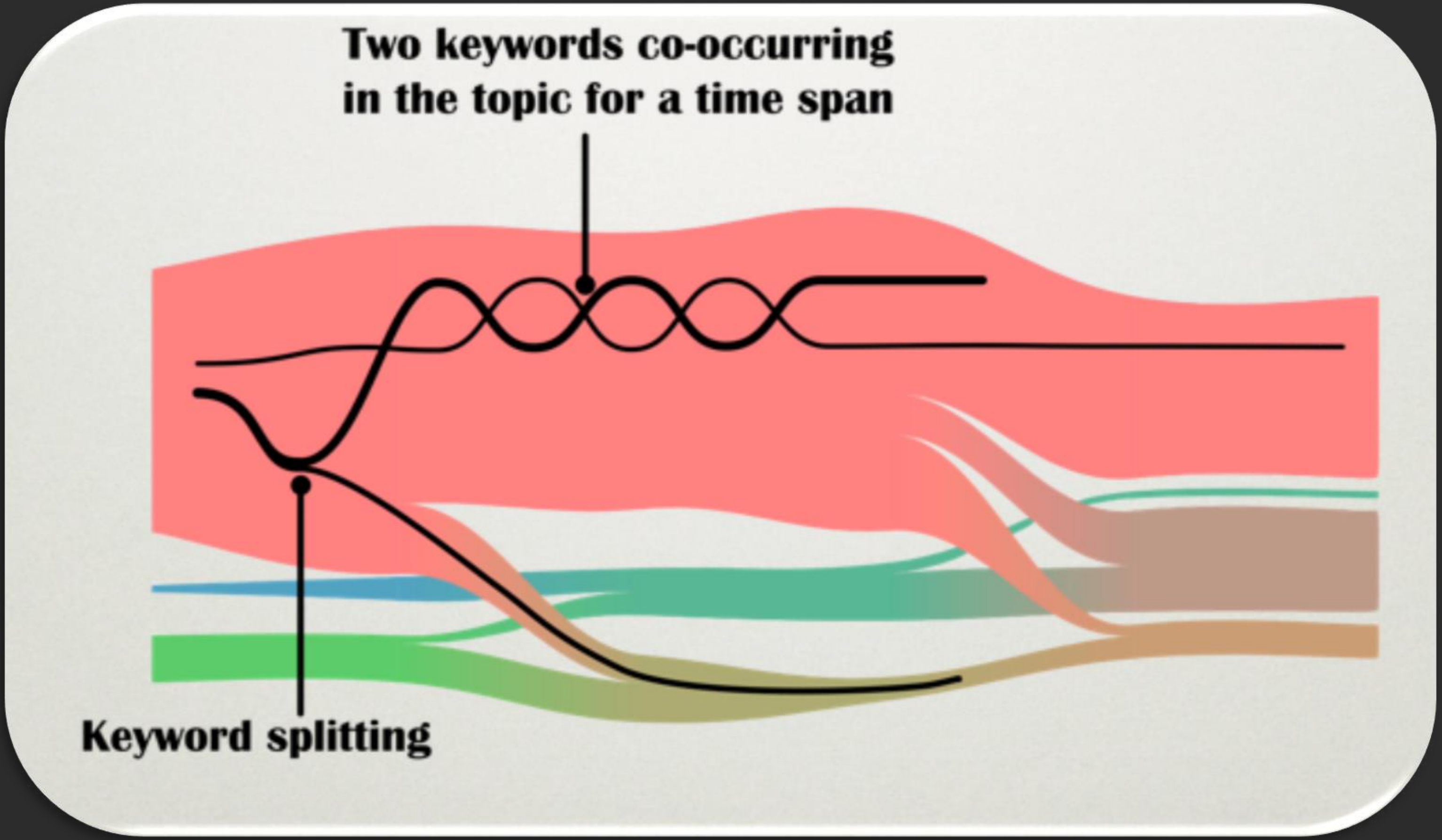
Keyword Correlation as Thread

- Alternatives

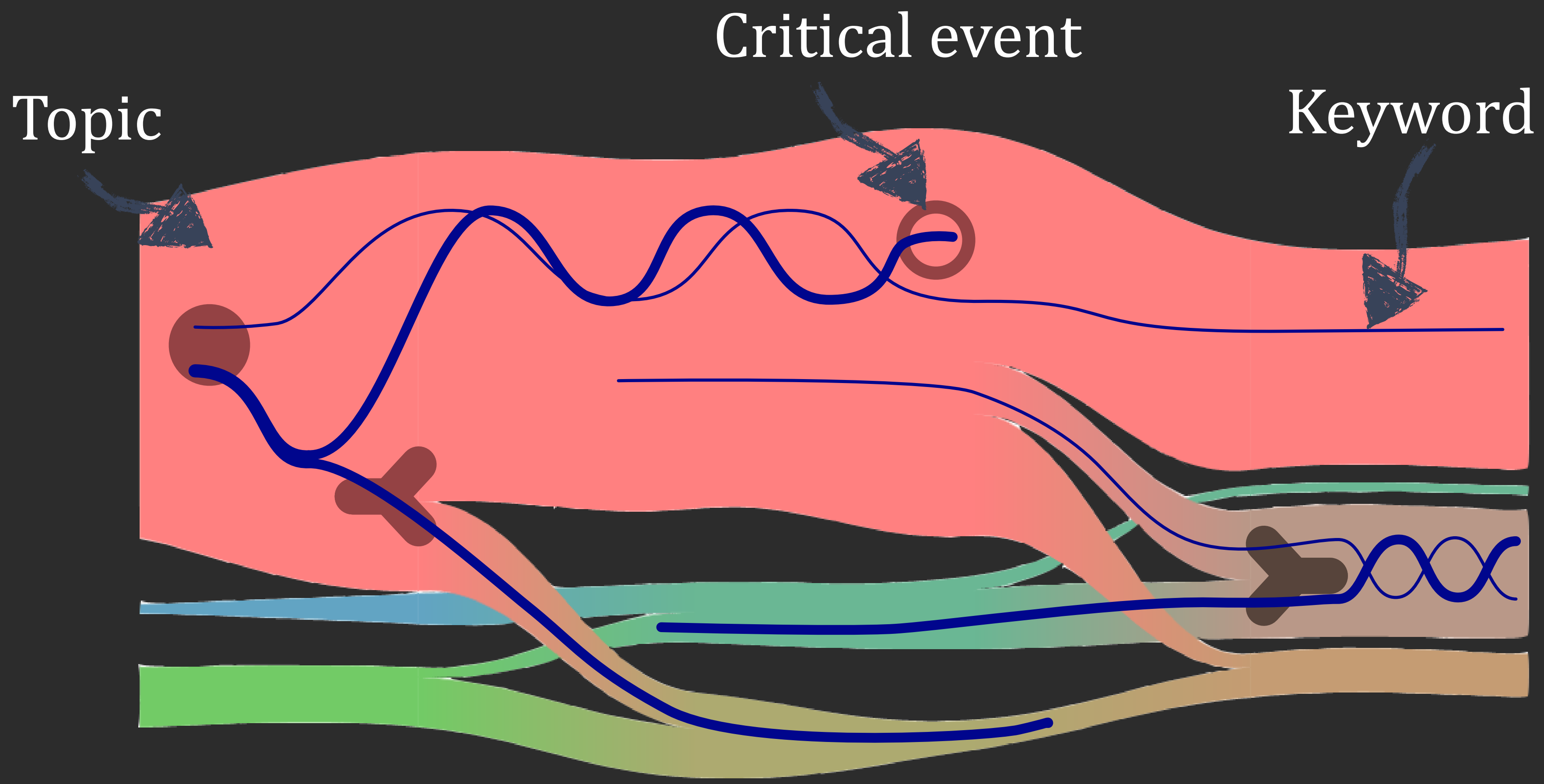


Keyword Correlation as Thread

- Intertwine to indicate co-occurrences

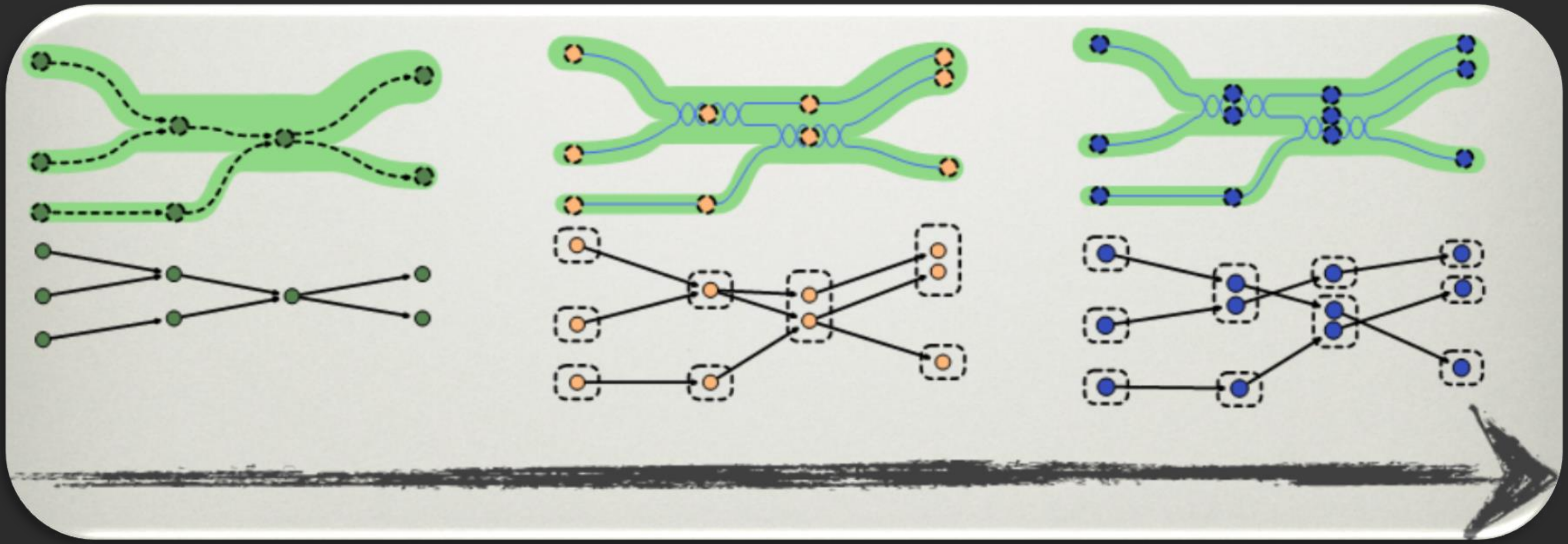


Visualization Design - Consistency



Layout Algorithm

- Three-level directed acyclic graph (DAG)



Adapt to Big Data - Challenges

- Generate evolving multi-branch trees as well as to model their evolution patterns over time
  - Fitness
  - Smoothness
- Handle huge document collection efficiently

Our Solution

- Bayesian online filtering framework
  - Fitness: adopt the fast Bayesian rose tree method (our work in KDD’12)
  - Smoothness: build a constraint tree

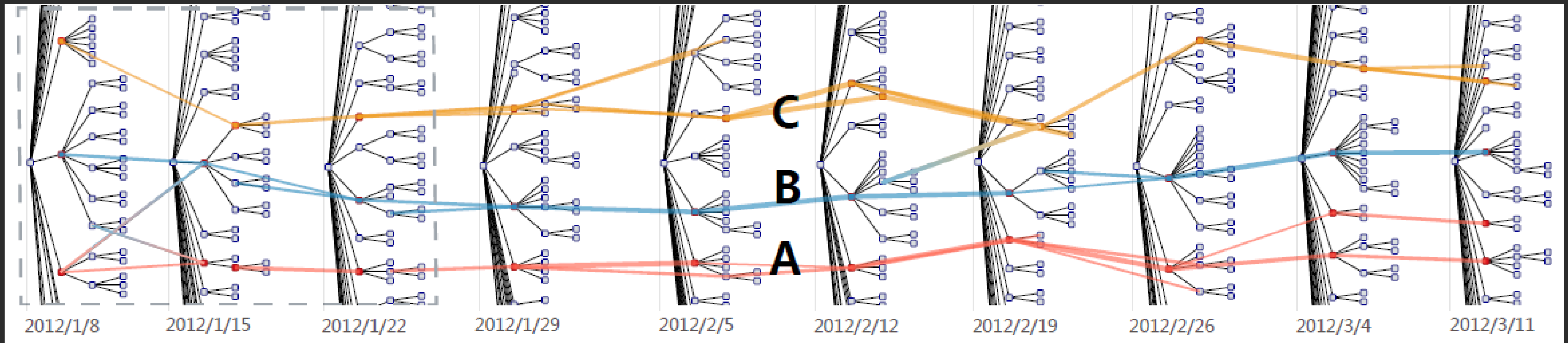
$$p(T^t|\mathcal{D}^t, T^{t-1}) \propto p(\mathcal{D}^t|T^t)\,p(T^t|T^{t-1})$$

- Efficiency: build a constraint tree and treat it as a tree index

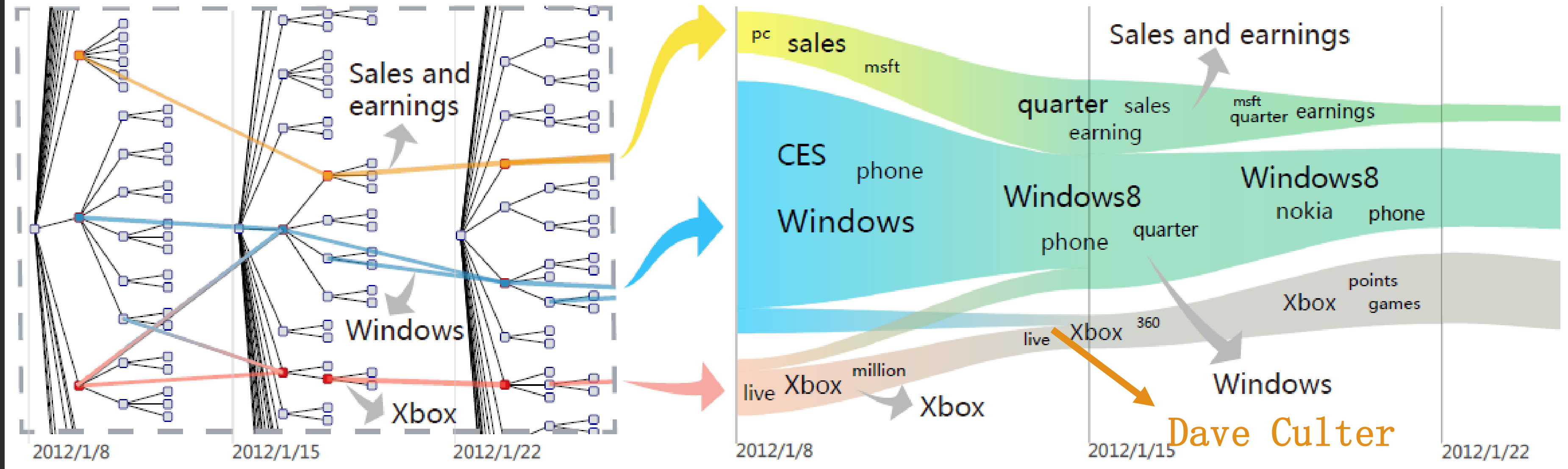
Microsoft Research 微软亚洲研究院

Example

66,528 news articles

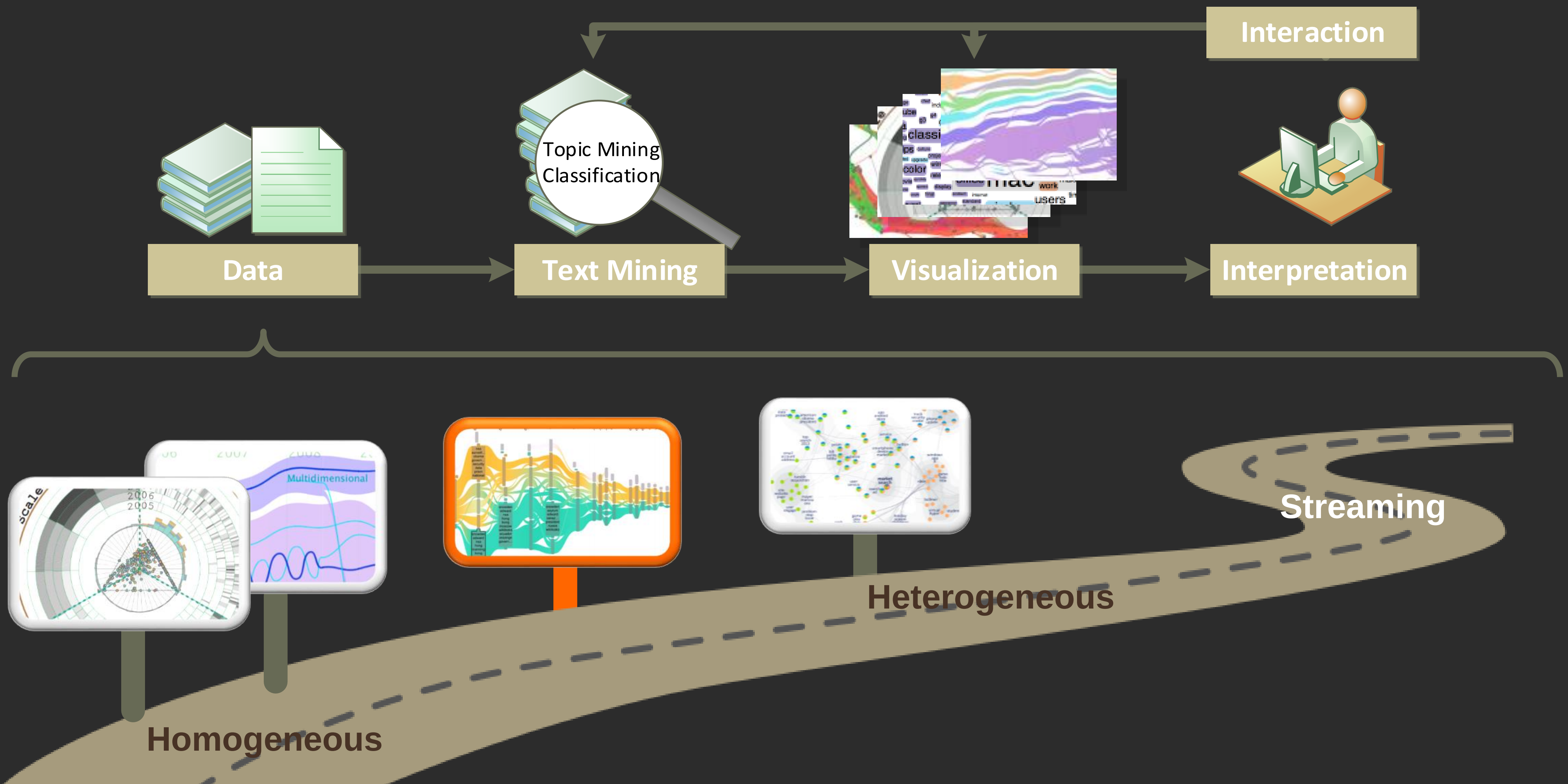


(a) Ten evolving trees from Jan. 8 to Mar. 17 and the highlighted topics “xbox”(A), “windows”(B) and “sales and earnings”(C)



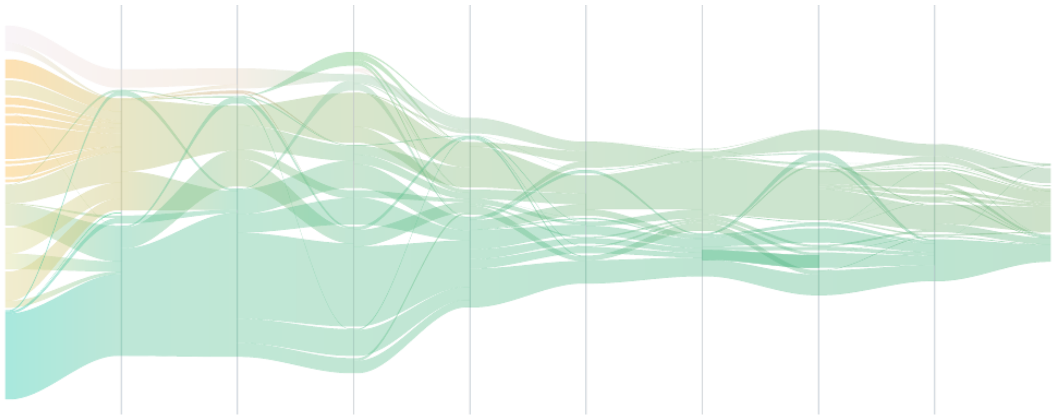
(b) Splitting/merging patterns of topics “xbox,” “windows,” and “sales and earnings” (Jan. 8 to Jan. 28) and the transition from the tree representation to the TextFlow visualization. With this visualization, the evolving splitting/merging relationships are clearly conveyed.

Our Focus: Data

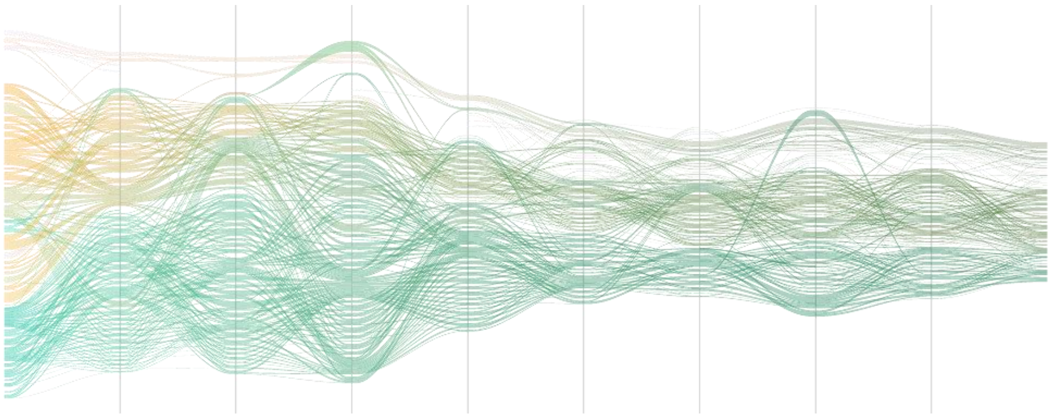


RoseRiver: How Hierarchical Topics Evolve in Large Text Corpora

- Dilemma of topic numbers when handling large collections



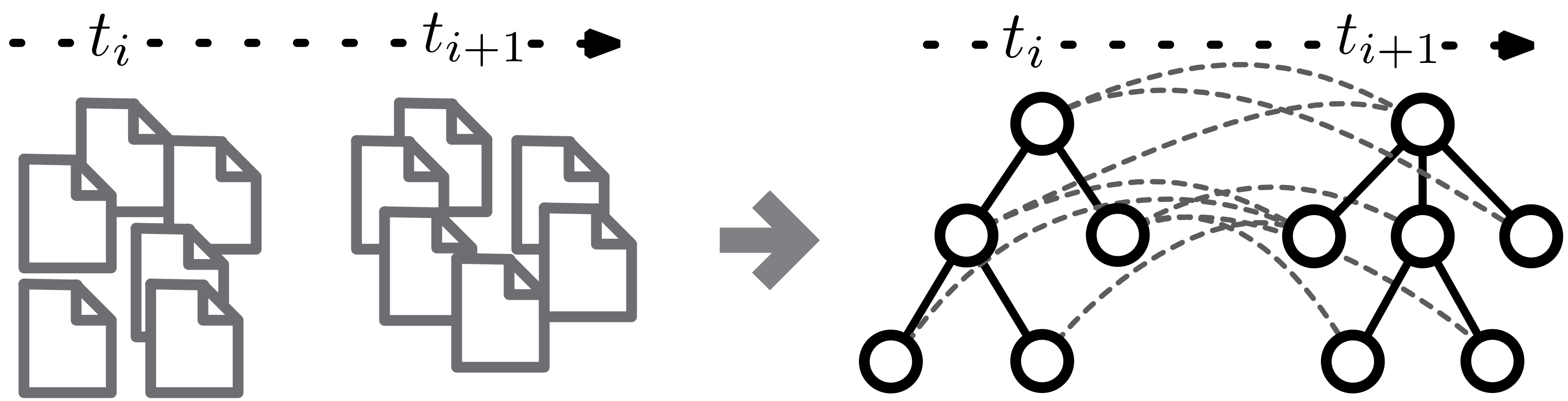
Too general



Too clutter

(X-axis represents time. Each colored stripe represents a topic.)

Our Solution: Evolving Topic Trees



- Interpretable topic results
  – Real-world text corpora are naturally organized as trees
- Intuitive navigation
  – From a global overview to local details

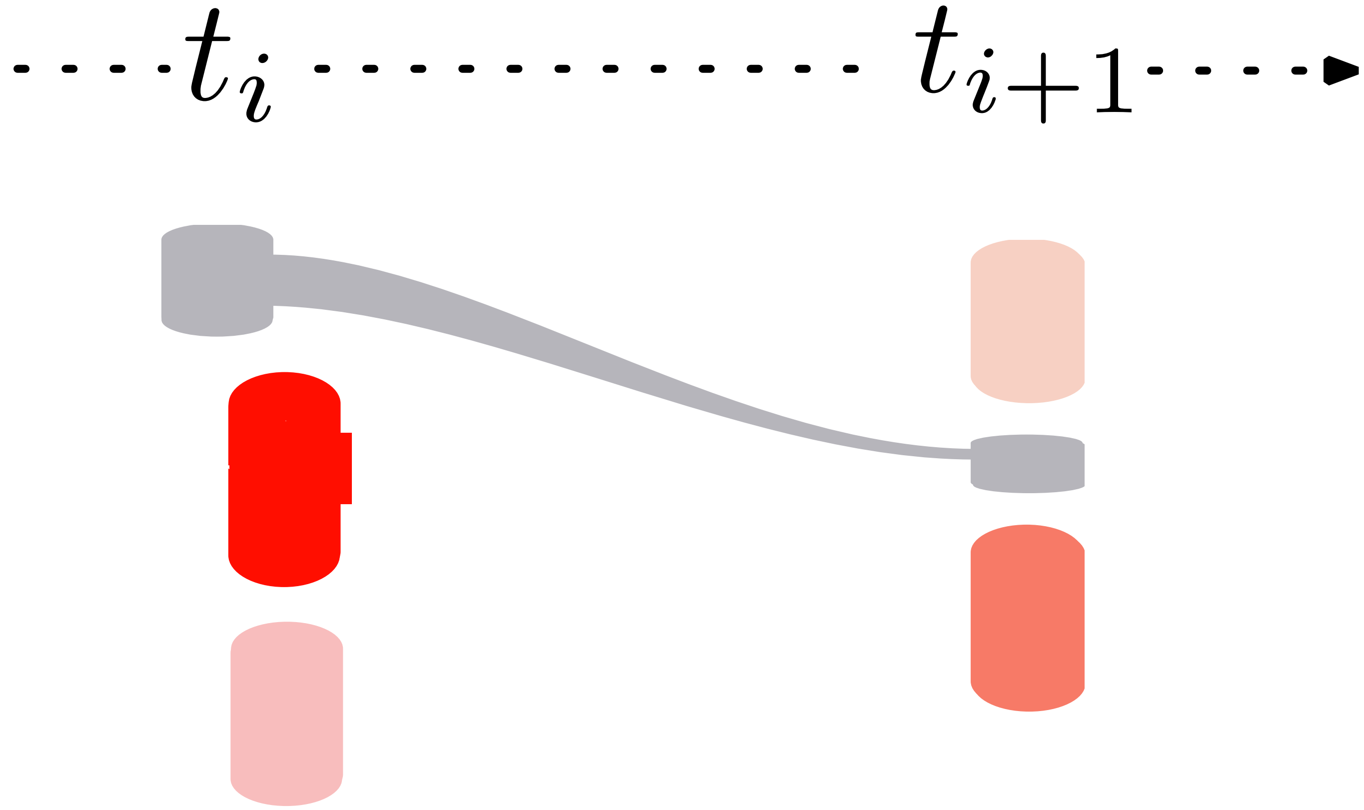
Visualization Challenges

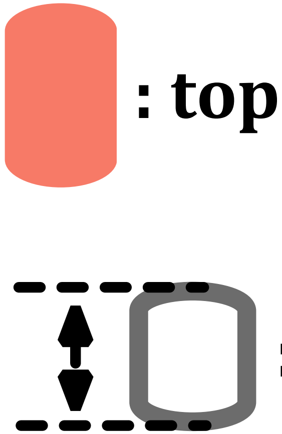
- Effective representation
  - Topic nodes
  - Topic relationships
- Progressive navigation
  - Obtain a full picture
  - Quickly focus on the information of interest
  - Preserve the mental-map

Case Study: Prism Scandal

- 69,867 news articles and 568,225 tweets
- June 3, 2013 to February 9, 2014
- Grouped into 36 topic trees by week
  - Tree depth: 3 – 9
  - Total nodes: 19 – 165
  - Nodes at first level: 3 – 20

Visual Encoding





: topic in the tree

: topic size

:similar document pairs

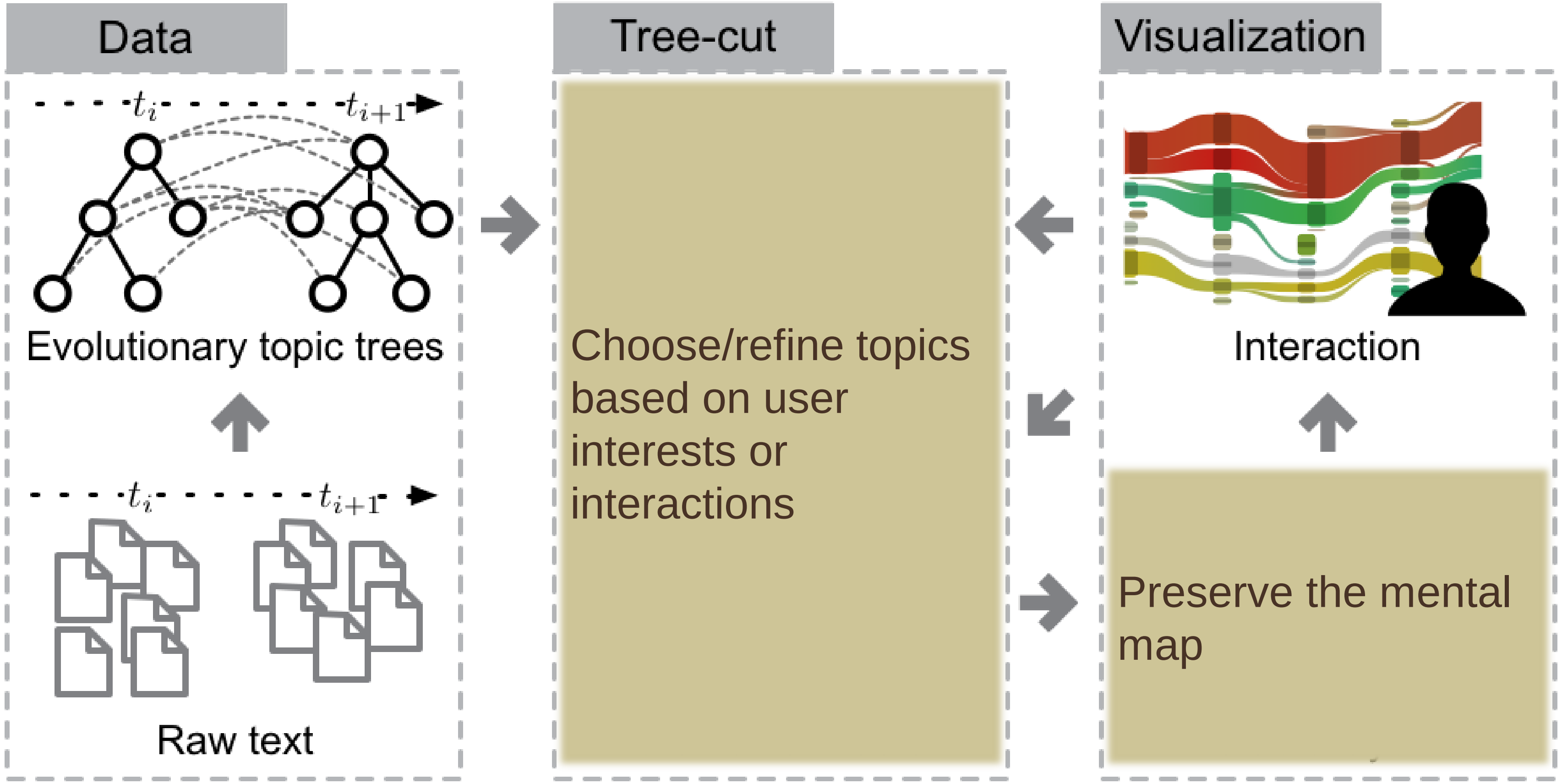
: relevance

RoseRiver

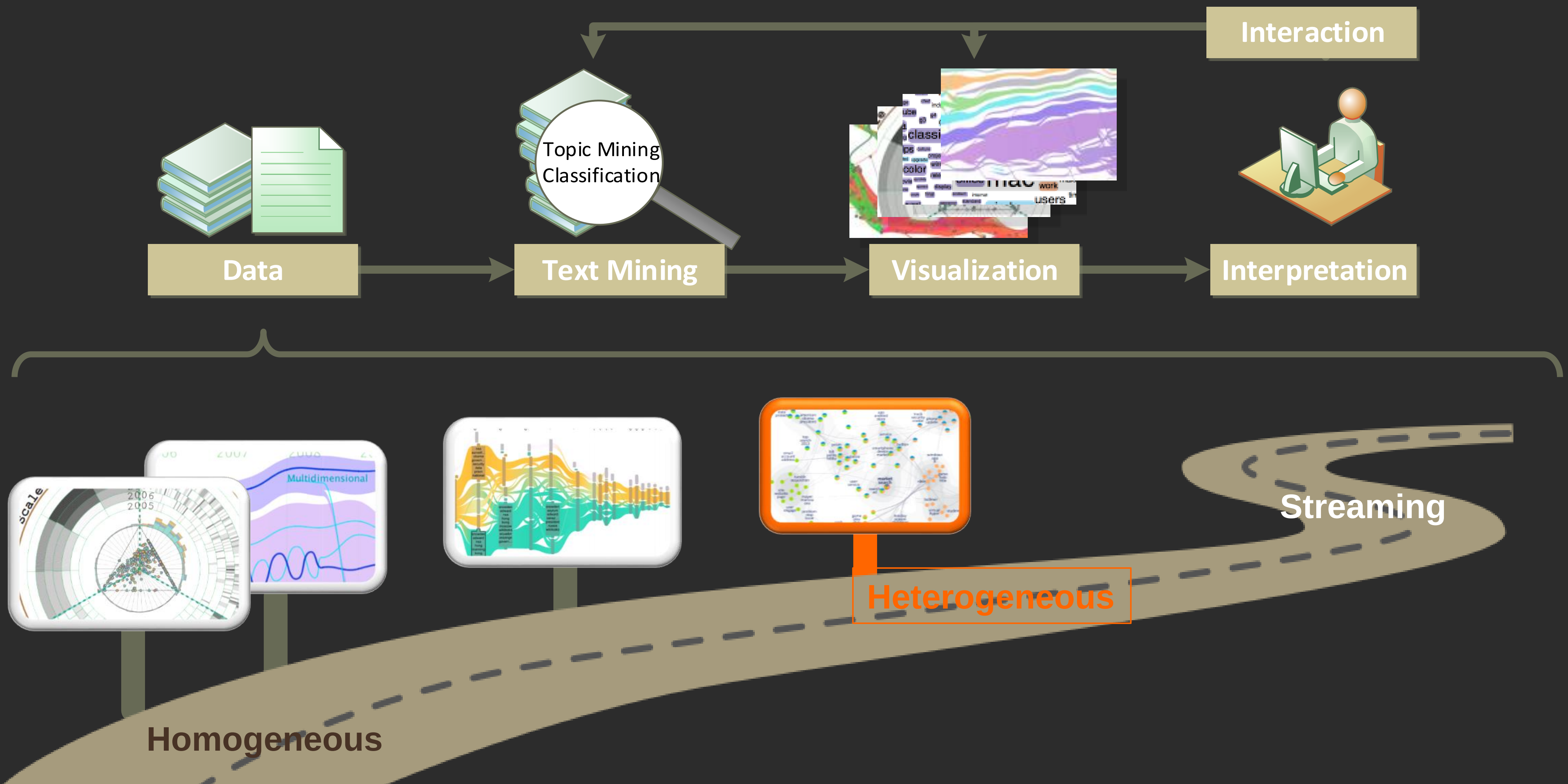
Case study on Prism data

69,867 news articles + 568,225 tweets
(from Jun. 5, 2013 to Feb. 9, 2014)

System Overview



Our Focus: Data



Motivation

- Relevant topics is heavily discussed in multiple sources
- Checking one source may take a part for the whole
- Gather separate pieces of information about these topics scattered in different sources and reconstruct the full picture

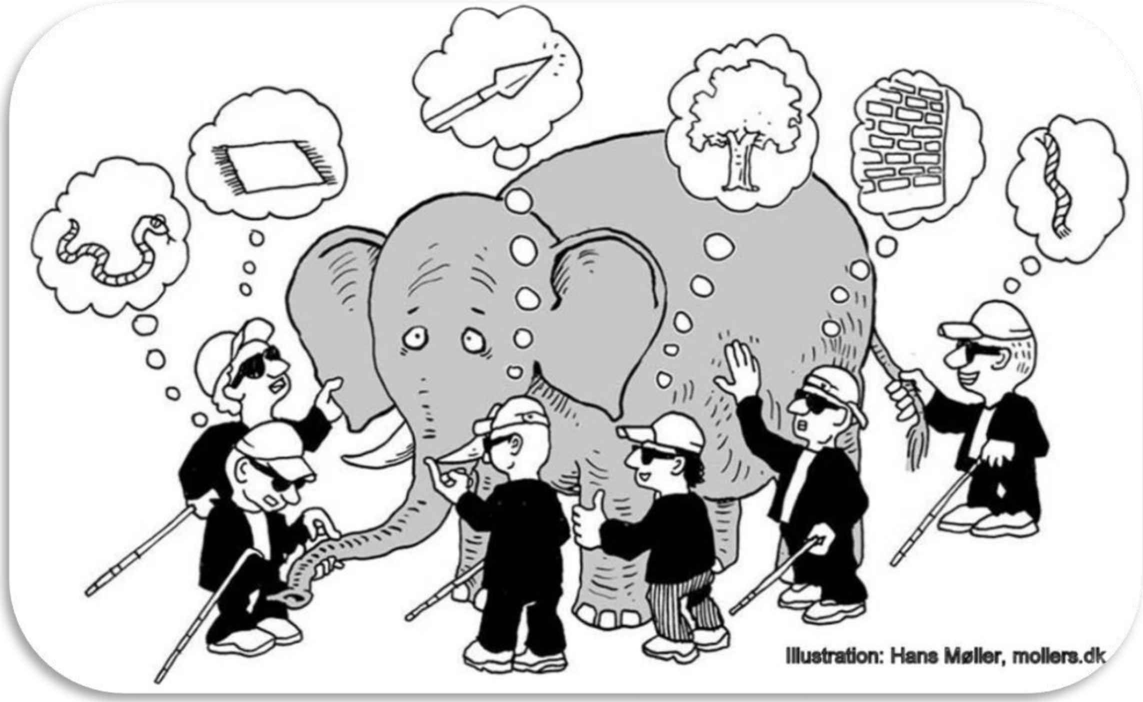
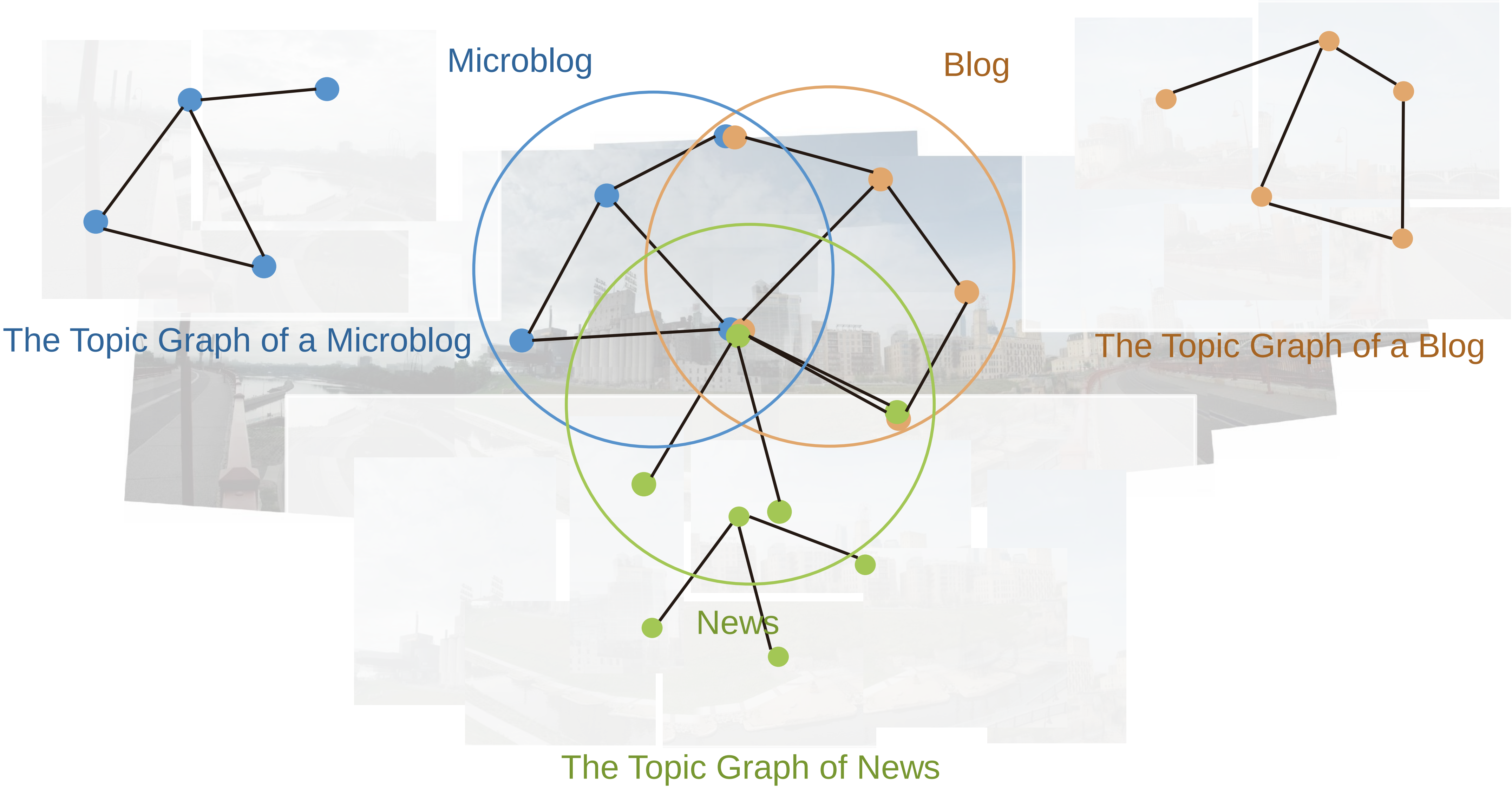
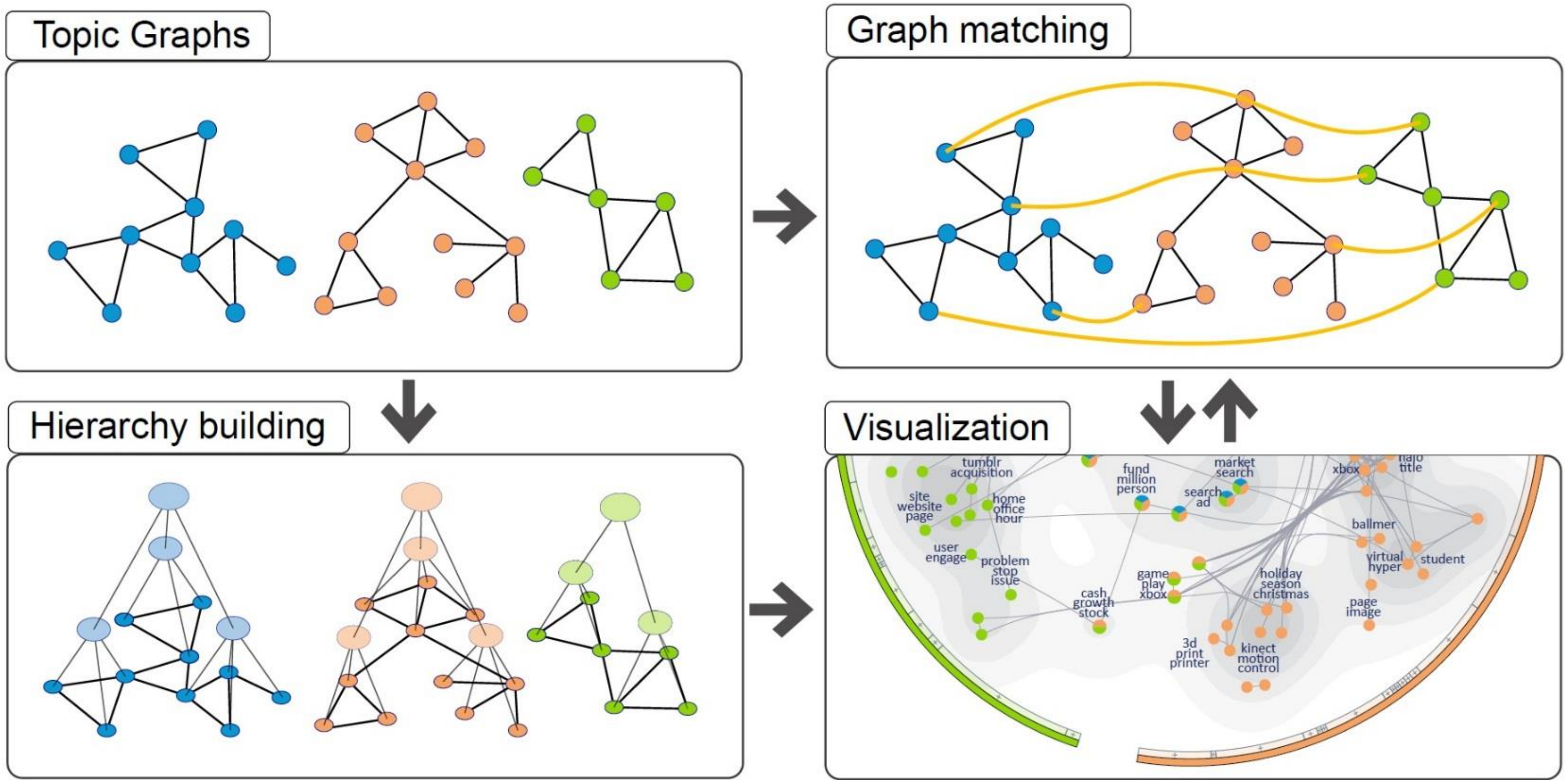


Image Panorama

TopicPanorama



TopicPanorama Overview



Graph Matching: Why

- A straightforward method
  - merged data + a topic graph construction method
- Deficiency
  - may fail to model the diversity across different corpora

Graph Matching: Our Solution

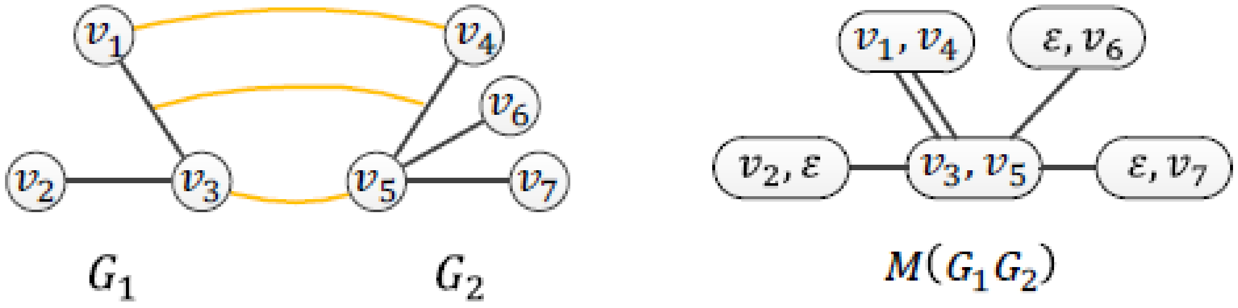
- Objective
  - Finding correspondence among multiple topic graphs
- Base
  - Graph-edit-distance-based matching method
- Minimize the cost of all pairwise graph matchings
  - s.t. all node mapping relationships are transitive

$$d(G_1, G_2, \ldots, G_N) = \min c(f_{G_1 G_2 \ldots G_N}), \quad c(f_{G_1 G_2 \ldots G_N}) = \sum_{i=1}^{N} \sum_{j=i+1}^{N} c(f_{G_i G_j})$$

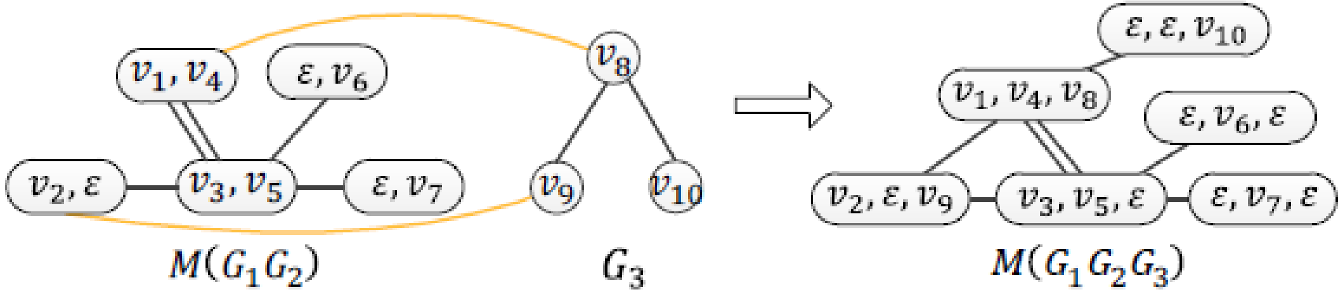
$$s.t.\ v_l \mapsto v_m, v_m \mapsto v_n \Rightarrow v_l \mapsto v_n$$

$$\forall G_i, G_j, G_k \in \{G_1, G_2, \ldots, G_N\}, \forall v_l \in \mathcal{V}_i, \forall v_m \in \mathcal{V}_j, \forall v_n \in \mathcal{V}_k,$$

Graph Matching: Metagraph



(a) $f_{G_1G_2}$ and its corresponding metagraph



(b) Derive the meta-graph $M(G_1G_2G_3)$ from $M(G_1G_2)$.

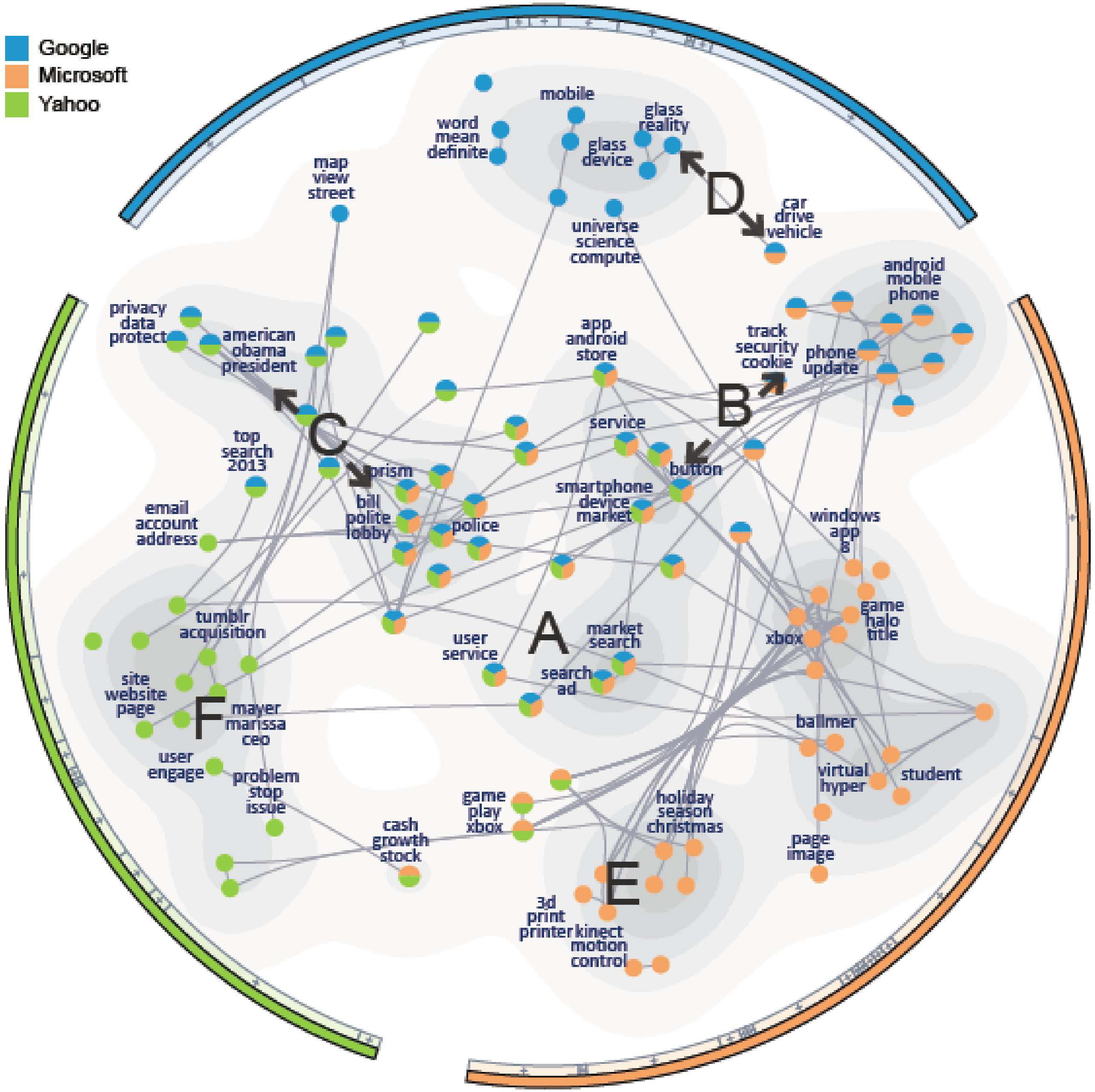
Incremental Algorithm

- Allow users to interactively modify the graph matching result

- Develop an incremental graph matching algorithm based on the incremental Hungarian algorithm

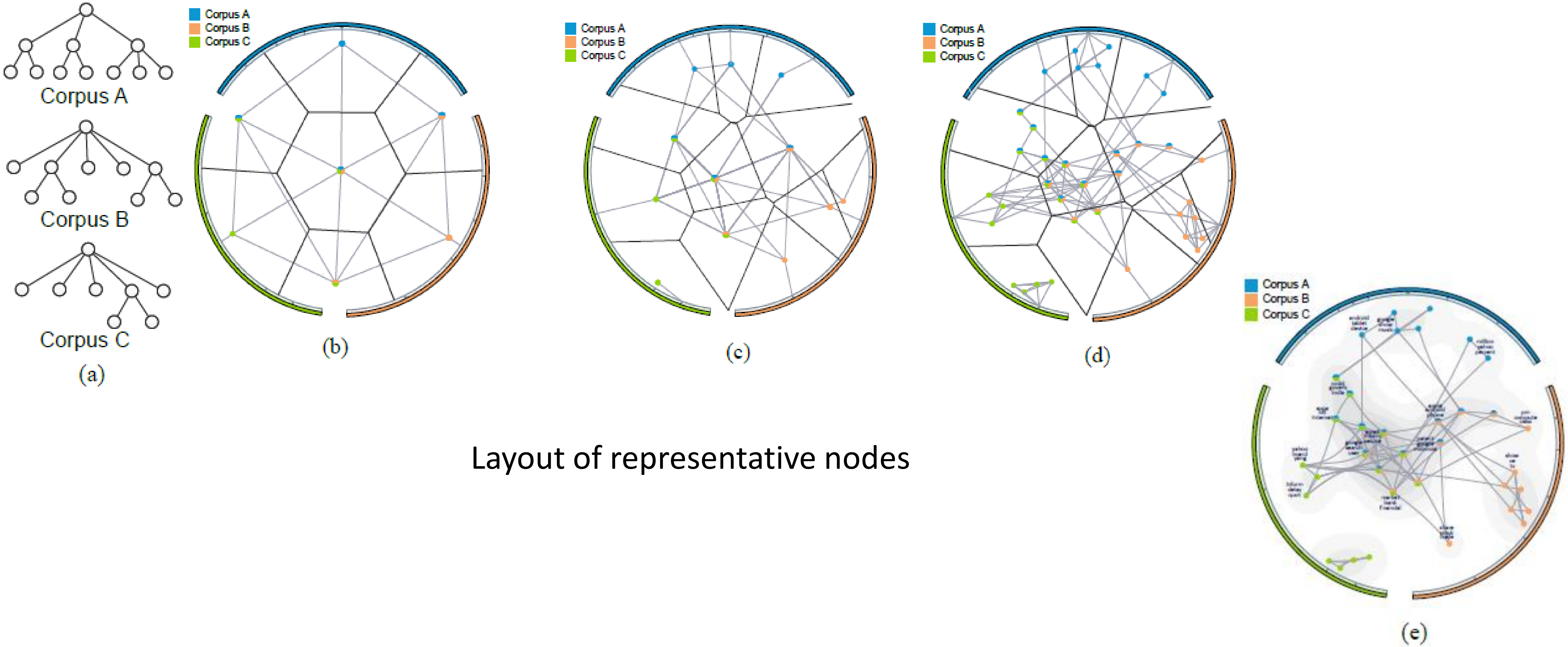
Visualization

- Graph matching as density-based graph visualization
- Topic hierarchy as stacked tree
- Coupling graph visualization with stacked tree



Layout Algorithm

Basic principle: the common parts-> the layout area of each area; the distinctive parts -> the corpus area



Layout of representative nodes

TopicPanorama: a Full Picture of Relevant Topics

刘世霞[1]　王希廷[1,2]　陈键飞[2]　朱军[2]　郭百宁[1]

[1] 微软亚洲研究院

[2] 清华大学

Microsoft Research 微软亚洲研究院

Visual User Behavior Analytics: Key Focuses

Multifaceted Behavior Analytics	Design advanced techniques to explore multifaceted behaviors and their relationships
Implicit Behavior Analytics	Design model-driven, interactive visual analytics methods to understand implicit behaviors
Explicit Behavior Analytics	Design effective, fundamental techniques and methodologies to visualize explicit behaviors

Microsoft Research 微软亚洲研究院

"Big Picture": Our Key Focuses

Multifaceted Behavior Analytics

Implicit Behavior Analytics

Explicit Behavior Analytics

Step 3

Step 2

Step 1

“Big Picture”: Our Key Focuses

Multifaceted Behavior Analytics — Step 3

Implicit Behavior Analytics — Step 2

Explicit Behavior Analytics — Step 1

Microsoft Research 微软亚洲研究院

"Big Picture": Our Key Focuses

Multifaceted Behavior Analytics

Step 3

Implicit Behavior Analytics

Step 2

Explicit Behavior Analytics

Step 1

“Big Picture”: Our Key Focuses

Multifaceted Behavior Analytics

Step 3

Implicit Behavior Analytics

Step 2

Explicit Behavior Analytics

Step 1

StoryFlow

LoyalTracker

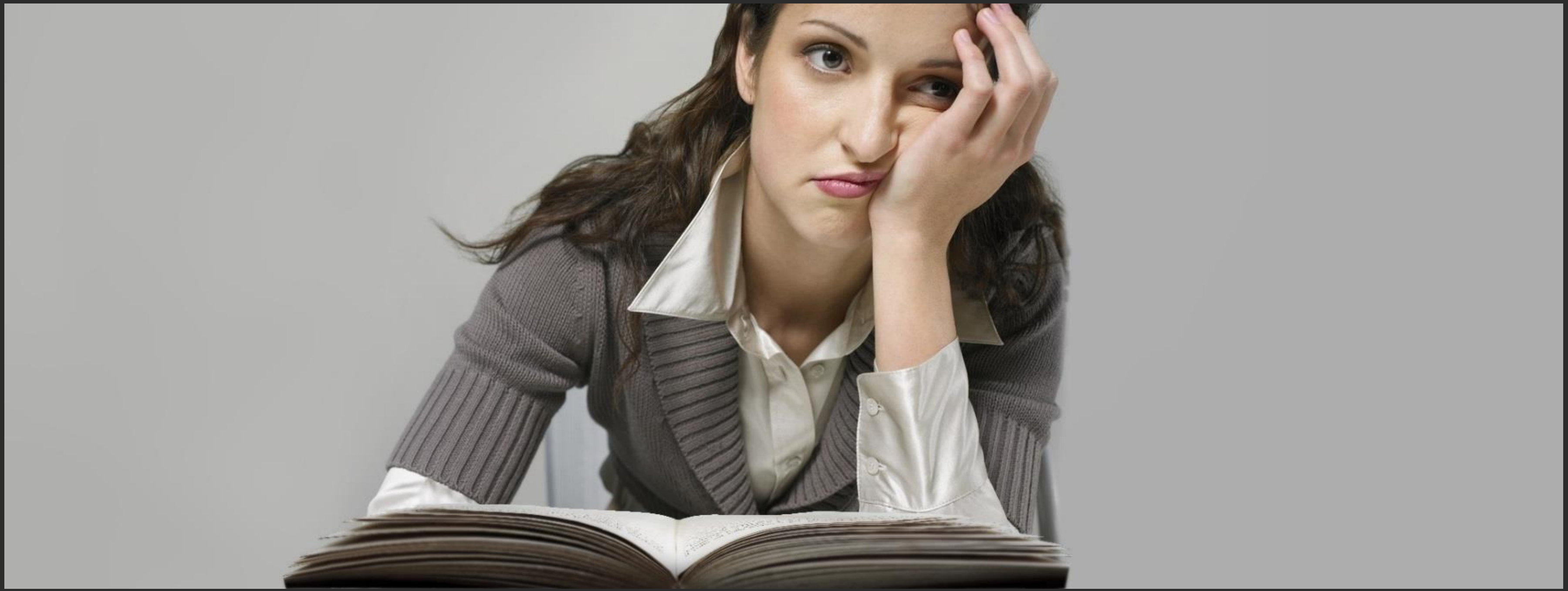
Storytelling

A Good Storyteller

Who, When, and Where

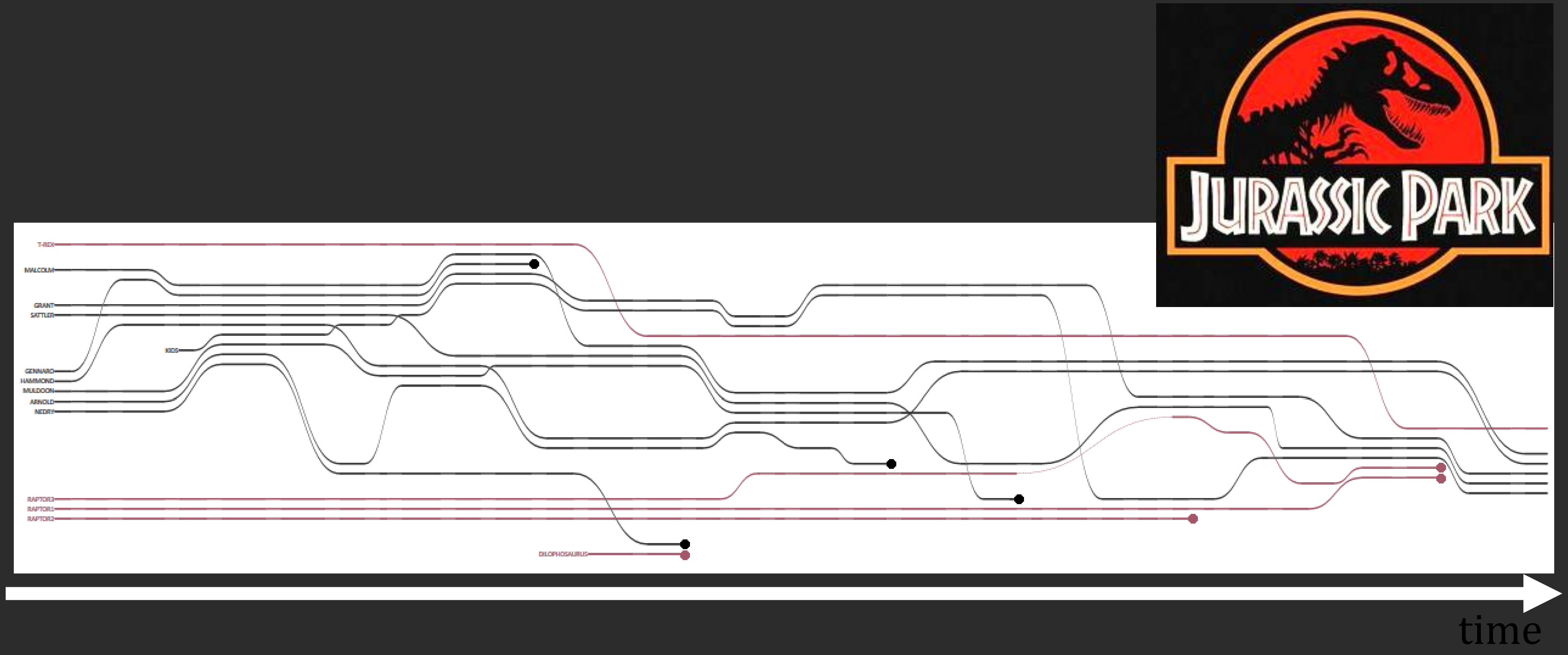
Stories Are Complicated

- The dynamic relationships of characters
- The complex structure of scenes



Microsoft Research 微软亚洲研究院

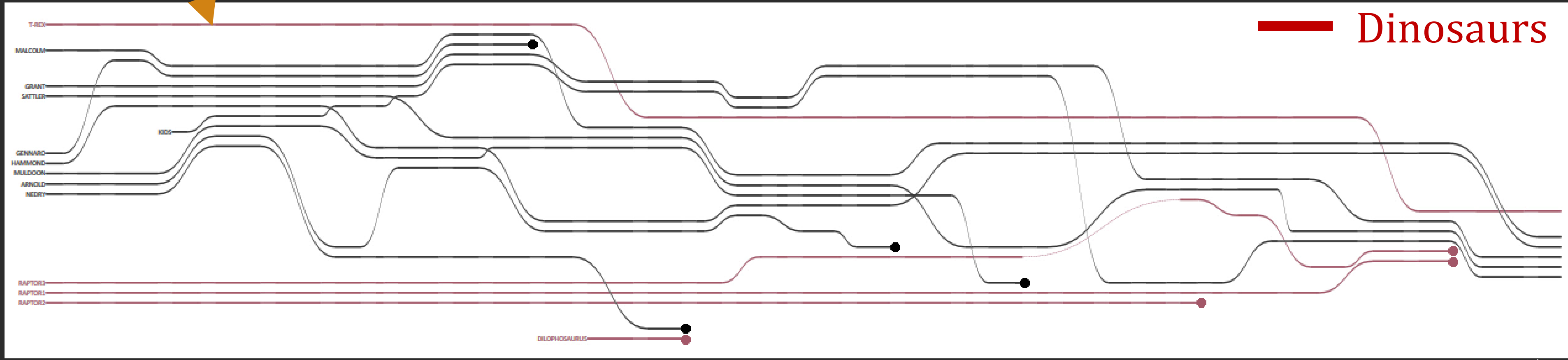
Storyline Visualization



Storyline Visualization

One character

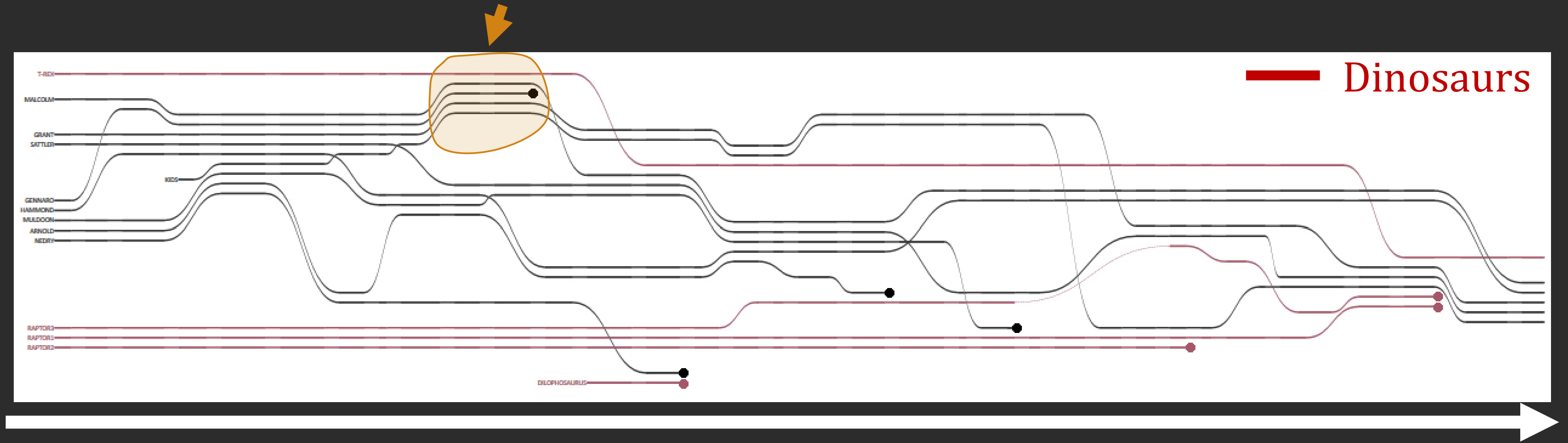
T-Rex



time

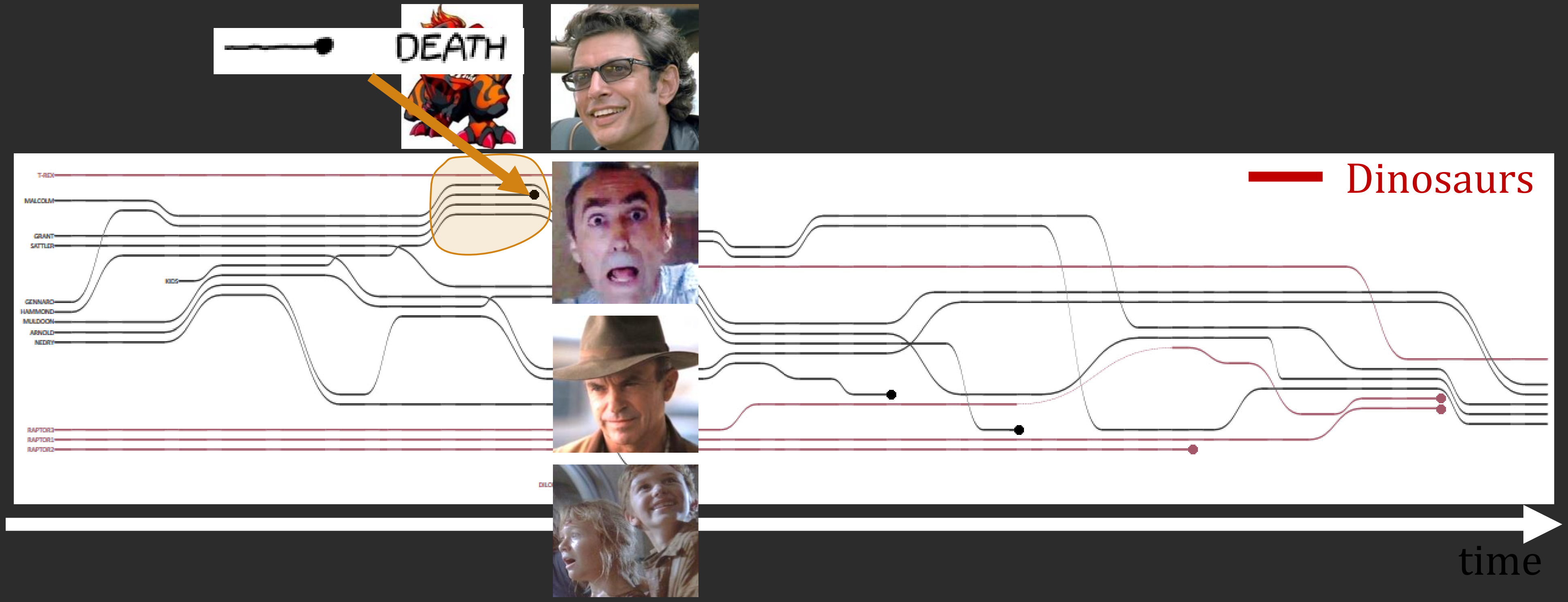
Microsoft Research 微软亚洲研究院

Storyline Visualization



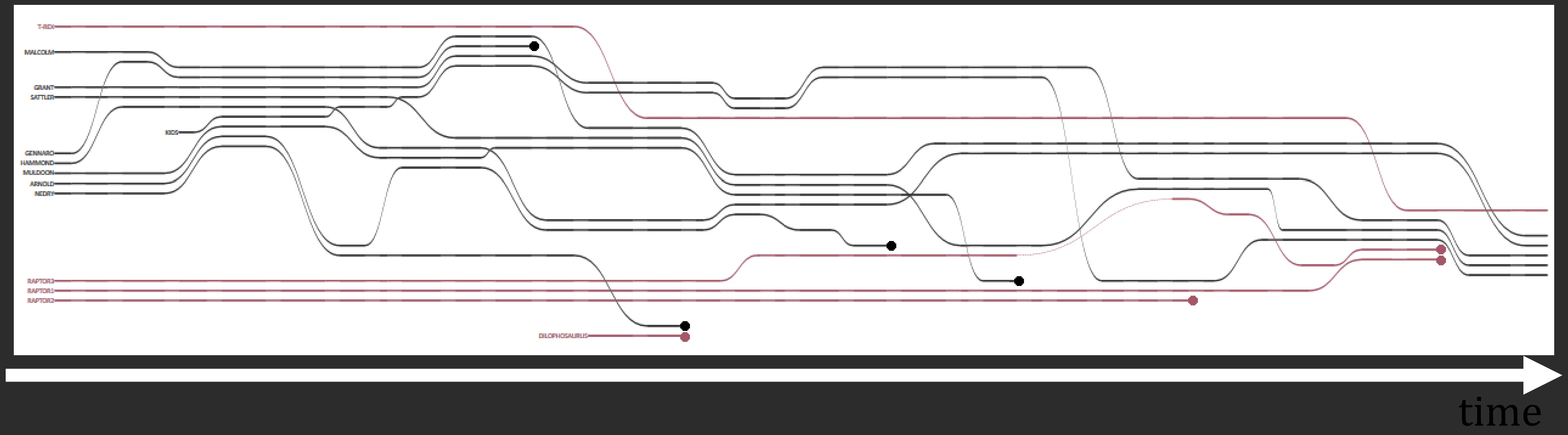
Microsoft Research 微软亚洲研究院

Storyline Visualization

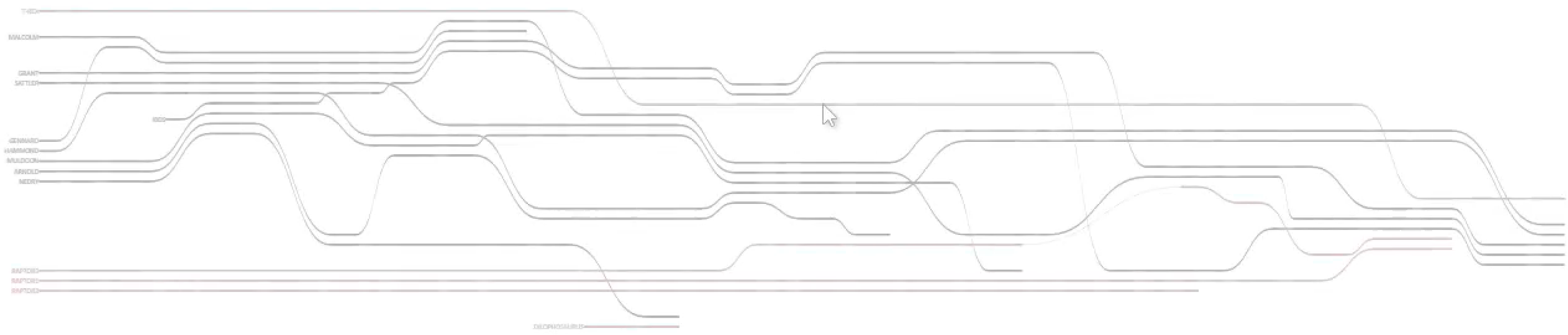


Microsoft Research 微软亚洲研究院

Storyline Visualization



Jurassic Park

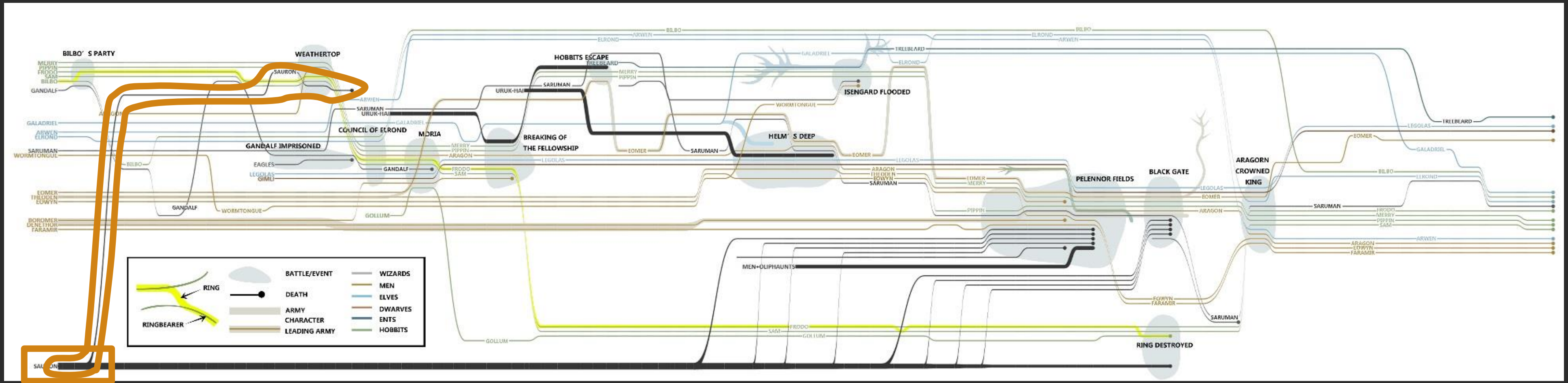


Microsoft Research 微软亚洲研究院

The Lord of the Rings Trilogy

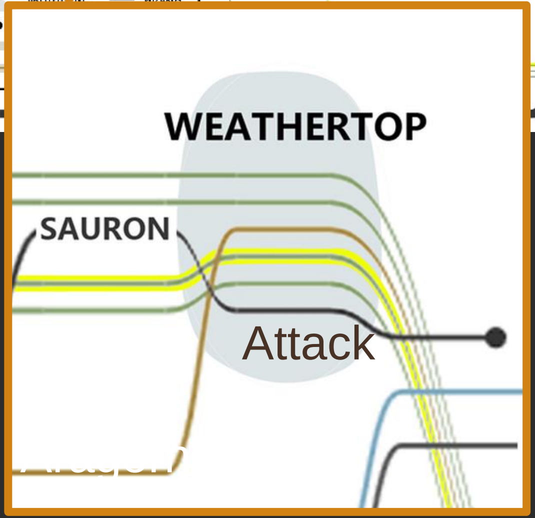
Microsoft Research
微软亚洲研究院

The Lord of the Rings Trilogy

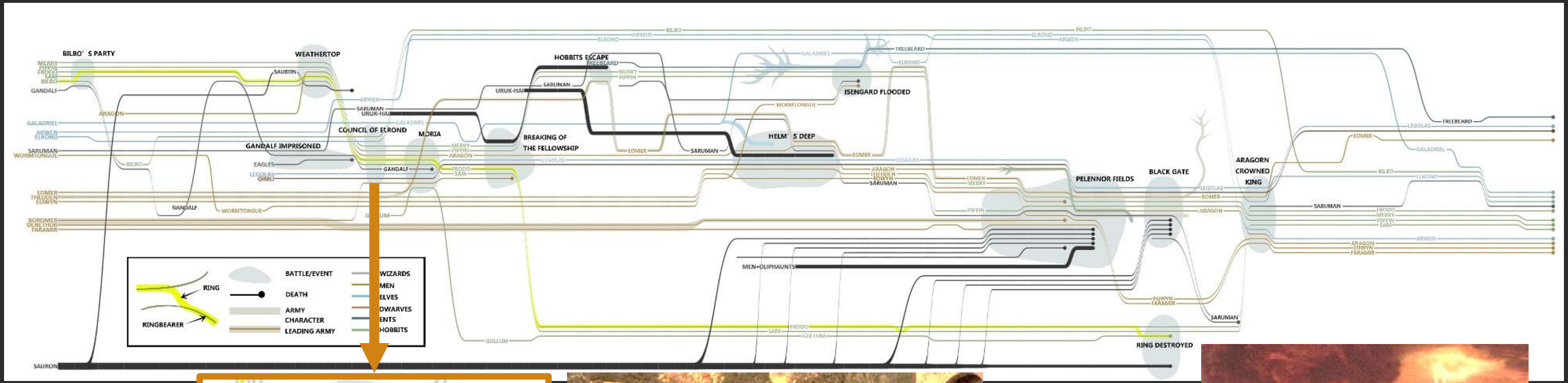


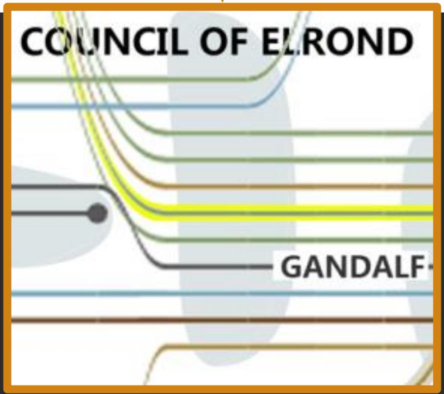
Microsoft Research 微软亚洲研究院

The Lord of the Rings Trilogy



The Lord of the Rings Trilogy

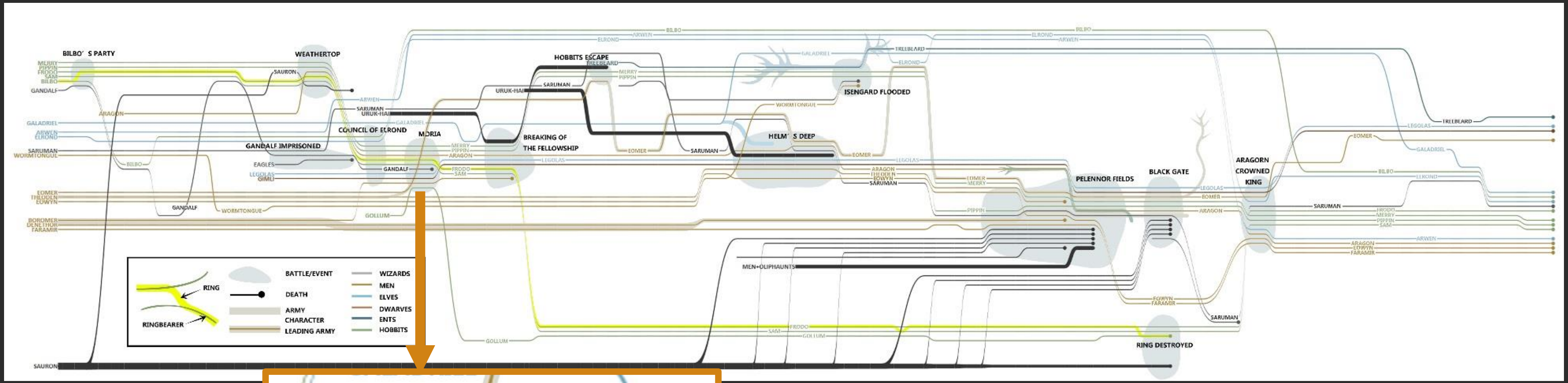


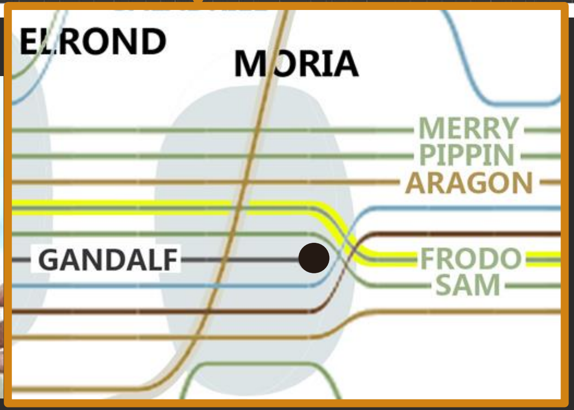


Mount Doom

Microsoft Research 微软亚洲研究院

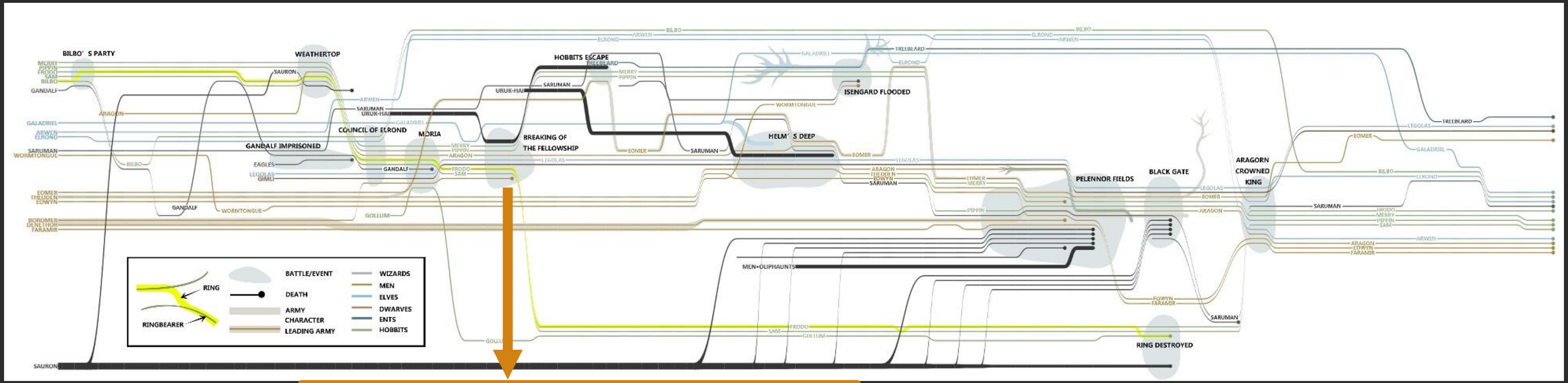
The Lord of the Rings Trilogy

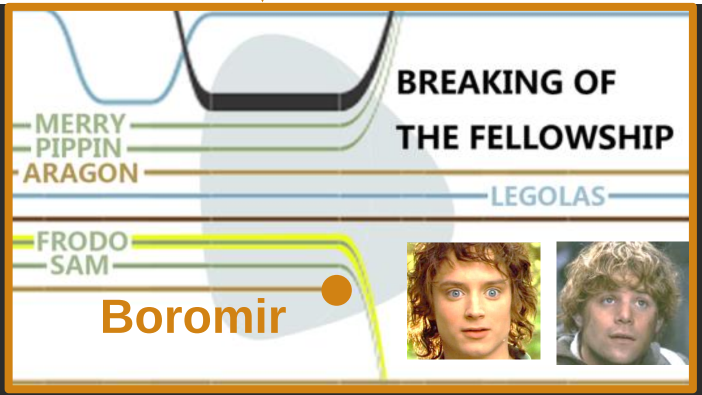




Microsoft Research 微软亚洲研究院

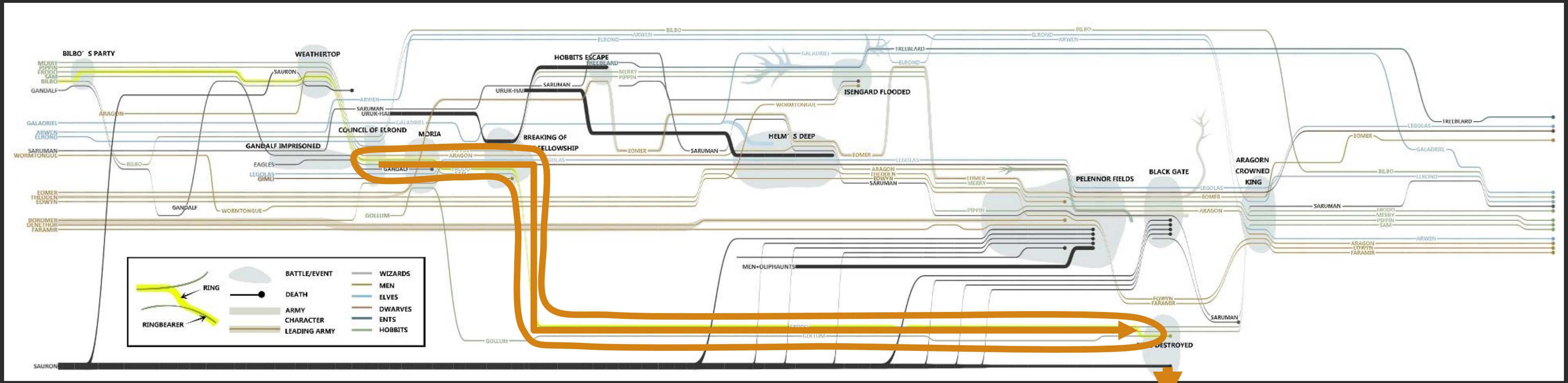
The Lord of the Rings Trilogy





Microsoft Research 微软亚洲研究院

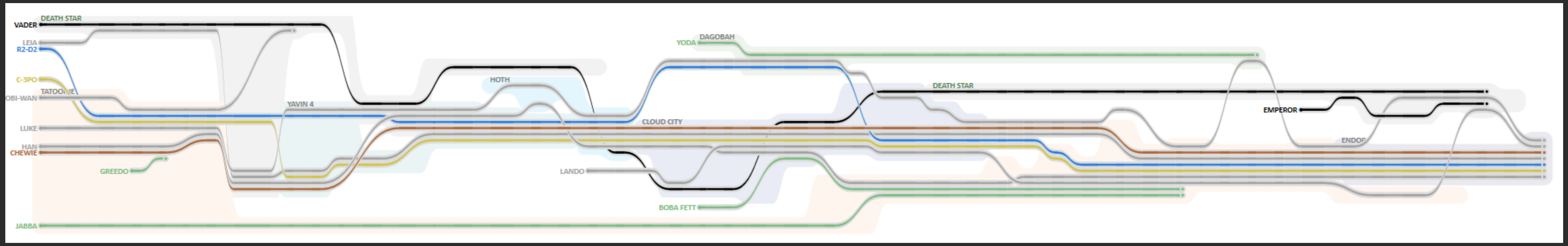
The Lord of the Rings Trilogy



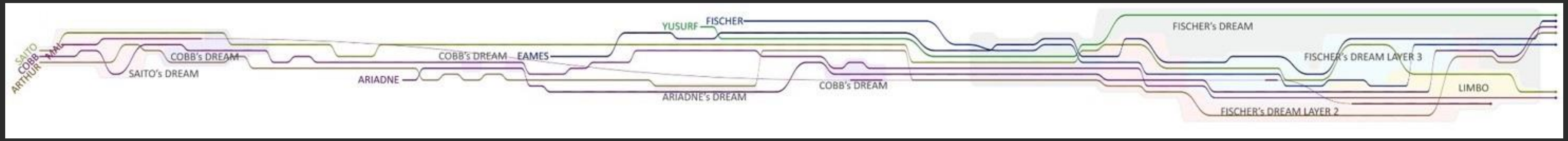
Microsoft Research 微软亚洲研究院

StoryFlow (InfoVis 2013)

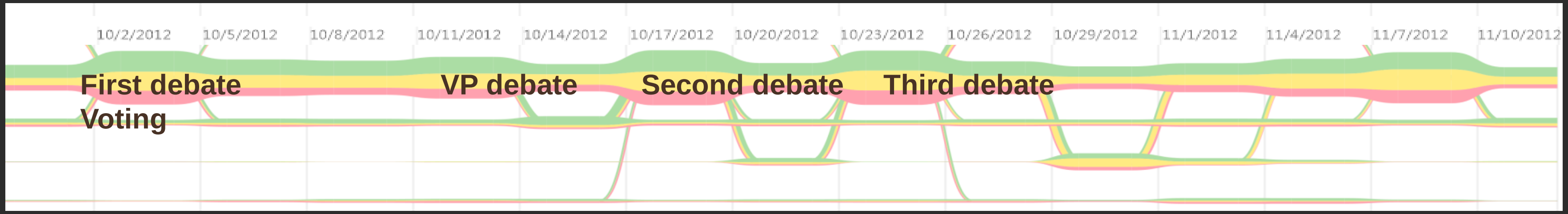
- Real-time



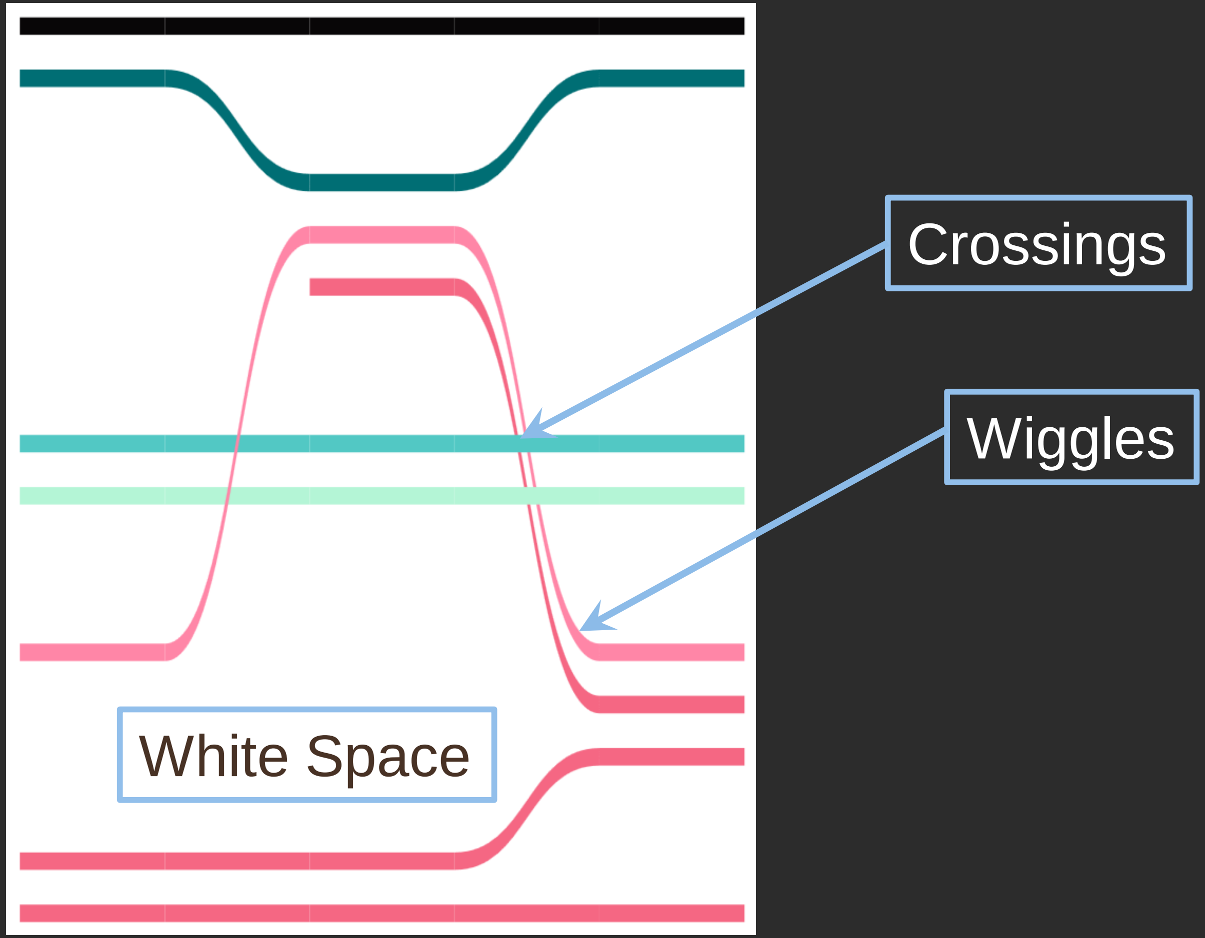
- Hierarchy



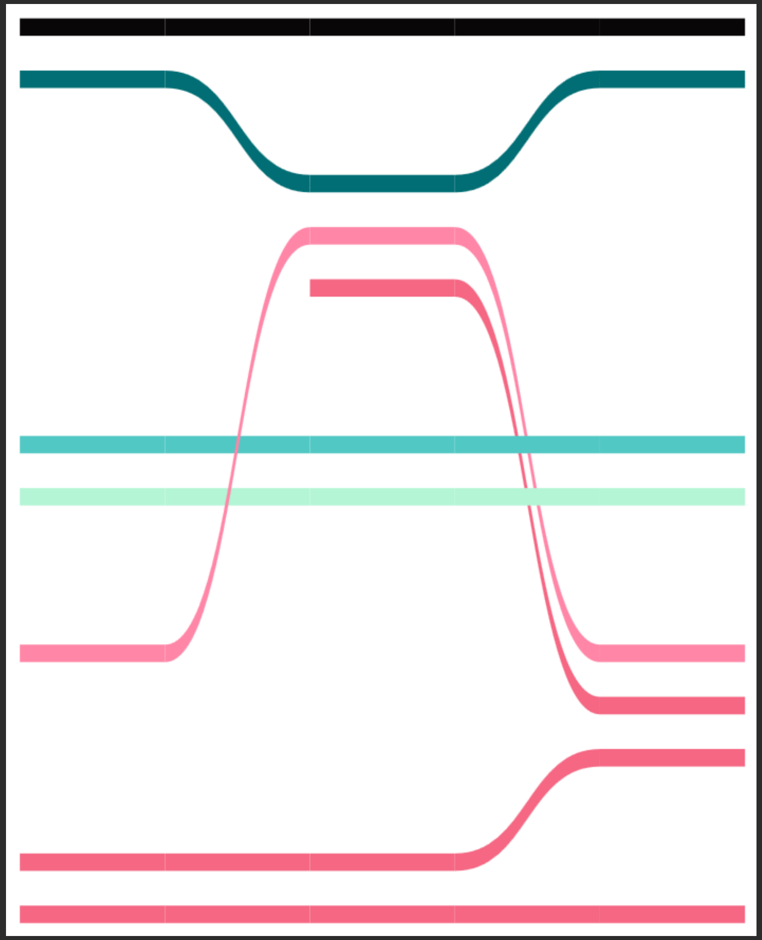
- Level-of-detail rendering



Objective



Strategy



Crossings

Wiggle number

Wiggles

Wiggles distance

White space

Importance decrease

Discrete

Continuous

Quantitative Analysis

- Intel i7-2600 CPU (3.4GHz), 8GB memory
- Comparison with TM's method

	Data		Time(s)		Crossings		Wiggles	
	#Entity	#Frame	Ours	TM	Ours	TM	Ours	TM
Star Wars	14	50	0.16	129.79	48	93	82	133
Inception	8	71	0.16	149.67	23	99	88	162
Matrix	14	42	0.16	172.47	14	43	54	94
MID	79	523	0.60	>10^5	1267	1871	831	874

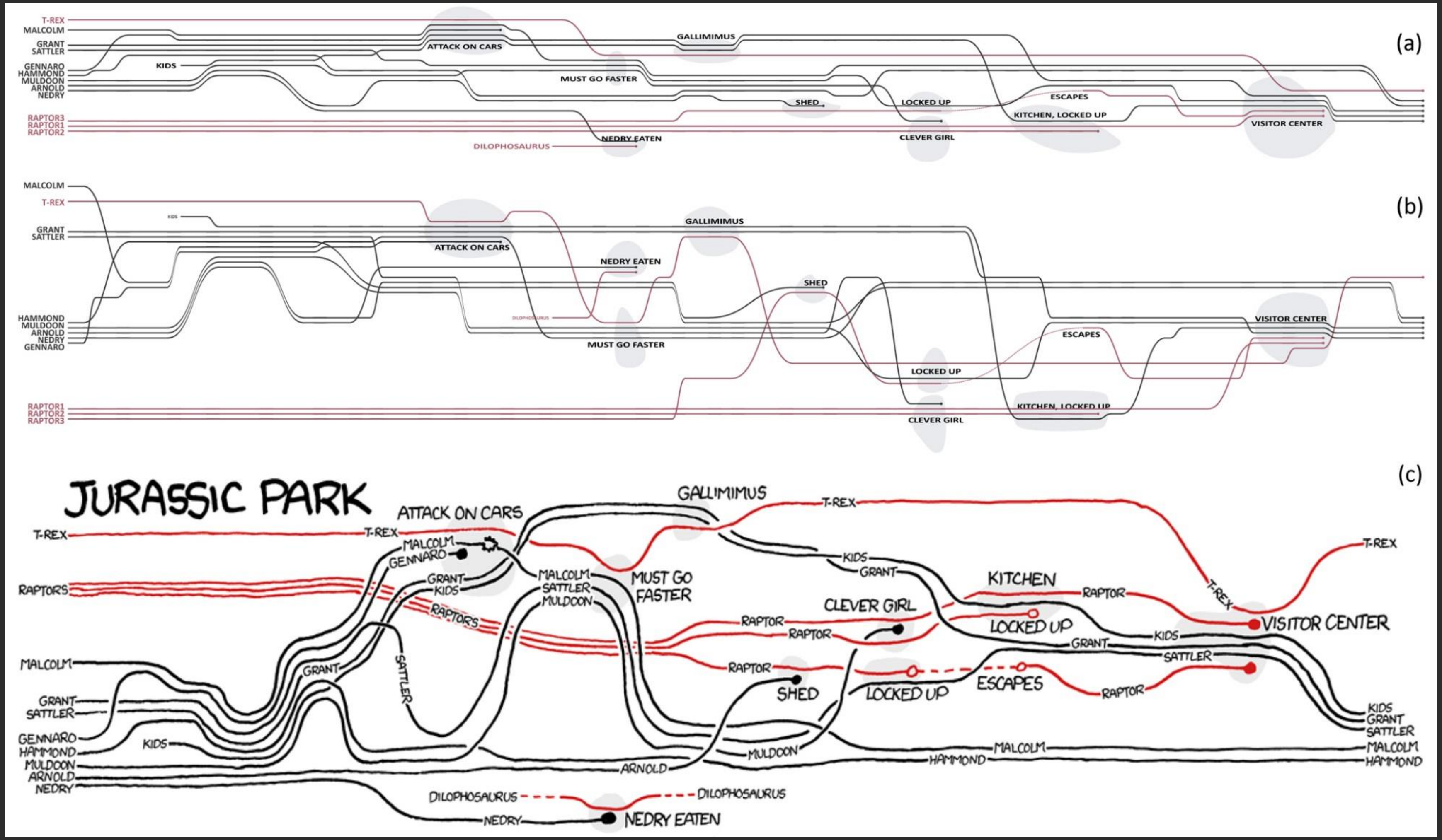
Quantitative Analysis

- Intel i7-2600 CPU (3.4GHz), 8GB memory
- Comparison with TM's method

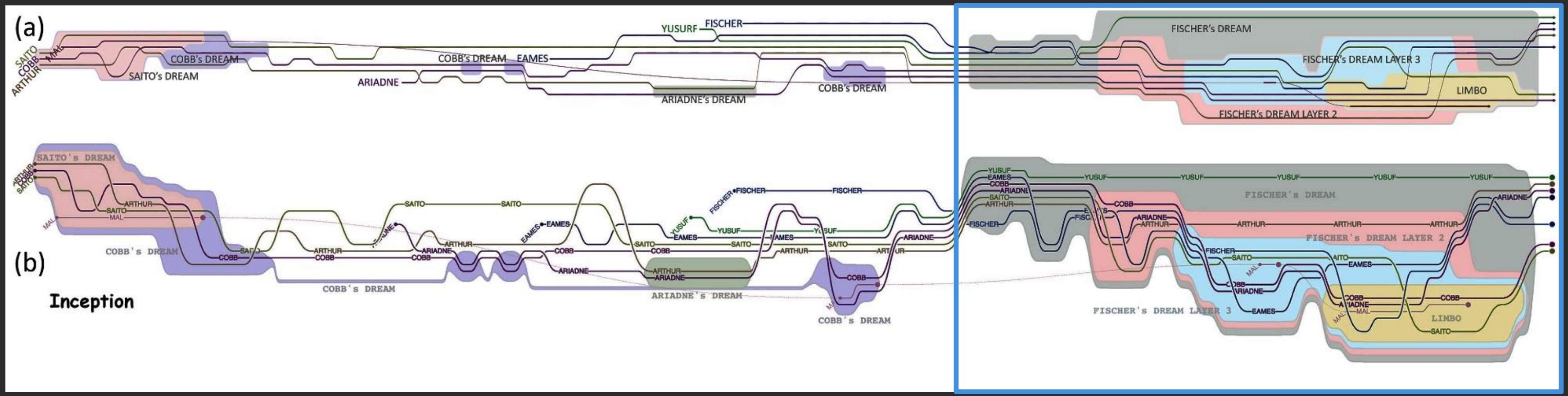
	Data		Time(s)		Crossings		Wiggles	
	#Entity	#Frame	Ours	TM	Ours	TM	Ours	TM
Star Wars	14	50	0.16	129.79	48	93	82	133
Inception	8	71	0.16	149.67	23	99	88	162
Matrix	14	42	0.16	172.47	14	43	54	94
MID	79	523	0.60	>10^5	1267	1871	831	874

Microsoft Research 微软亚洲研究院

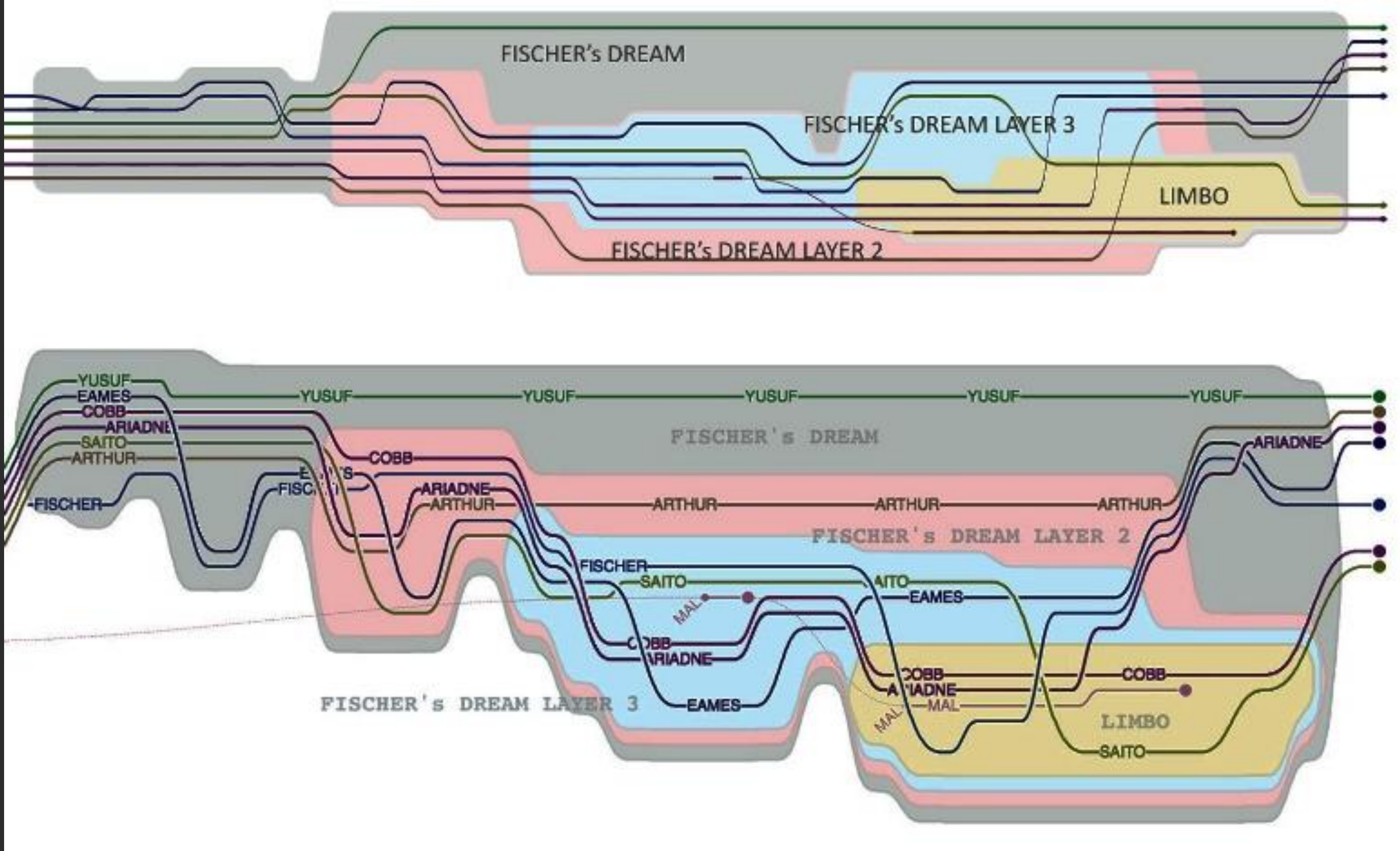
Jurassic Park



Inception

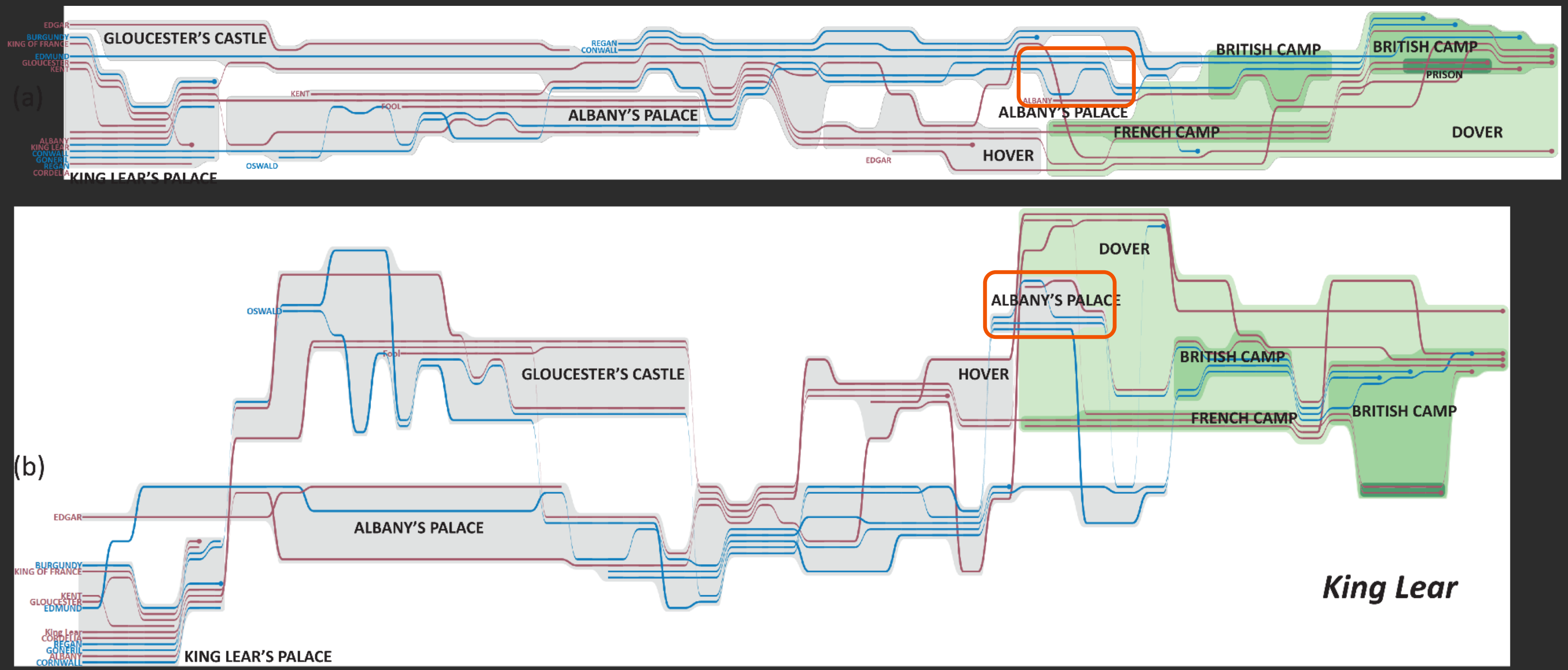

(a)
(b)
Inception
SAITO's DREAM
COBB's DREAM
ARIADNE's DREAM
FISCHER's DREAM
FISCHER's DREAM LAYER 2
FISCHER's DREAM LAYER 3
LIMBO
YUSUF
FISCHER
EAMES
ARIADNE
COBB
SAITO
ARTHUR
MAL

Inception



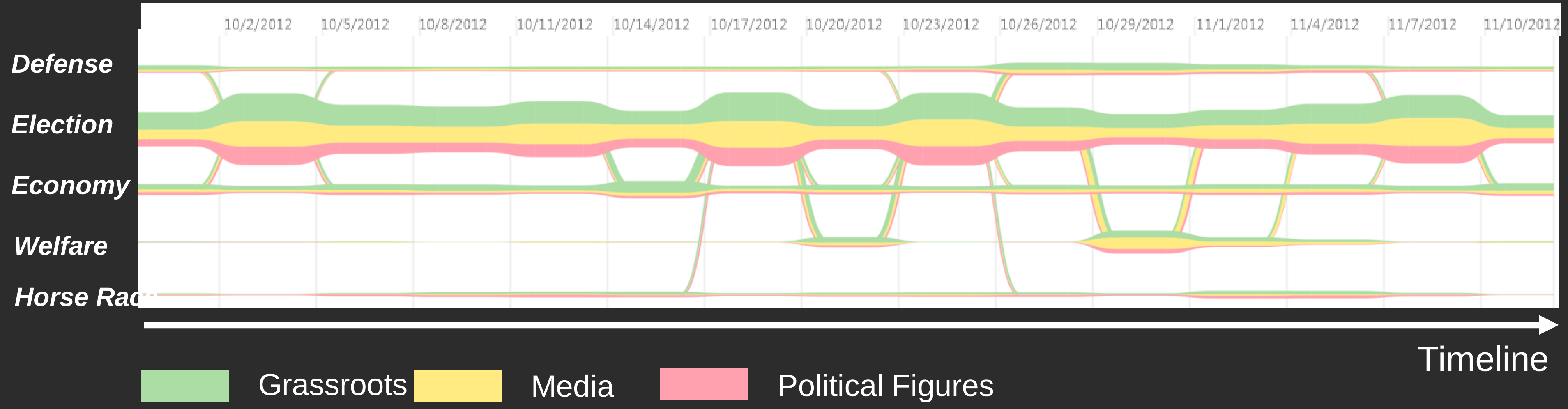
FISCHER's DREAM
FISCHER's DREAM LAYER 3
LIMBO
FISCHER's DREAM LAYER 2
YUSUF
EAMES
COBB
ARIADNE
SAITO
ARTHUR
FISCHER
FISCHER's DREAM
FISCHER's DREAM LAYER 2
MAL
LIMBO
FISCHER's DREAM LAYER 3

King Lear


(a)
GLOUCESTER'S CASTLE
ALBANY'S PALACE
ALBANY'S PALACE
BRITISH CAMP
BRITISH CAMP
PRISON
FRENCH CAMP
DOVER
HOVER
KING LEAR'S PALACE
EDGAR
BURGUNDY
KING OF FRANCE
EDMUND
GLOUCESTER
KENT
FOOL
REGAN
CONWALL
OSWALD
EDGAR
(b)
DOVER
ALBANY'S PALACE
BRITISH CAMP
FRENCH CAMP
BRITISH CAMP
HOVER
GLOUCESTER'S CASTLE
ALBANY'S PALACE
KING LEAR'S PALACE
OSWALD
EDGAR
BURGUNDY
KING OF FRANCE
KENT
GLOUCESTER
EDMUND
CORNWALL
King Lear

Overall Patterns (1/2)

- Three user groups focus mainly on *Election*
  - Grassroots focus on Economy and switched frequently
  - Political figures are more focused
  - Media occasionally switched

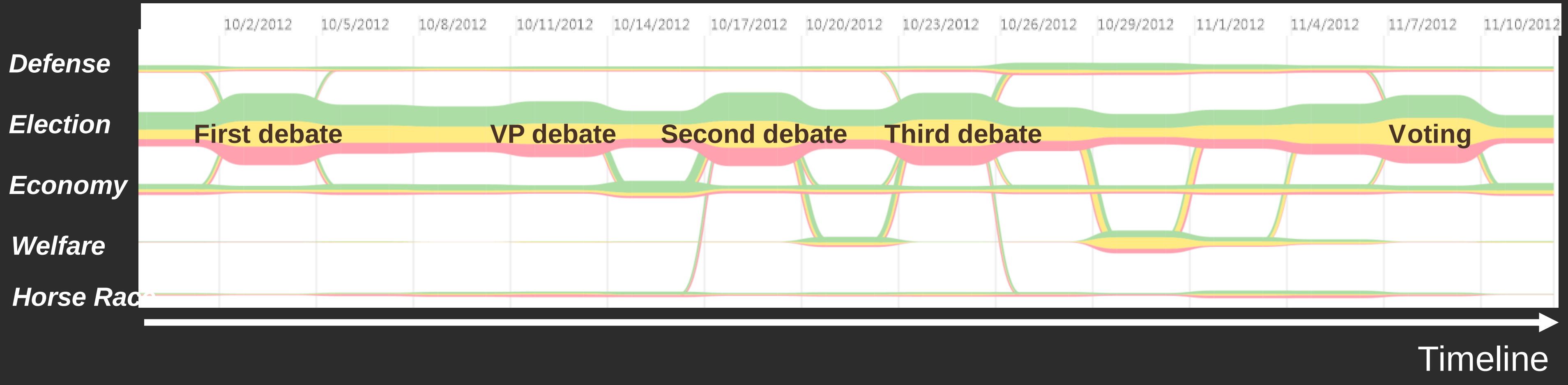


Microsoft Research 微软亚洲研究院

Overall Patterns (2/2)

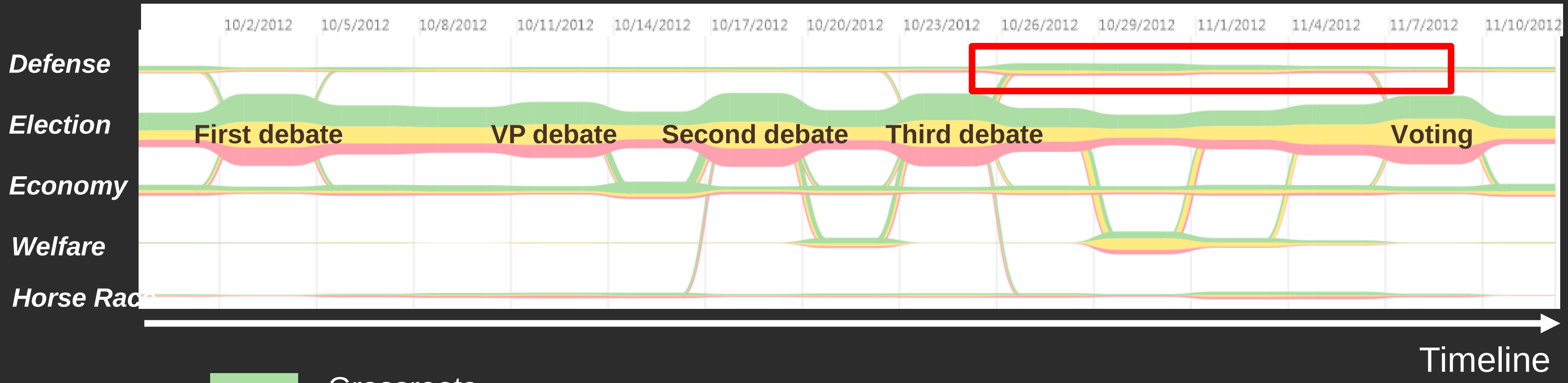
- Five significant peaks on *Election*
  - Third debate on foreign affairs

Grassroots
Media
Political Figures



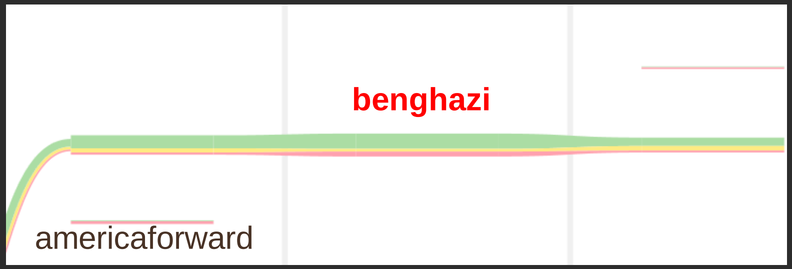
Significant transitions (1/3)

- Transition from *Election* to *Defense*



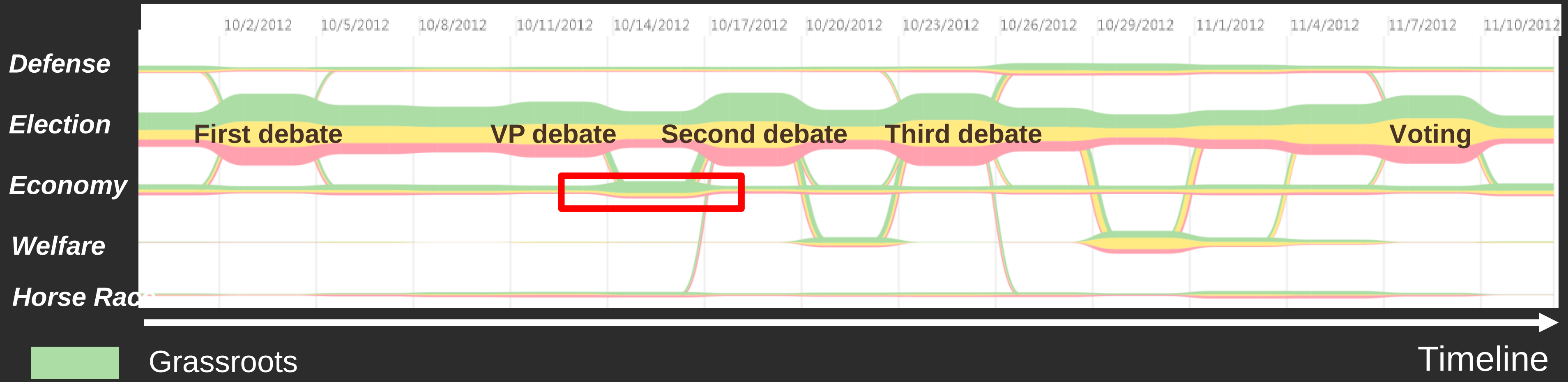
Grassroots
Media
Political Figures

‘doesn’t obama owe us answers (on #benghazi)?



Significant transitions (2/3)

- Transition from *Election* to *Economy*

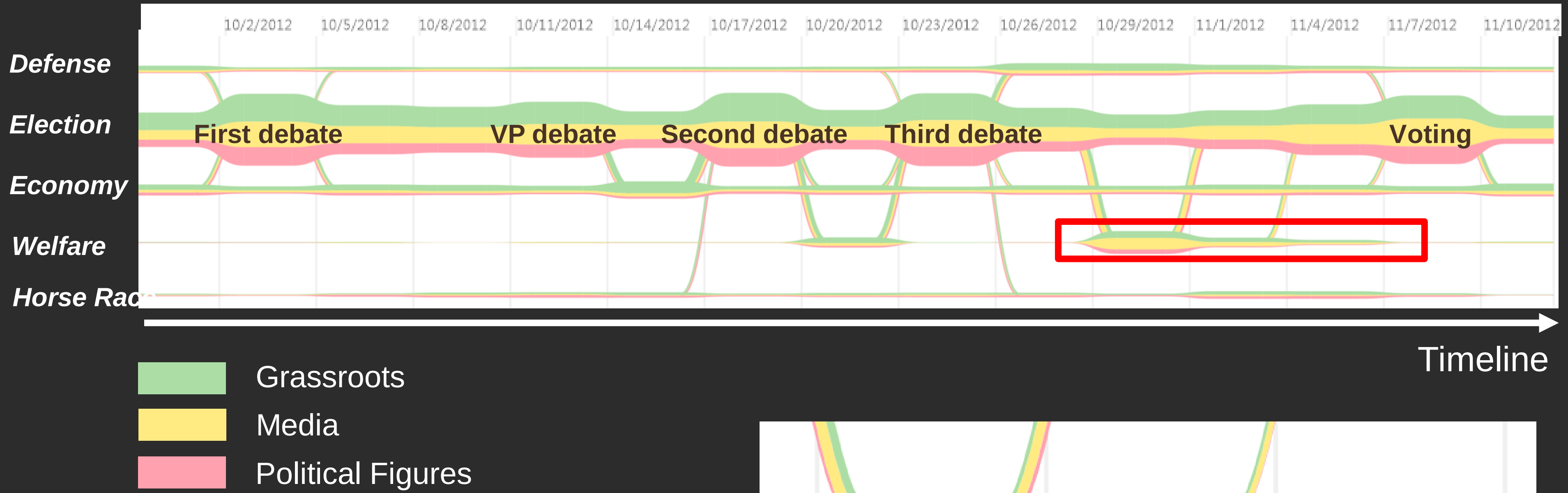


Grassroots
Media
Political Figure

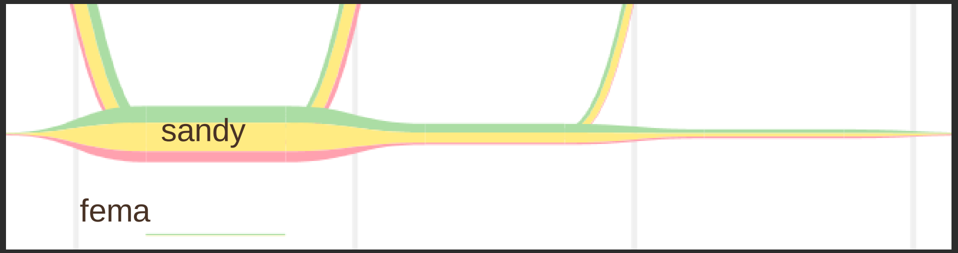
think Romney is tough on china? ask the workers of #sensata about that as they train their Chinese replacements

Significant transitions (3/3)

- Transition from *Election* to *Welfare*



Issue-attention cycle



Microsoft Research 微软亚洲研究院

"Big Picture": Our Key Focuses

Multifaceted Behavior Analytics

Step 3

OpinionFlow

SocialFlow

Implicit Behavior Analytics

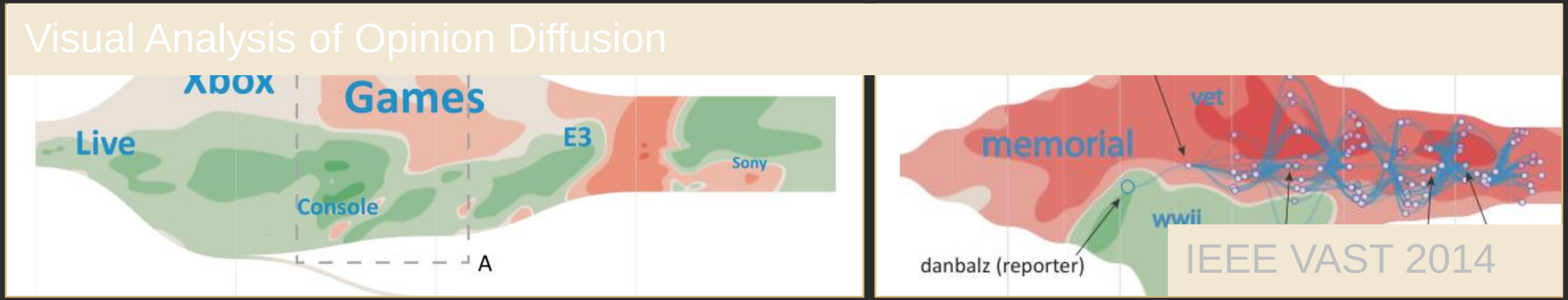
Step 2

Explicit Behavior Analytics

Step 1

Implicit Behavior Analytics: OpinionFlow

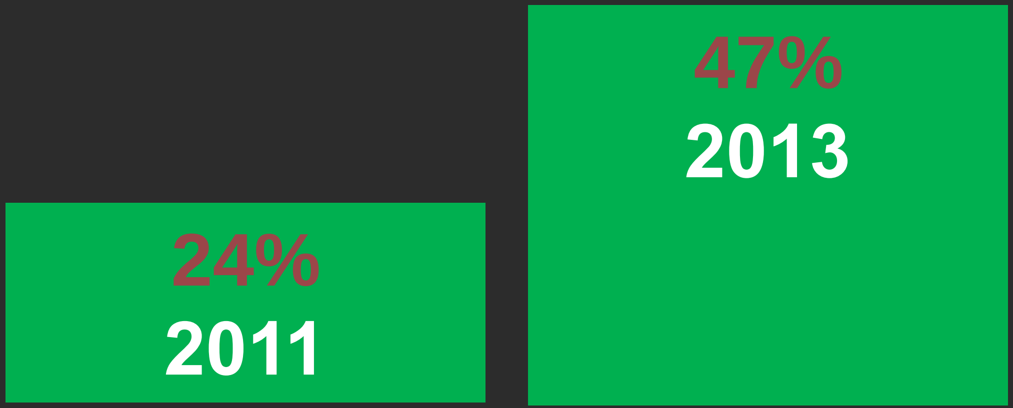
- OpinionFlow: visual analysis and exploration of implicit opinion propagation on social media



Microsoft Research 微软亚洲研究院

Opinion on Social Media

- Forrester 2009: 78% consumers trust peer recommendations on social media
- Facebook impact on purchase

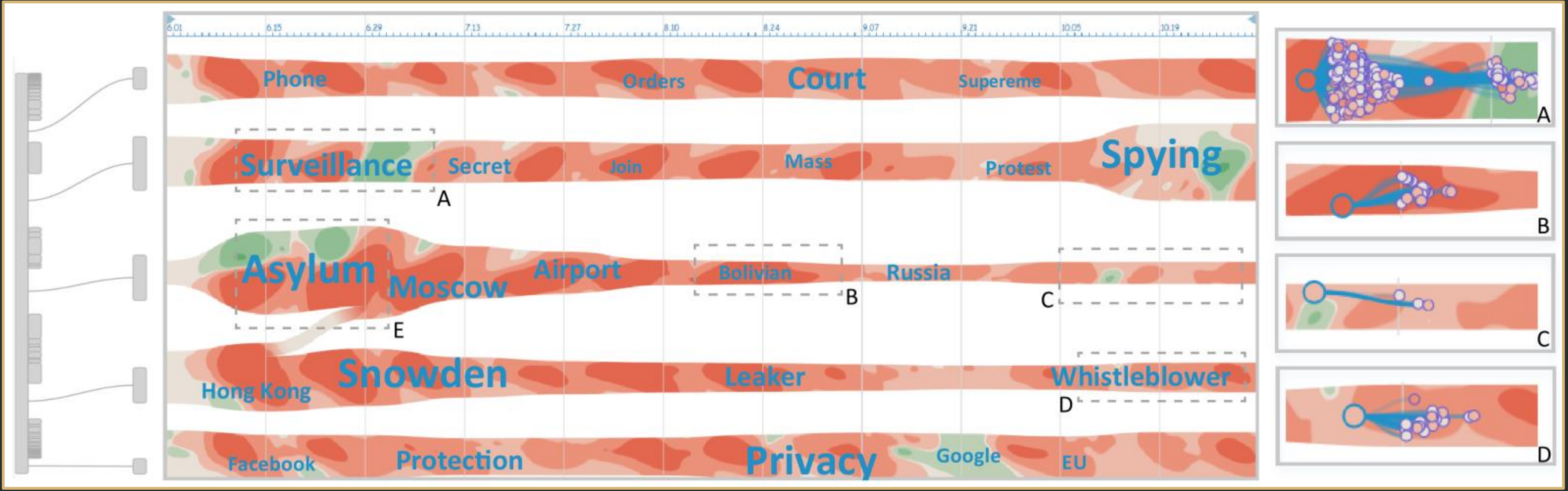


Analyzing opinion diffusion is important

Major Challenges

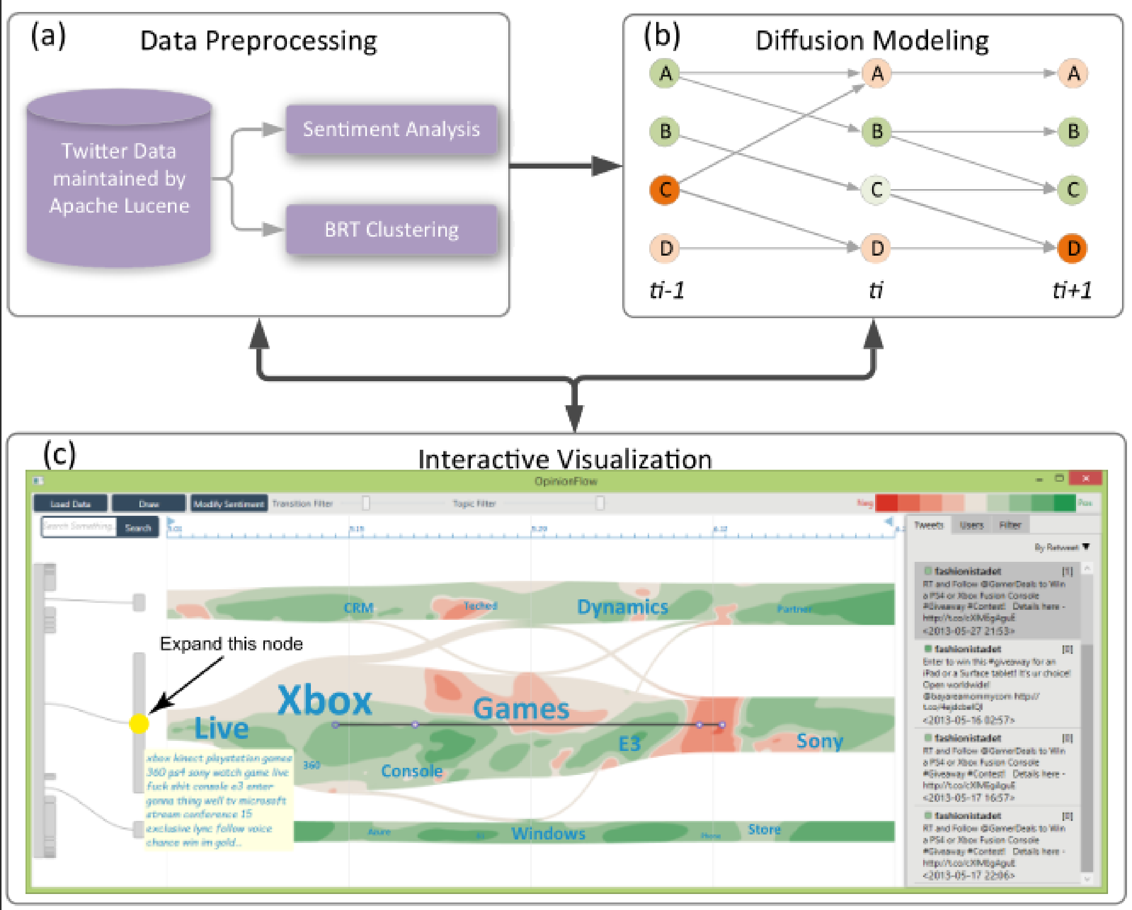
Key Focus of OpinionFlow

Visually trace and analyze opinion diffusion on social media in large-scale events



System Overview

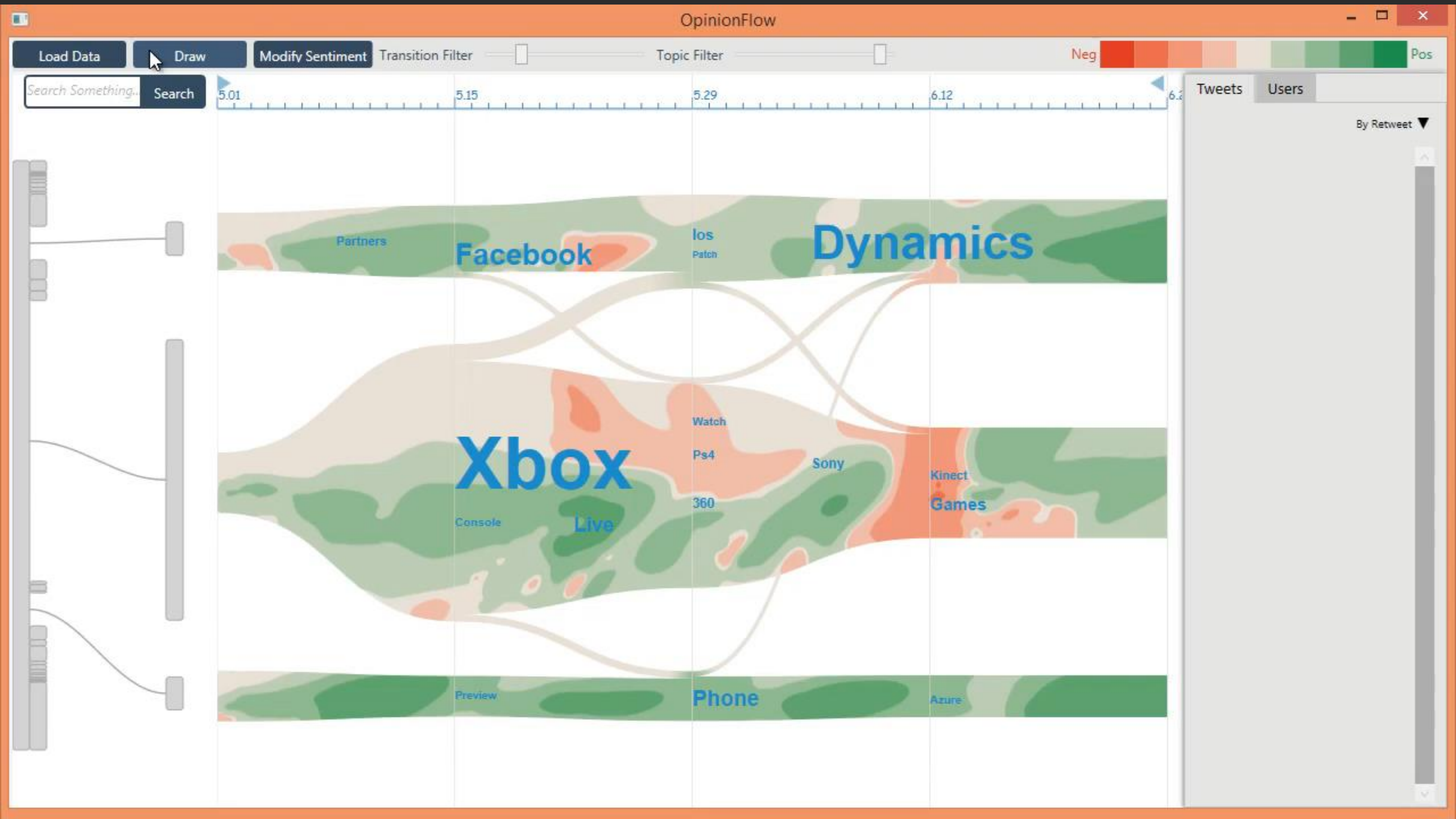
- Data analysis + diffusion modeling + interactive visualization



Questions

- Question A: overview of topics and opinions
  - What are the major opinions in a large-scale event on social media?
  - How are they changing over time?

- Question B: diffusion of opinions
  - What different roles do individual users play in opinion diffusion?
  - How does the attention transition behavior of users relate to opinion diffusion?



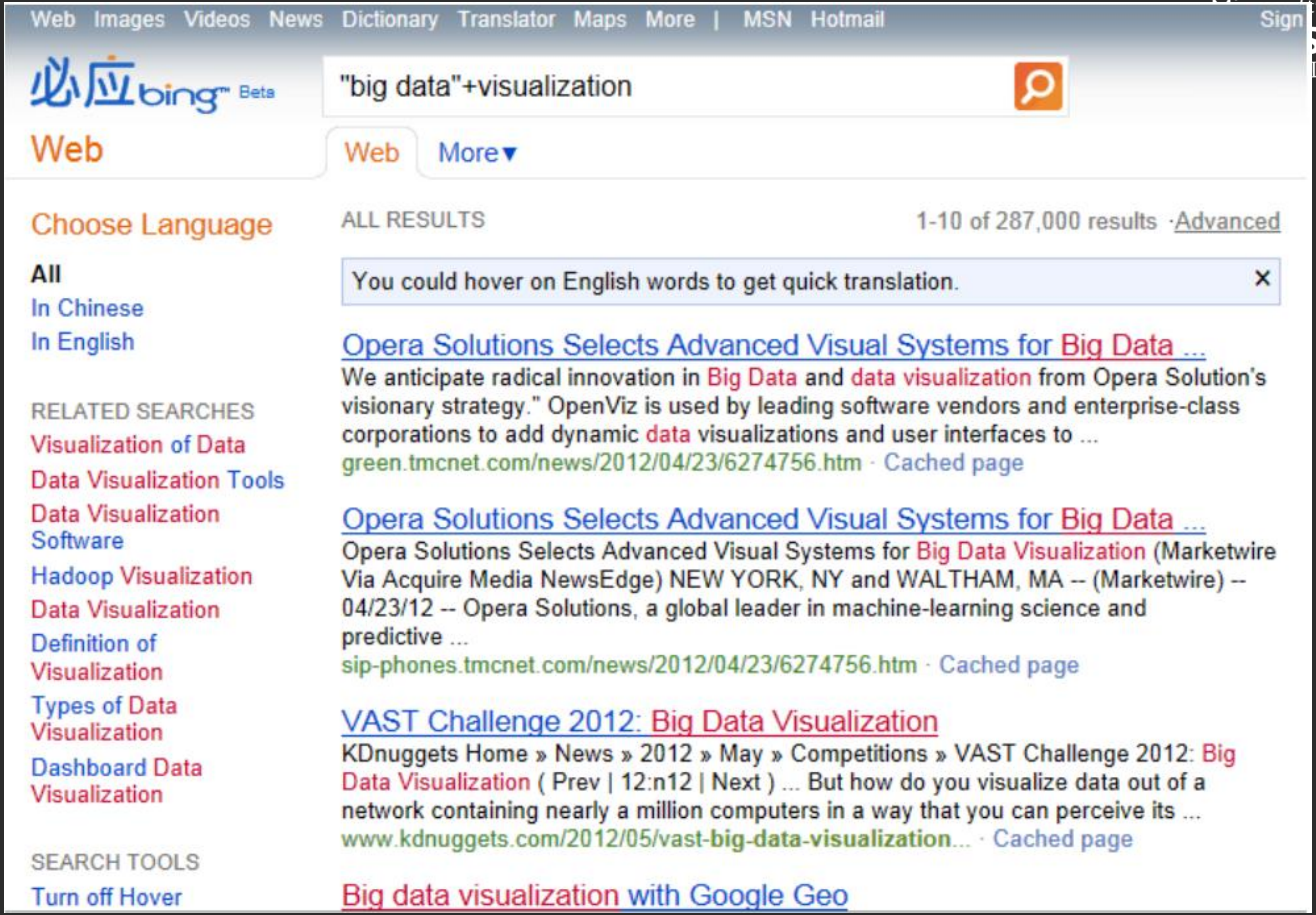
OpinionFlow
Load Data
Draw
Modify Sentiment
Transition Filter
Topic Filter
Neg
Pos
Search Something...
Search
5.01
5.15
5.29
6.12
Tweets
Users
By Retweet
Partners
Facebook
Ios
Patch
Dynamics
Xbox
Watch
Ps4
360
Sony
Kinect
Games
Console
Live
Preview
Phone
Azure

Outline

- Definition
- Objective
- History
- InfoVis 1.0
- InfoVis 2.0
- InfoVis 3.0
- **Opportunities and challenges**

Microsoft Research 微软亚洲研究院

Massive Data - Golden Mountain



search
亚洲研究院

Big Data's Hot Cousin

Data visualization: Big Data's hot cousin

Clever coders make data fun and easy, as vendors add tools to catch up

By Simon Sharwood, 26th April 2012

1

Moritz Stefaner calls himself a "Truth and Beauty Operator". But if you don't understand what that means, he'll translate to the safer "Freelance information Visualizer" and explain that "Large companies approach me with large data sets they want visualized. I take their data and turn it into something beautiful."

Beauty is important to Stefaner because he feels it is a way to address the problem of data overload.

NEWS

Analysts: Data visualization tools key to 'big data' analytics success

Mark Brunelli, Senior News Editor
Published: 30 Nov 2011

Demand for data visualization tools is rising sharply, partly as a result of more companies seeking to gain valuable business insights through "big data" analytics initiatives. But achieving success with data visualization often requires fresh thinking about how to present information to business users, especially in big-data environments, according to data management analysts.

Keys to big data in the brain, not the computer, former NSA exec says

- By Rutrell Yasin
- Apr 26, 2012

Data visualization tools are critical to the success of big data analytics. But achieving that success will require more than people with engineering skills; it will require input from cultural anthropologists and social scientists, a former National Security Agency executive told an audience in the Washington, D.C. metro area.

Richard Schaeffer, former information assurance director for NSA and head of the consulting firm Riverbank Associates, said April 26 in a keynote address at SAP National Security Services' Big Data Day in McLean, Va.

Big Data You Can Touch

By Rich | Published July 9, 2012

Steven Overly at the Washington Post writes that a new Startup called Augaroo aims to help companies make sense of Big Data through visualization. The company is developing a product called ZoomData that will allow businesses to create data charts and graphics on iPads and other Web-enabled devices.

Jeffrey MacMillan/Capital Business - Justin Langseth of Augaroo makes a pitch for funding at Capital

Obama’s big data plans

- The Department of Energy is getting in on the big data frenzy, too, investing $25 million to develop Scalable Data Management, Analysis and Visualization Institute, which aims to develop techniques for visualizing the incredible amounts of data generated by the department’s team of supercomputers.

Obama's big data plans

- A $2 million award for a research training group in big data that will support training for undergraduate and graduate students and postdoctoral fellows using novel statistical, graphical, and visualization techniques to study complex data.

Microsoft Research 微软亚洲研究院

Challenges

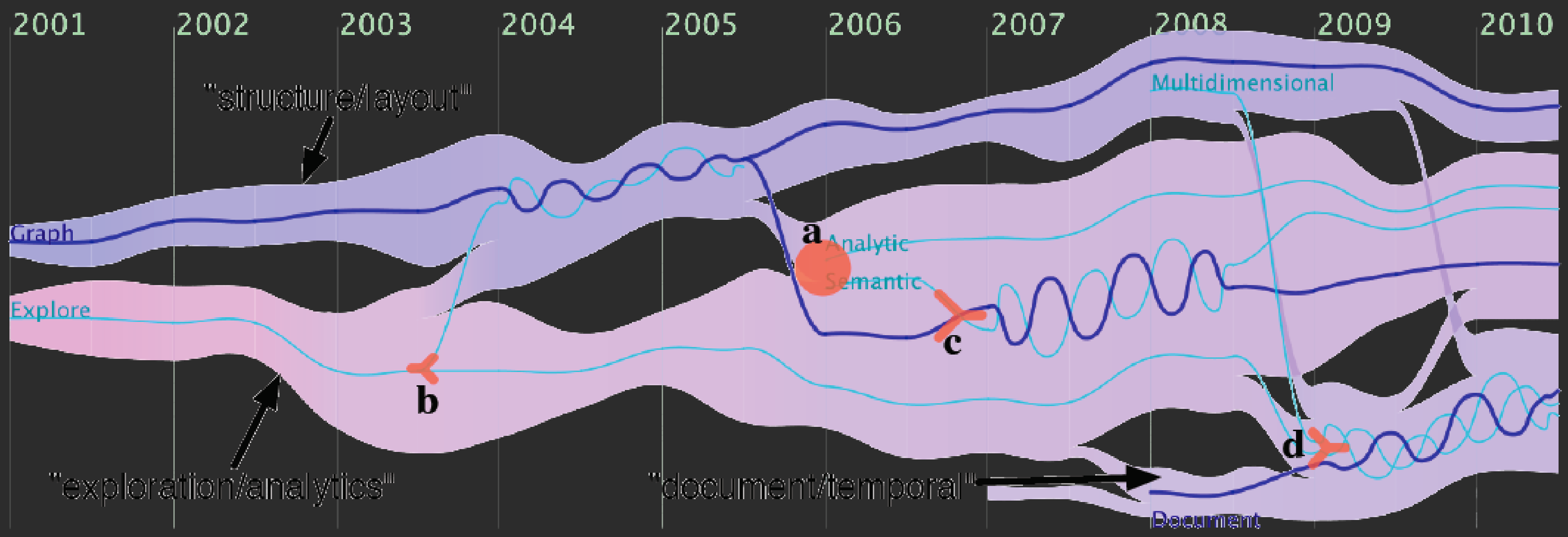
- Big data availability
- Data quality is not high
- Collaboration with domain
- Multiple skills
- Different evaluation systems
- Scalability is a big problem

Acknowledgements

- HKUST
  - Conglei Shi, Panpan Xu, Prof. Huamin Qu
- Georgia Tech
  - Mengdie Hu, Prof. John Stasko
- Tsinghua University
  - Jianfei Chen, Prof. Jun Zhu
- MSRA colleagues and interns
  - Weiwei Cui, Yingcai Wu, Yangqiu Song, Furu Wei, Xin Tong, Xiting Wang, Mengchen Liu, Kai Yan,…

Microsoft Research 微软亚洲研究院

Thanks a lot for your attention!



- Homepage: http://research.microsoft.com/en-us/um/people/shliu/
- Email: shixia.liu@microsoft.com