ContextWing : Pair-wise Visual Comparison for Evolving Sequential Patterns of Contexts in Social Media Data Streams

Yuheng Zhao, Xinyu Wang, Chen Guo, Min Lu, Siming Chen

School of Data Science, Fudan University (复旦大学大数据学院)

Speaker - Yuheng Zhao (赵宇恒)

Content

- **Introduction**
  - Background
  - Goals
- **Related Work**
- **Overview**
  - Analysis Tasks
- **Modeling**
  - Topic and keywords extraction
  - Pair-wise Correlation
  - Public Attention
  - Pattern Generation
  - Streaming Analysis
- **Visualization Design**
  - Design rationales
- **System overview**
  - Layout and Interaction
- **Evaluation**
  - Case study
- **Discussion & Conclusion**

Background

Social media is a common and popular opinion sharing platform.

With the explosive increase of text data, we need to analyze and extract content to know the public opinions in an event.

Visualization provides us a quick, clear understanding of the information.

created_at	replies_count	retweets_count	likes_count	tweet
2020-10-23 19:35:17	7	147	181	Here's why, in the final #debate, Trump's "China" attacks on Biden blew up in his face: Trump withdrew $15M from China from a secret bank account. The cash was deposited in 2017 AFTER he be
2020-10-23 09:32:27		2	3	#Trump can't stop lying. I'm so tired of it. Trump has secret bank account in China. #Russia pays Trump. #PresidentialDebate2020
2020-10-23 09:29:13	6	52	100	Trump throws Fauci under the bus, blames Biden for back room deals, and won't answer questions directly. What is he hiding? Why can't we get an actual straight answer from Trump? Joe answers
2020-10-23 18:56:42		3	6	Kristen Welker: " A report this week, which was referenced, does indicate that your company has a bank account in China. So how can voters know that you don't have any foreign conflicts of intere
2020-10-23 09:41:19	1	21	41	No, Don. Nobody knew about your bank account in China, which you took millions from after you became President. That's why it's news. https://t.co/DFNoluupl0 #Debates2020 #Debate

But how to use visualization to help quickly understand and summarize messages' opinions?

- **Give visual summary and gain insights by visual comparison and exploration.**

Goals

A. What do people discuss in an event?

→ Texts → Topics → Keywords → Contexts

Goals

B. Can we organize keywords and know more?

Keywords → Contextual sequence

→ I have a pen and an apple.

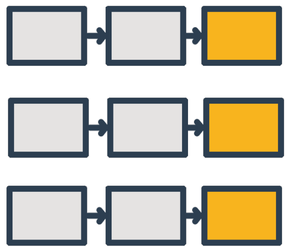
- Pen is a keyword
- Have, apple are contexts of pen.

→ If we know the sequence "Have-pen-apple" in order, we can know the overview meaning of the sentence.

Goals

C. Are there some frequent contextual sequences to form a pattern?

Contextual sequence



Pattern

Have-pen-apple × 99

Have-pen-apple × 99

Have-pen-apple × 99

A Pattern

Diversified patterns

Goals

D. What are the relationships between contents and key entities?

Texts → Themes → Keywords → Contexts (patterns)

→ **Key entities**

→ **Data Streams**

Goals

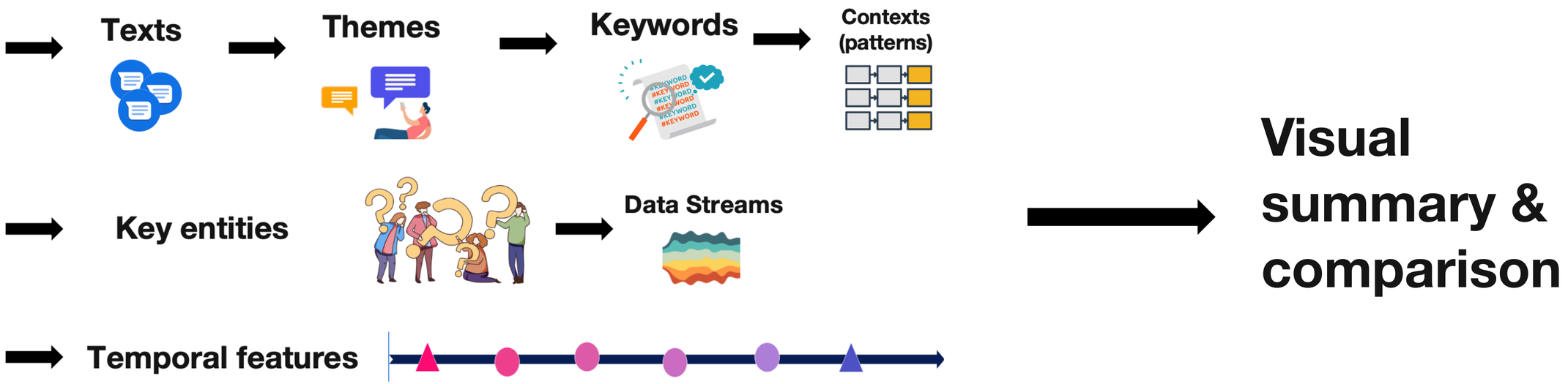
E. How these contents and relationships evolve over time?

Texts → Themes → Keywords → Contexts (patterns)

Key entities → Data Streams

Temporal features

Goals

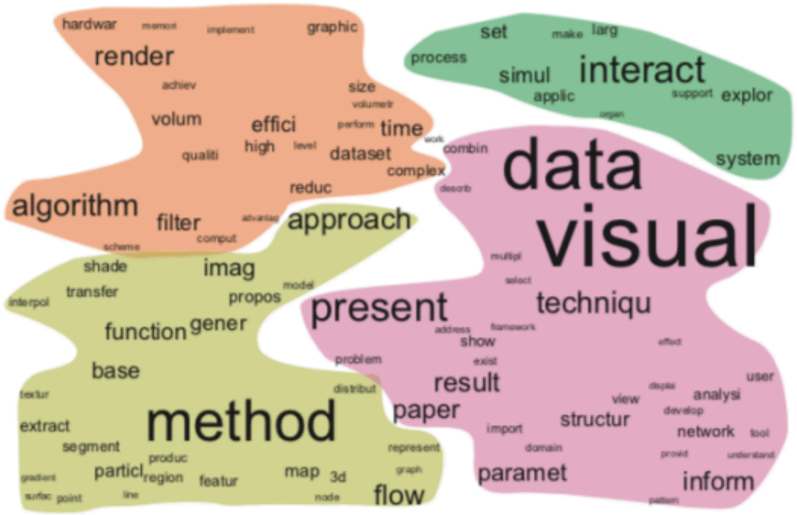


- A **novel visualization** technique for pairwise visual comparison.
- A **visual analysis system** to explore patterns of contexts in both semantic and temporal features.
- An **in-depth evaluation** of ContextWing.

Related work

Static text visualization

- Word cloud or word tree
- Not support **contextual connection** to represent sentence-level meaning
- Difficult to support **temporal comparison**

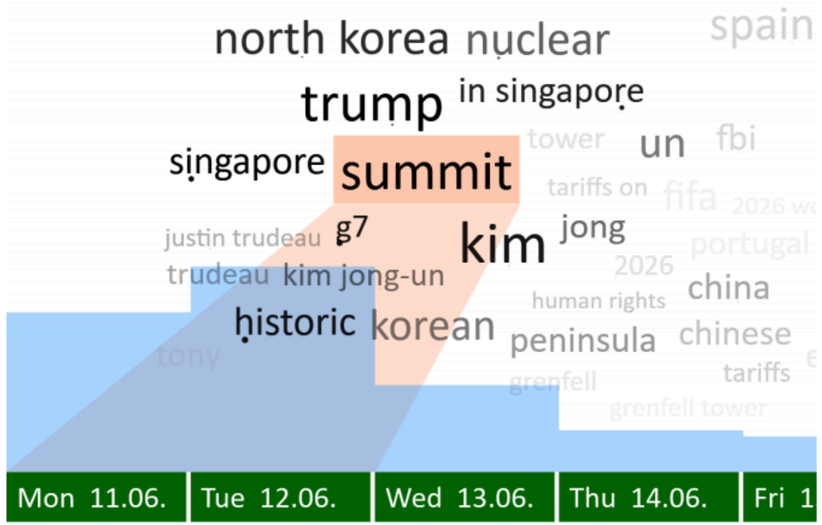


[1] Y. Wu, T. Provan, F. Wei, S. Liu, and K.-L. Ma, “Semantic preserving word clouds by seam carving,” in Computer Graphics Forum, vol. 30, no. 3, 2011, pp. 741–750.
[2] M.Hu,K.Wongsuphasawat,andJ.Stasko.2017.Visualizing Social Media Content with SentenTree.IEEE Transactions 23,1(2017),621–630.

Related work

Time-varying text visualization

- Timeline + word cloud
- Timeline + parallel tag cloud
- Not focused on **contextual information** and **pair-wise comparison**

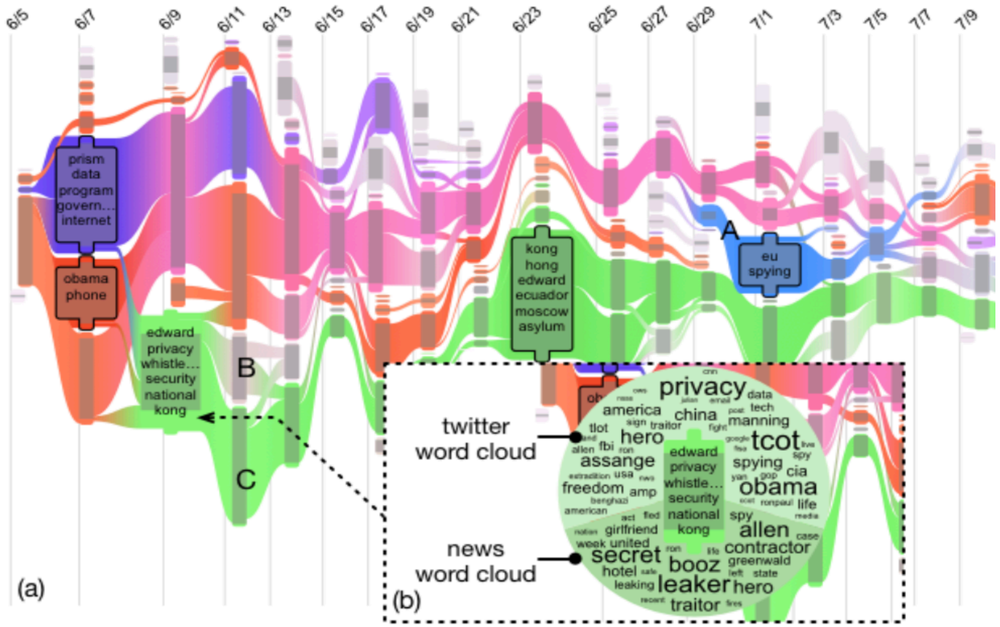


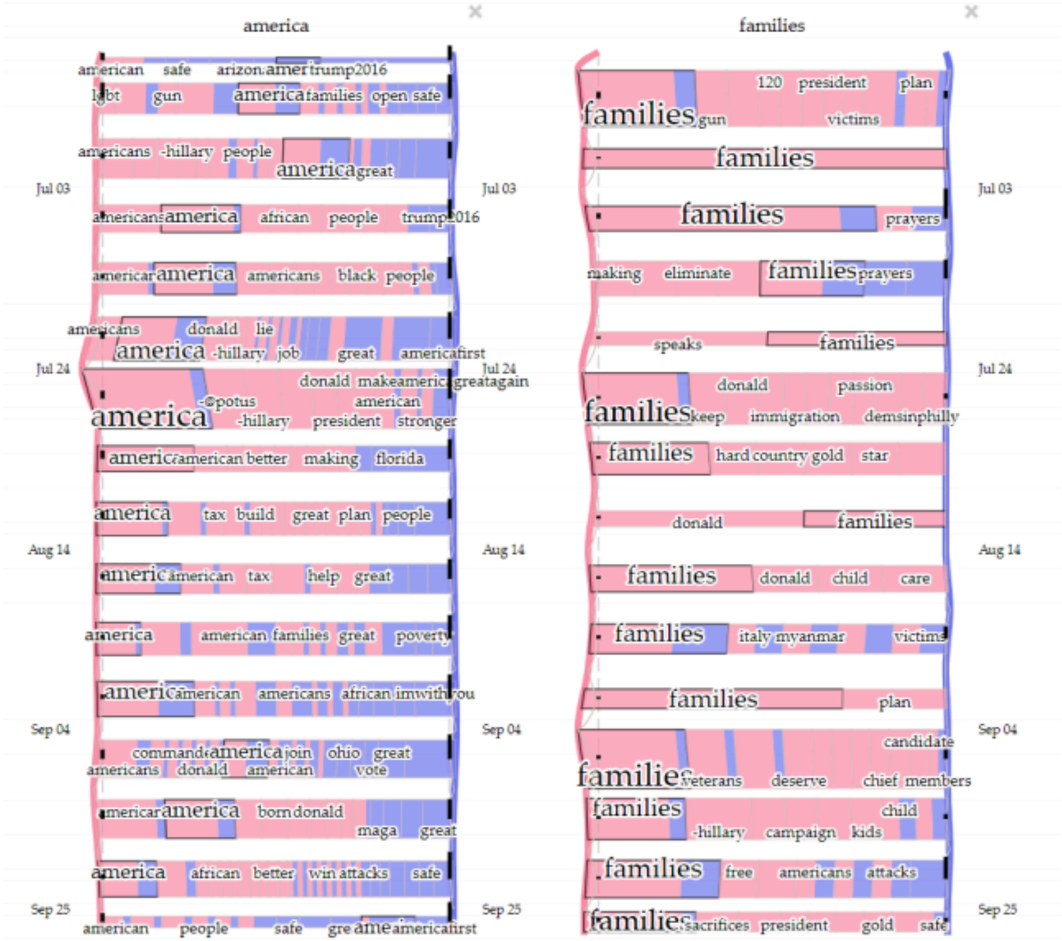
[1] J. Knittel, S. Koch, and T. Ertl. 2020. PyramidTags: Context-, Time- and Word Order-Aware Tag Maps to Explore Large Document Collections. IEEE Transactions(2020),1–1.
[2] C. Collins, F. B. Viegas, and M. Wattenberg. 2009. Parallel Tag Clouds to explore and analyze faceted text corpora. In 2009 IEEE Symposium on Visual AnalyticsScienceandTechnology.91–98.

Related work

Data stream comparison

- Different data steams + word cloud
- Pair-wise data steams + juxtaposition
- Not focused on **contextual information**





[1] Weiwei Cui, Shixia Liu, Zhuofeng Wu, and Hao Wei. 2014. How Hierarchical Topics Evolve in Large Text Corpora. Visualization and Computer Graphics,IEEETransactionson20(122014),2281–2290.
[2] S. Chen, N. Andrienko, G. Andrienko, J. Li, and X. Yuan. 2021. Co-Bridges: Pair-wise Visual Connection and Comparison for Multi-item Data Streams.IEEETransactionson27,2(2021),1612–1622.

Related work

✏️ Brief summary

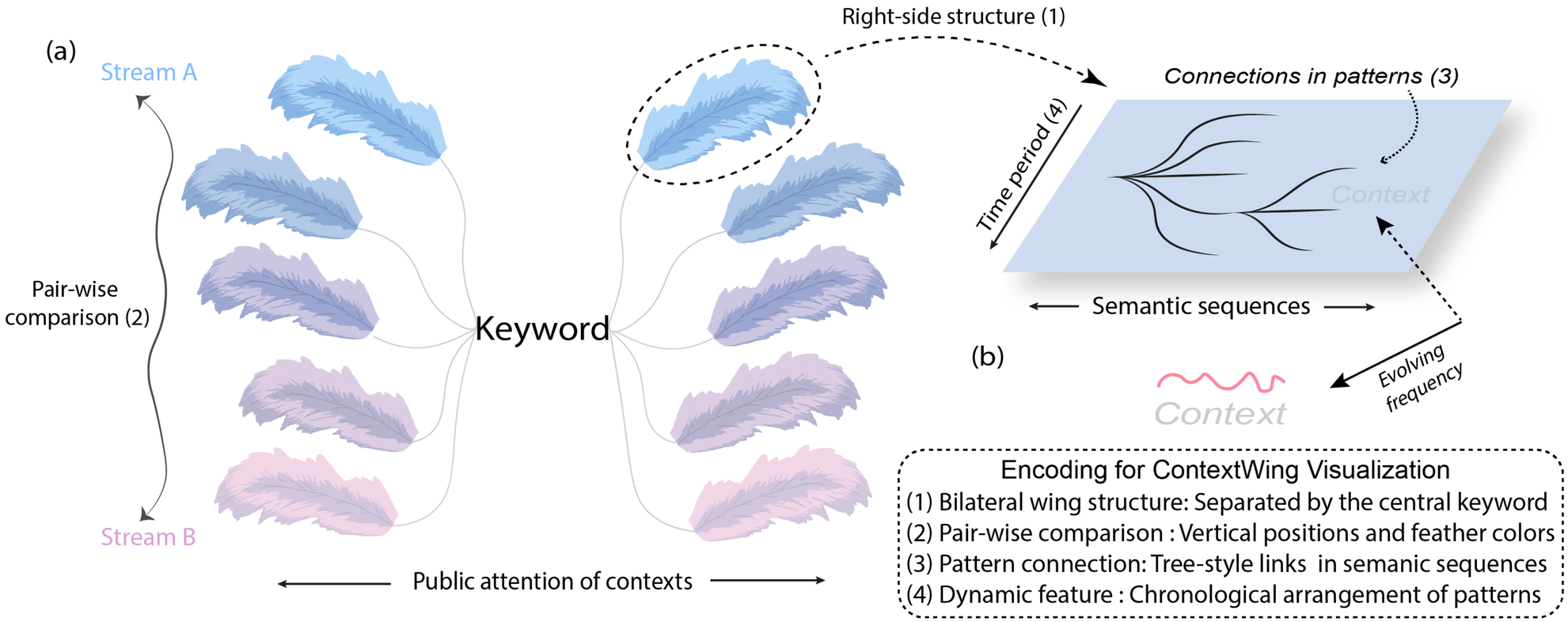
There are few visualizations that focus on context sequence,

most of which combine some simple word cloud visualizations,

and do not support：

sequence, pair-wise data streams and time at the same time.

Visual Design



Encoding for ContextWing Visualization

(1) Bilateral wing structure: Separated by the central keyword
(2) Pair-wise comparison : Vertical positions and feather colors
(3) Pattern connection: Tree-style links in semanic sequences
(4) Dynamic feature : Chronological arrangement of patterns

Overview

Analysis Tasks

Modeling

Visual analysis

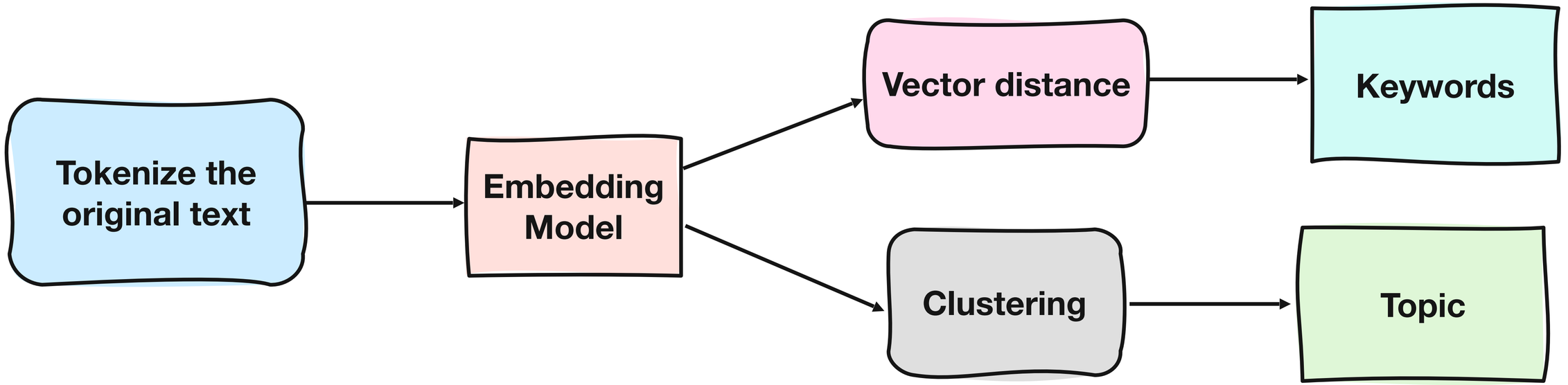
Target Users

- Computing semantics such as topics and patterns for analysis.

- Analyzing temporal information to know how the discussions change.
- Comparing semantic information to know the difference between discussions.
- Pair-wisely comparing the semantics in two data streams.
- Analyzing under streaming setting.

Modeling

1. Topic Modeling



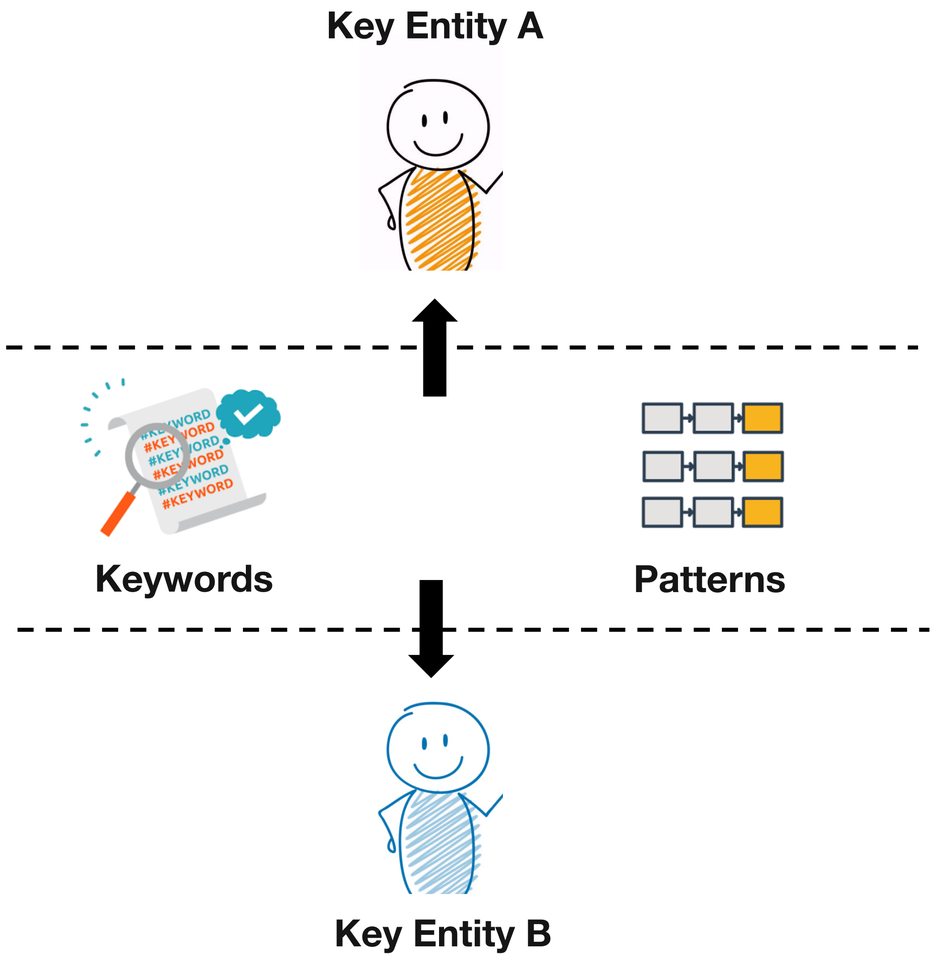
Modeling

2. Pair-wise correlation

Definition:

Correlation of word i to the key entities (j_0, j_1), denote as C_i^t and $C_i^t \in [0,1]$.

$$C_i^t = \frac{r\left(\beta_{ij_0}^t - \beta_{ij_1}^t\right)}{N^t}, \quad \forall i \in W^t$$

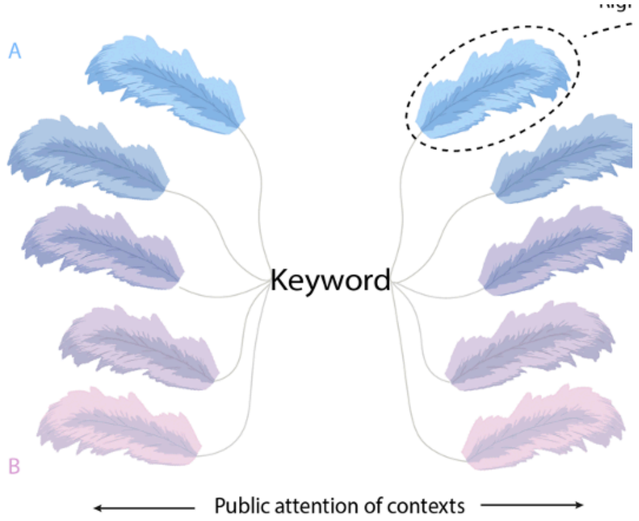


Modeling

3. Public Attention

The idea is derived from **retweets number** and **co-occurrence of two keywords**.

$$\mathscr{A}(c,k) = Log\left[\frac{\sum_{i=1}^{n} u_i(c,k)\eta_i \cdot \sum_{i=1}^{n} u_i(c,k)r_i \,/\sum_{i=1}^{n} u_i(c,k)}{\sum_{i=1}^{n} u_i(c,-k)\eta_i \cdot \sum_{i=1}^{n} u_i(c,-k)r_i \,/\sum_{i=1}^{n} u_i(c,-k)}\right]$$



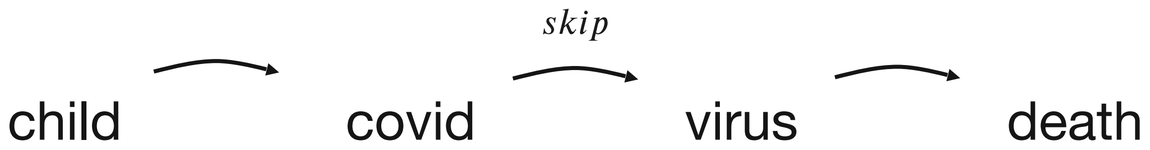
Contextual keyword

Distance

Central keyword

Modeling

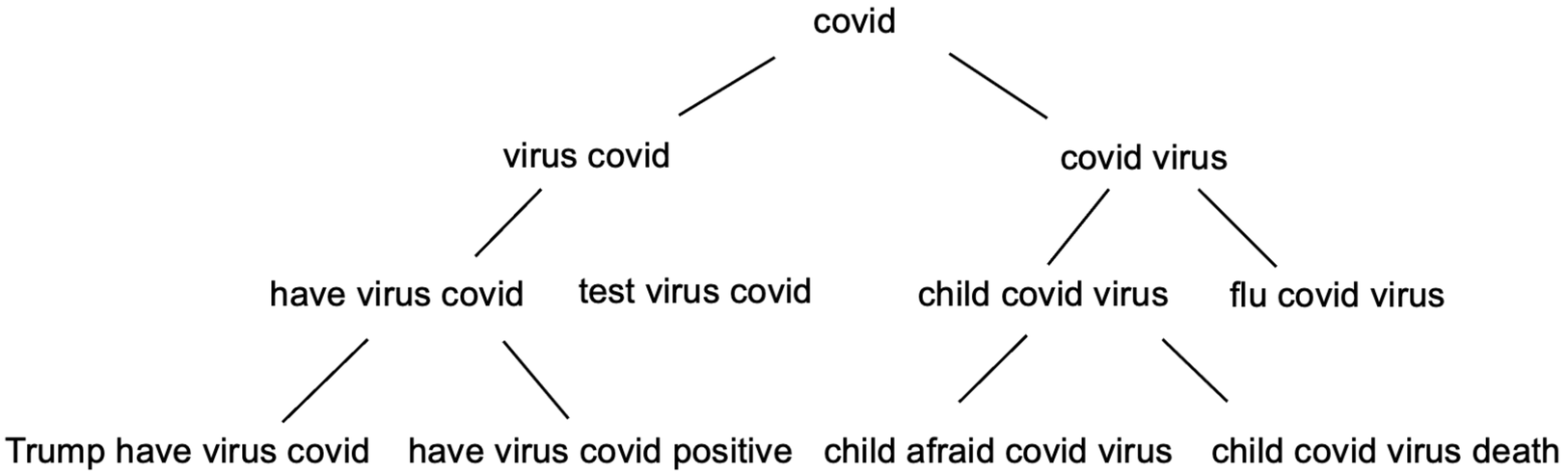
4. Pattern Generation



An example of **pattern generation process**:

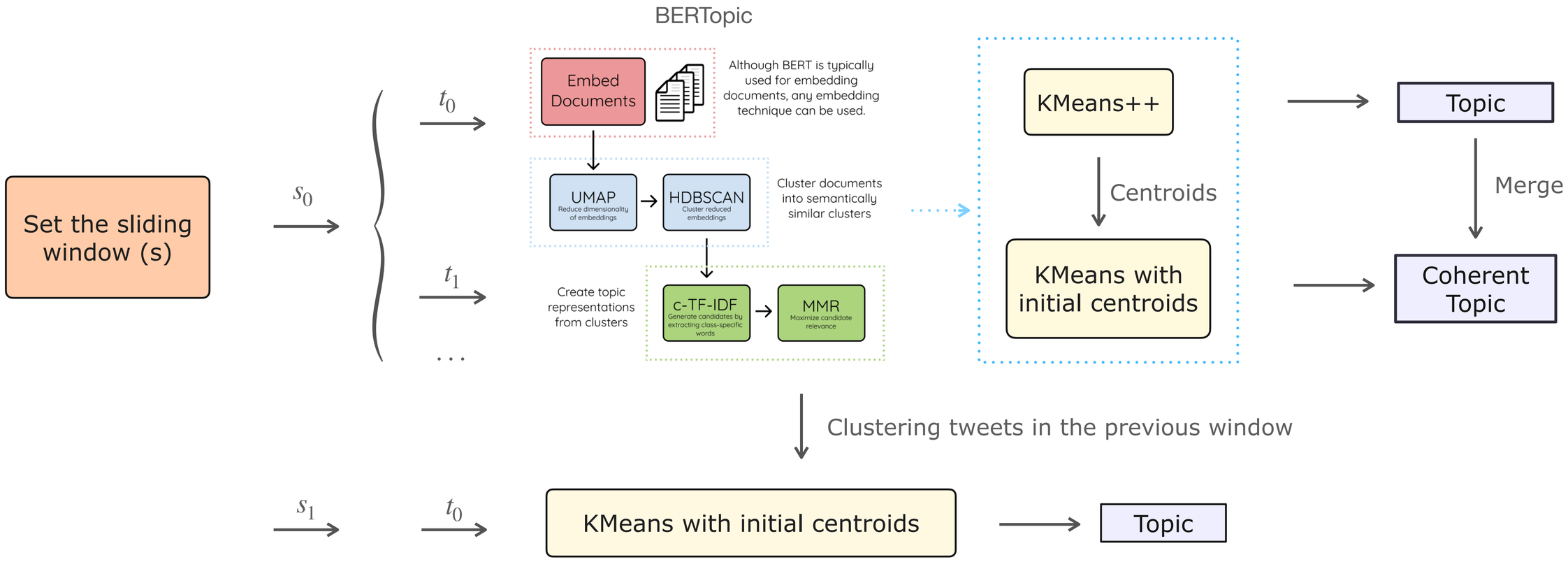
Select top k patterns by rank co-occurrence frequency

$$\eta_P^t = Rank([\alpha_{P_1}, \alpha_{P_2}, ..., \alpha_{P_n}], k)$$



Modeling

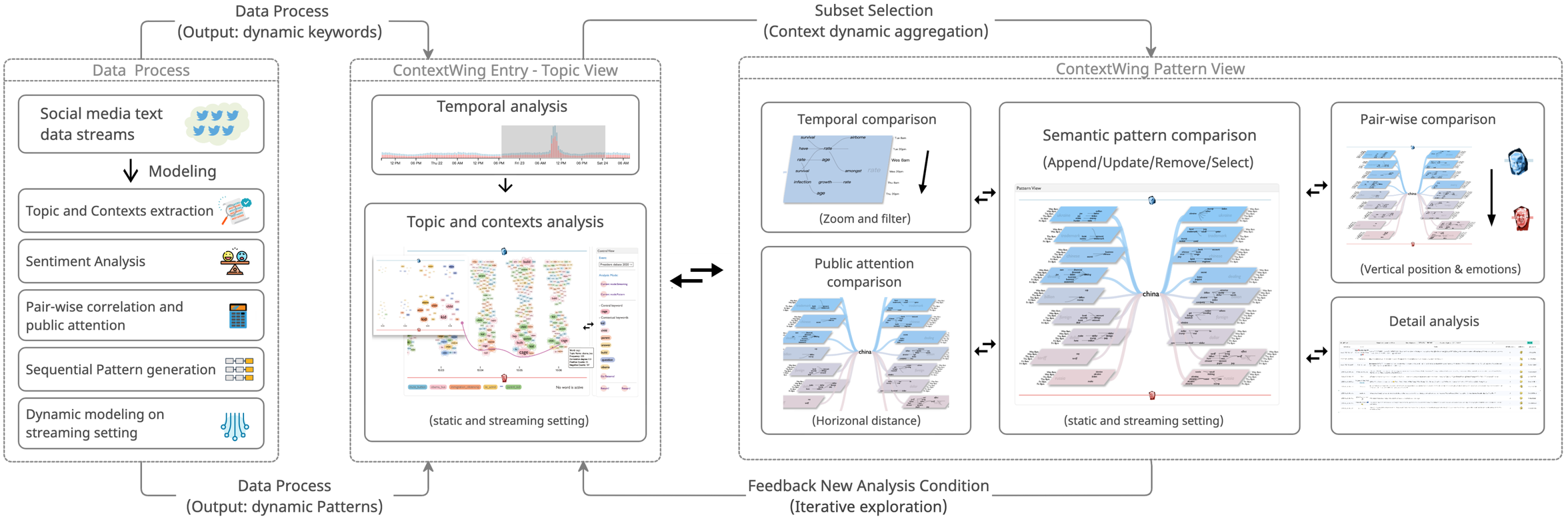
5. Streaming analysis - Topic modeling



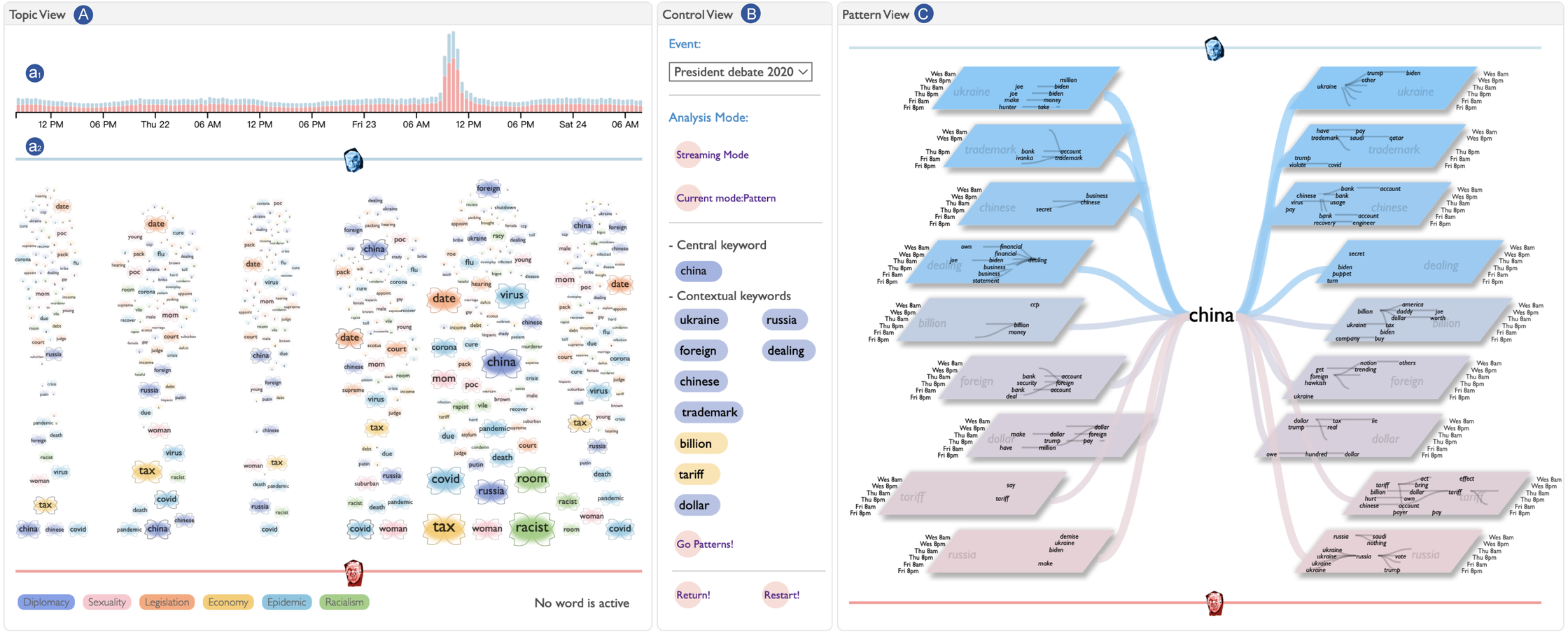
Grootendorst, Maarten. "BERTopic: Neural topic modeling with a class-based TF-IDF procedure." arXiv preprint arXiv:2203.05794 (2022).

System Overview

System architecture



System Overview



Detail View D

Results delivered in 74 ms

Time Options: 2020-10-21 20:00:00

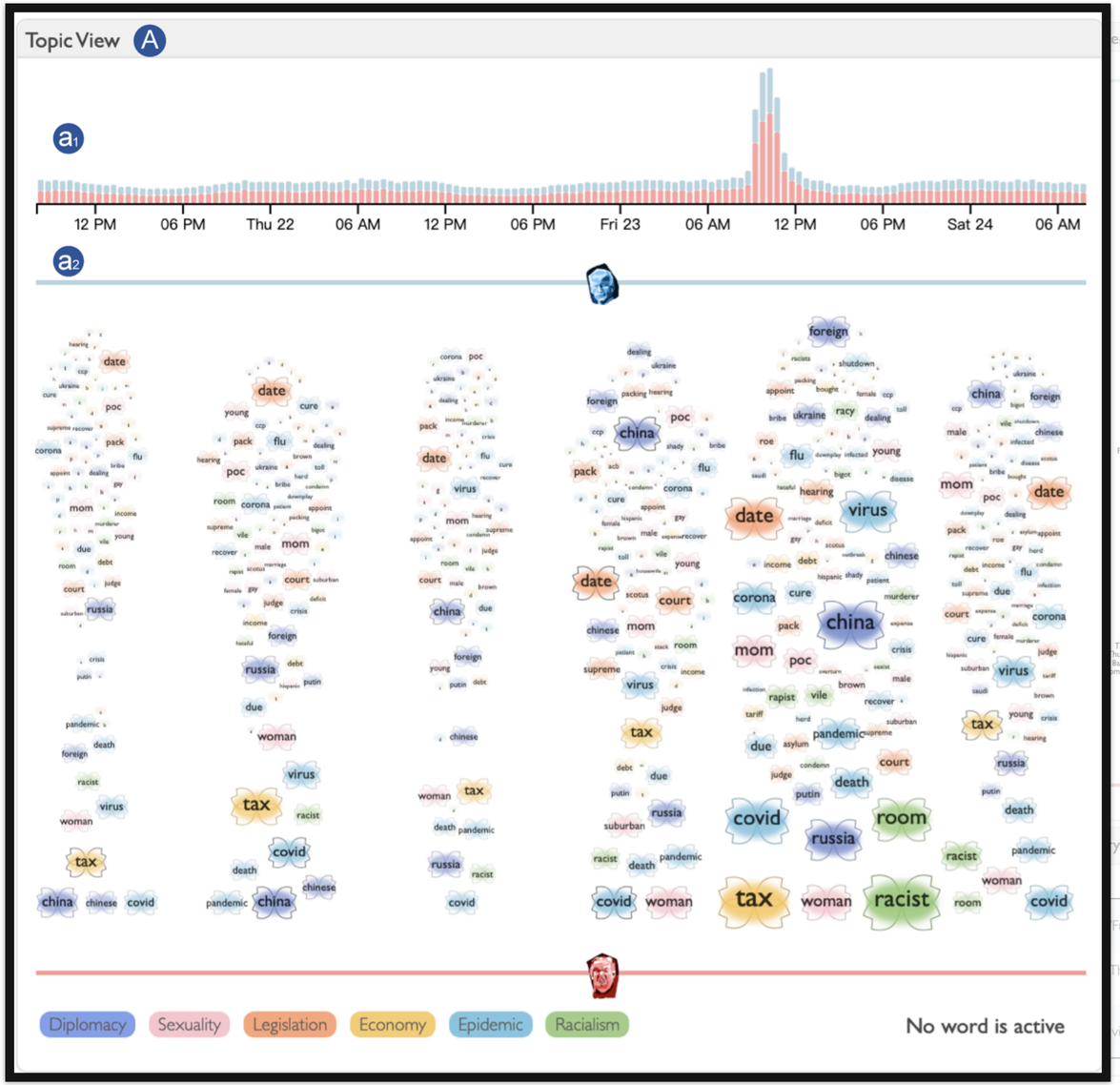
Keyword Query: joe biden dealing china

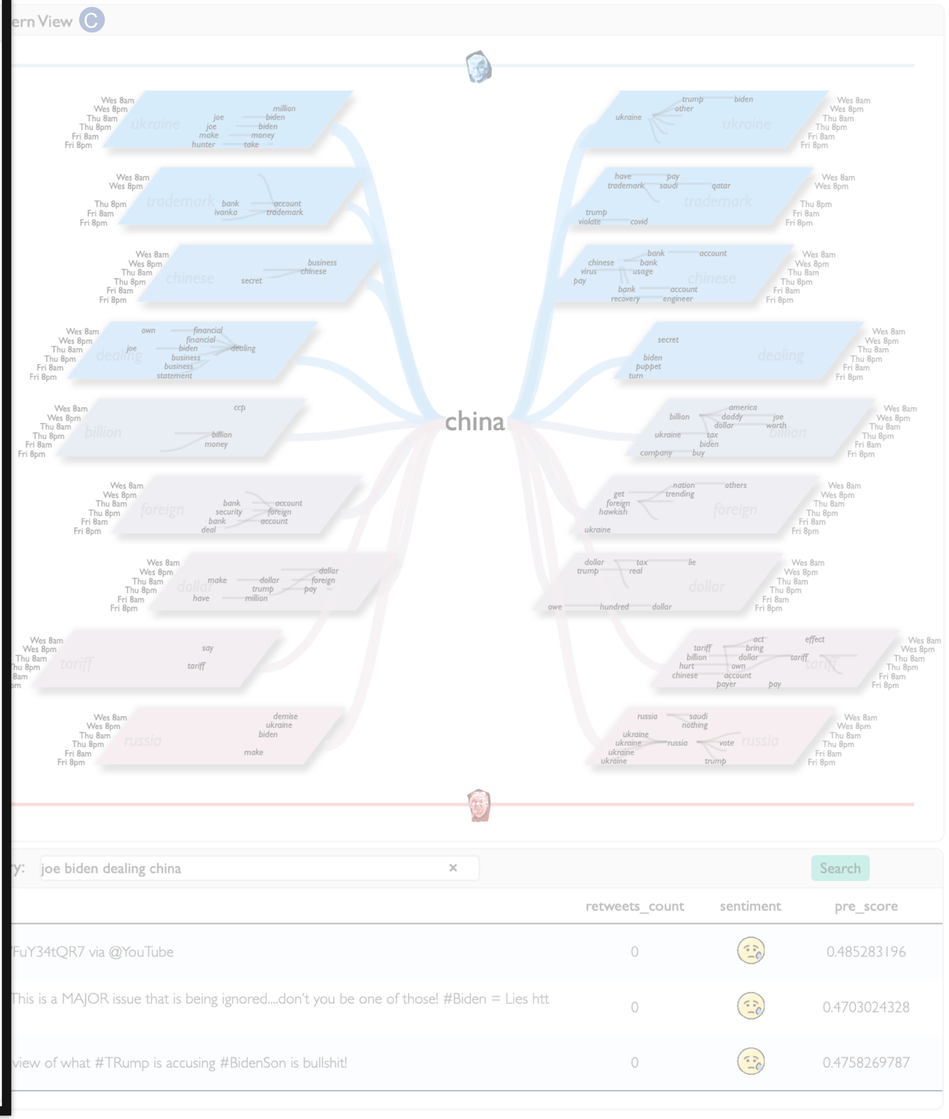
Search

created_at	name	tweet	retweets_count	sentiment	pre_score
2020/10/22 00:23:22	Janeen Calaway	#TrumpIsALier #Biden #UNITE #VoteEarly What Trump's Taxes Reveal About His Dealings In China | Morning Joe | MSNBC https://t.co/VFuY34tQR7 via @YouTube	0		0.485283196
2020/10/22 03:19:42	Ted Okerson MD FACE	Joe Biden has an obligation to answer questions about his son's influence-peddling and his own financial dealings—notably regarding China.This is a MAJOR issue that is being ignored....don't you be one of those! #Biden = Lies https://t.co/2xOEMBQZK0	0		0.4703024328
2020/10/22 03:48:05	icjr	What Trump's Taxes Reveal About His Dealings In China | Morning Joe | MSNBC https://t.co/sFU7waCU7Z via @YouTube Showing the full view of what #TRump is accusing #BidenSon is bullshit!	0		0.4758269787

System Overview

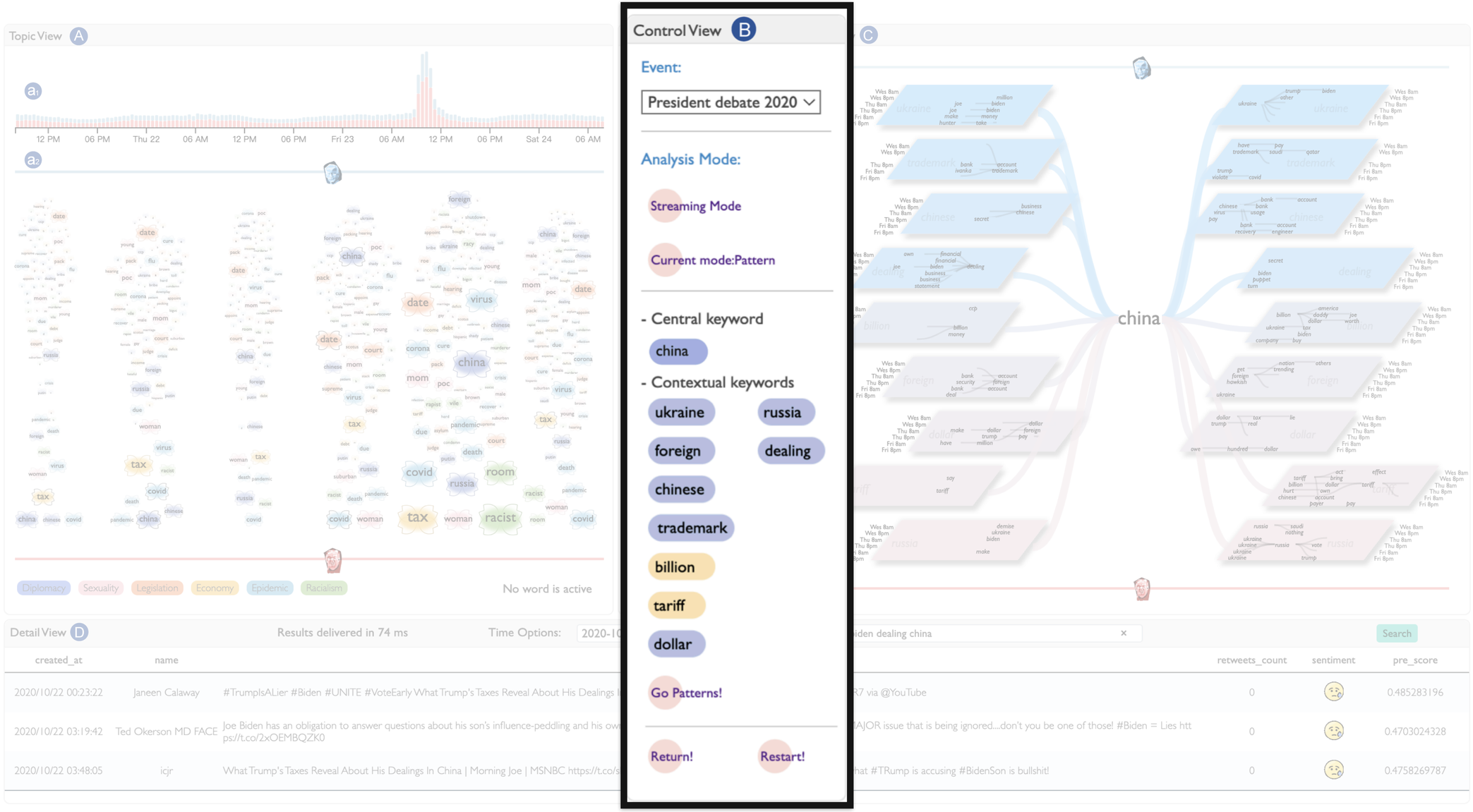
A. Topic view





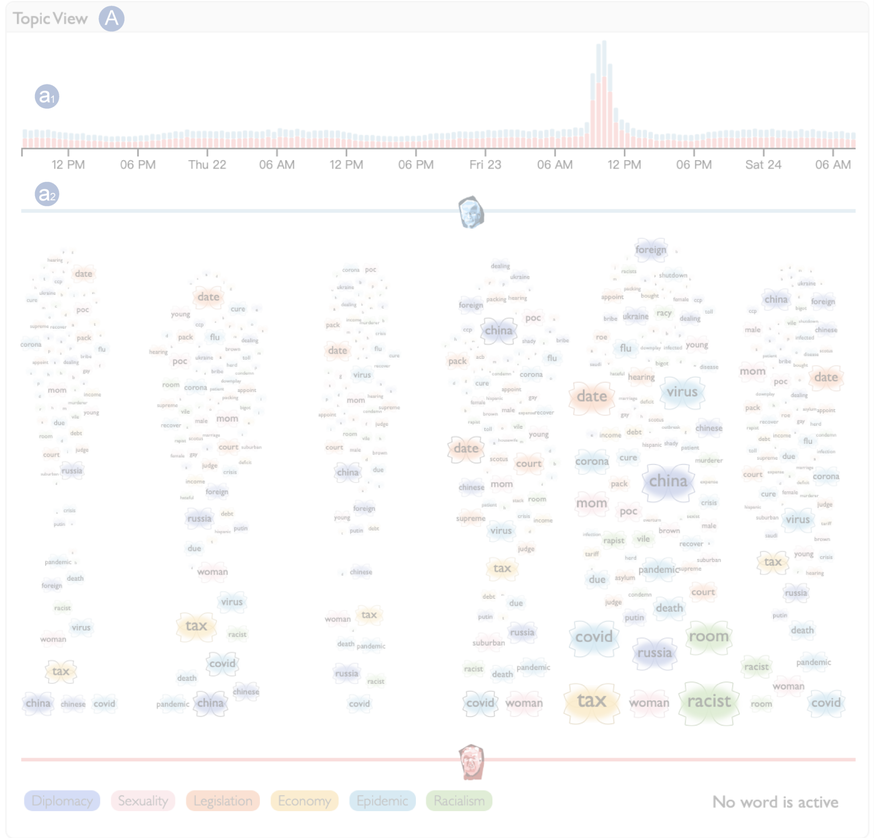
System Overview

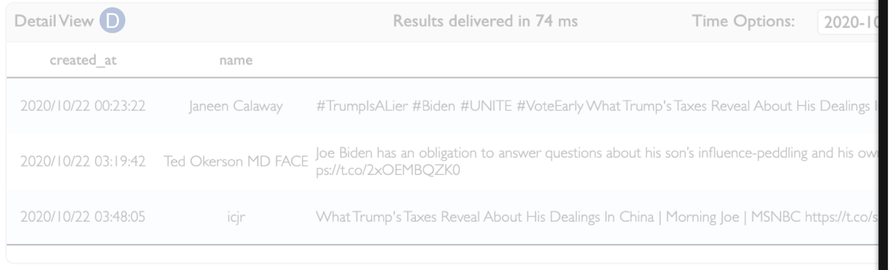
B. Control view

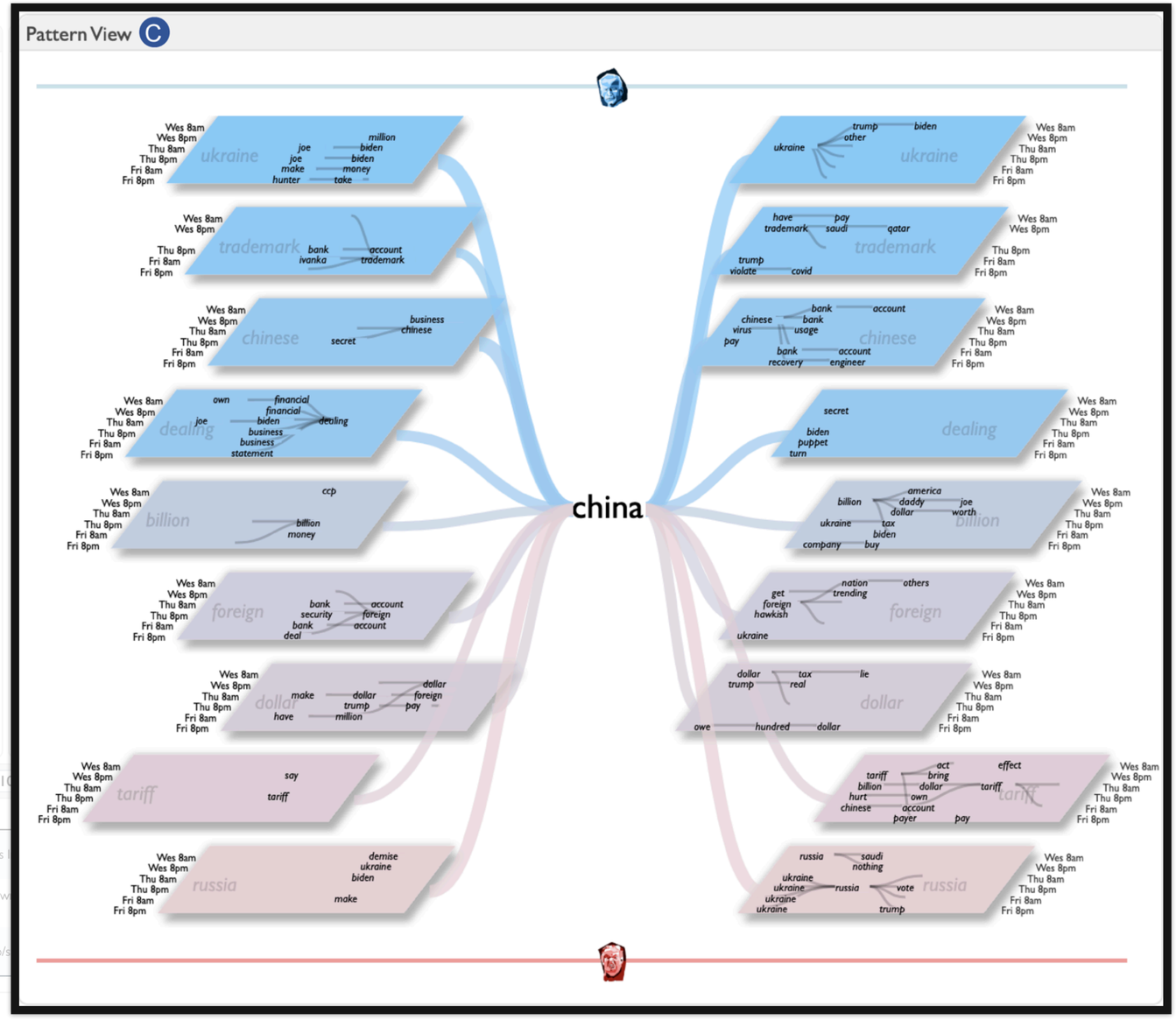


System Overview

C. Pattern view

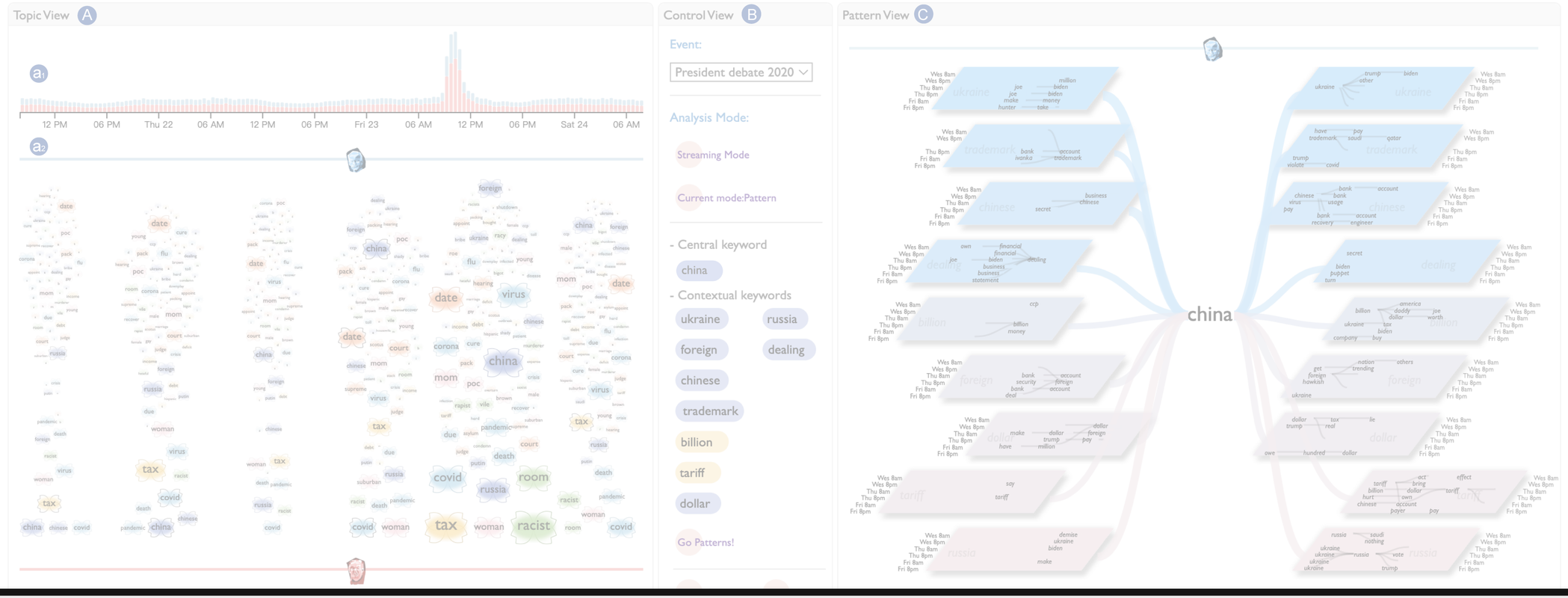






System Overview

D. Detail view



Detail View D

Results delivered in 74 ms

Time Options: 2020-10-21 20:00:00

Keyword Query: joe biden dealing china

Search

created_at	name	tweet	retweets_count	sentiment	pre_score
2020/10/22 00:23:22	Janeen Calaway	#TrumpIsALier #Biden #UNITE #VoteEarly What Trump's Taxes Reveal About His Dealings In China | Morning Joe | MSNBC https://t.co/VFuY34tQR7 via @YouTube	0	😢	0.485283196
2020/10/22 03:19:42	Ted Okerson MD FACE	Joe Biden has an obligation to answer questions about his son's influence-peddling and his own financial dealings—notably regarding China.This is a MAJOR issue that is being ignored....don't you be one of those! #Biden = Lies https://t.co/2xOEMBQZK0	0	😢	0.4703024328
2020/10/22 03:48:05	icjr	What Trump's Taxes Reveal About His Dealings In China | Morning Joe | MSNBC https://t.co/sFU7waCU7Z via @YouTube Showing the full view of what #TRump is accusing #BidenSon is bullshit!	0	😢	0.4758269787

Evaluation

Case study

- **The analysis workflow**
  - The 2020 U.S. Presidential debate
  - About Presidential assets

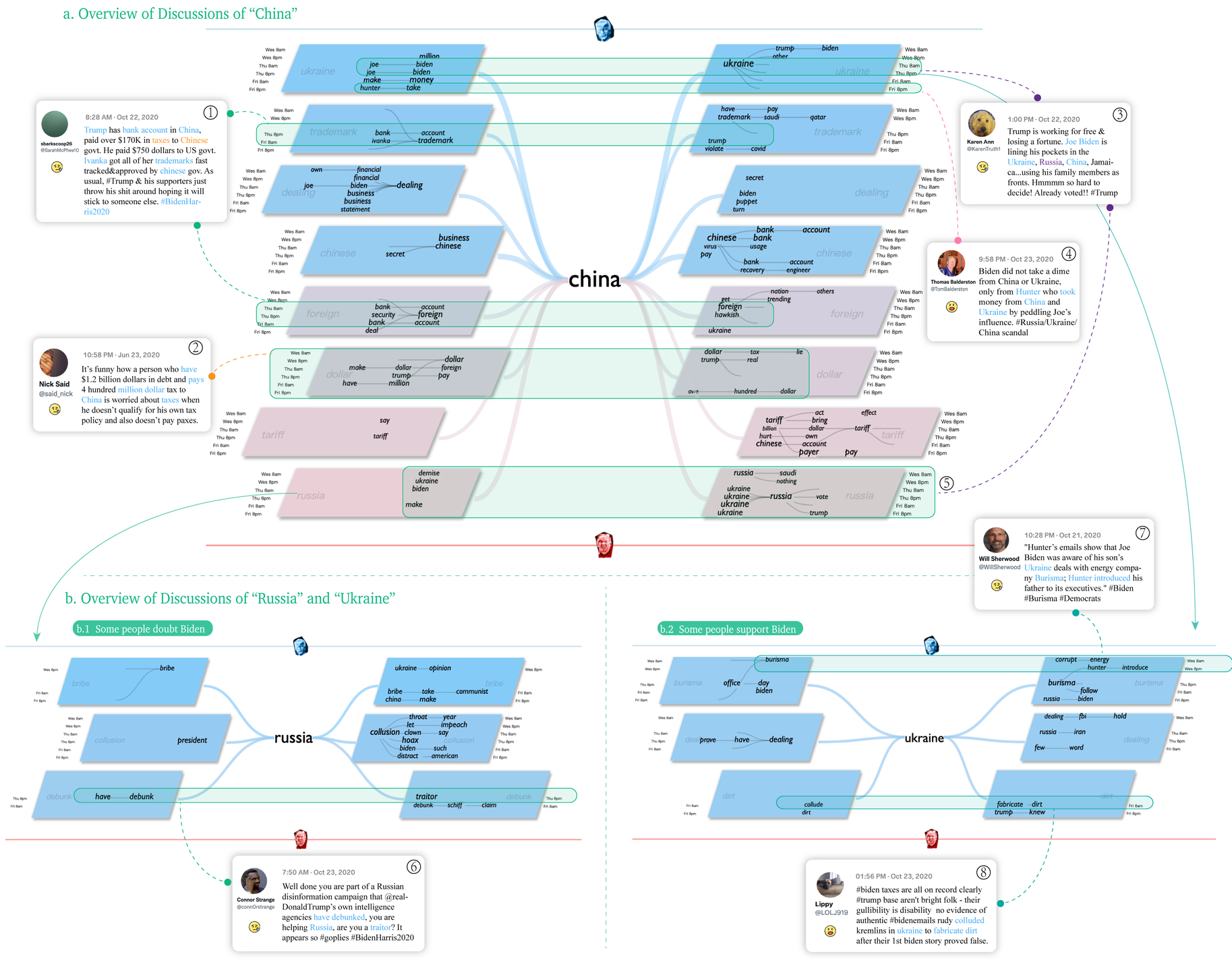
- **The streaming analysis**
  - The 2020 U.S. Presidential debate
  - About “Who built the cages?”

Evaluation

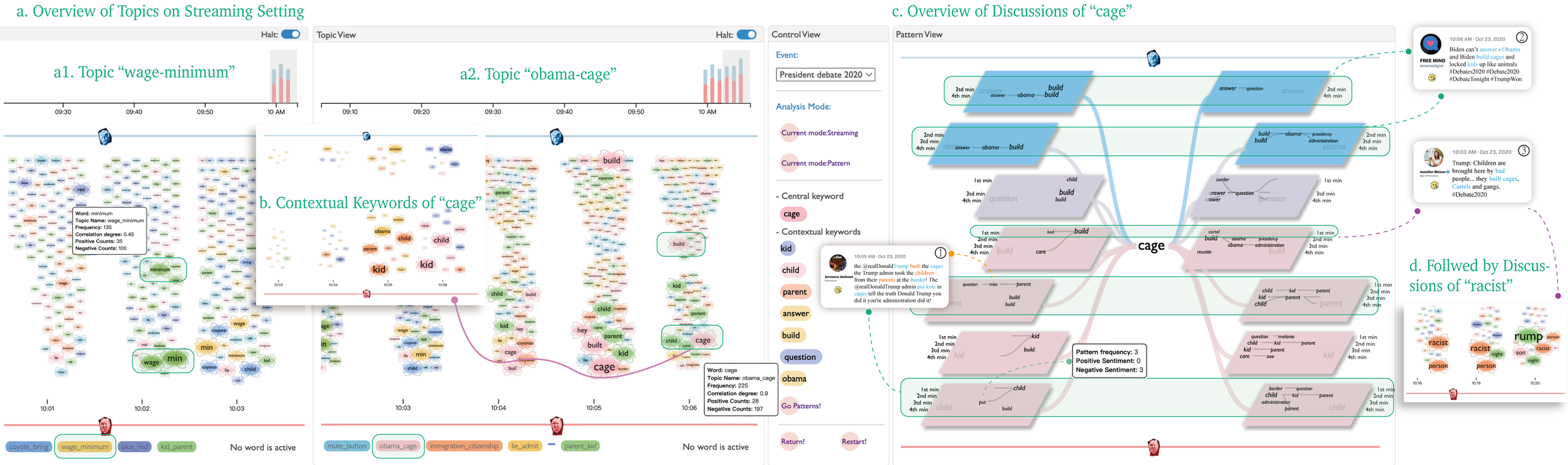
Case study 1

The 2020 U.S presidential election debate.

Personal assets and trades of two candidates.



Case 2 - Streaming setting

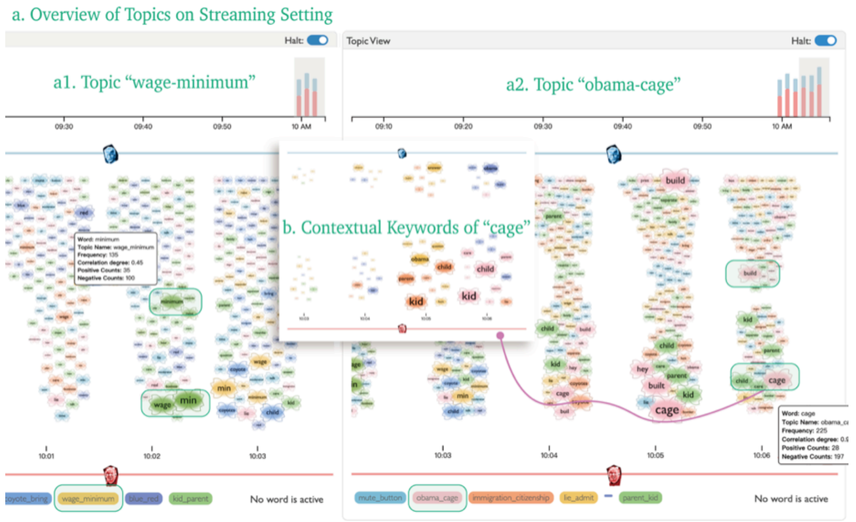


Discussion

Implications

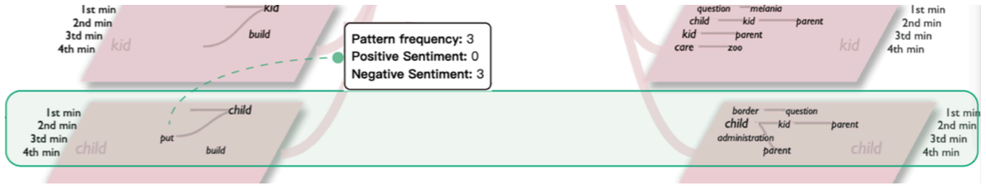
💡**The streaming analysis is needed.**

- Consider the practicability in real cases



💡**Sentiment adds contexts.**

- Not only “relating to ", reveals attitudes is also important for better understanding.



Discussion

Limitations

The number of data streams.

- Multiple streams: adjusting the computational model and rearranging the design of semantic sequence patterns intuitively.

2 or more?

The setting of parameters.

- More users customized, freedom exploration

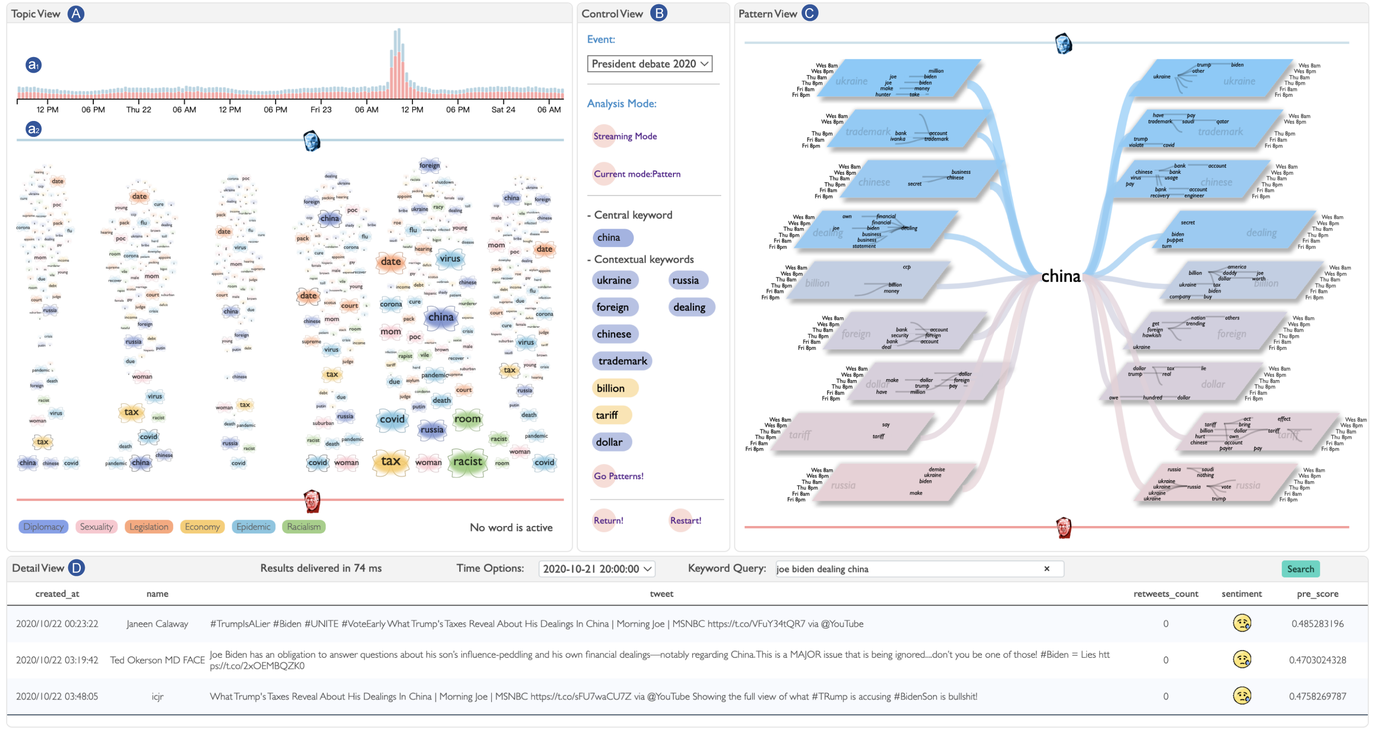
More/ less Keywords in the sequence?

The time granularity.

- This requires online re-aggregation and computing and is worth further studying in the future.

Time interval

Conclusion



- A **novel visual metaphor** for pairwise visual comparison.
- A **visual analysis system that**
  - supports exploration of patterns of contexts from semantics and temporal aspects.
  - supports analysis on both static and streaming settings.

Thanks for your listening!

Welcome to our website: http://fduvis.net/

Email: yuhengzhaocn@gmail.com