

Visual Analytics for Emerging Data Intensive Problems

Kwan-Liu Ma
University of California at Davis

ICRI 2023, Academia Sinica

Visual Analytics

is the science of analytics reasoning
guided by
statistical analysis, machine learning
and interactive visualization

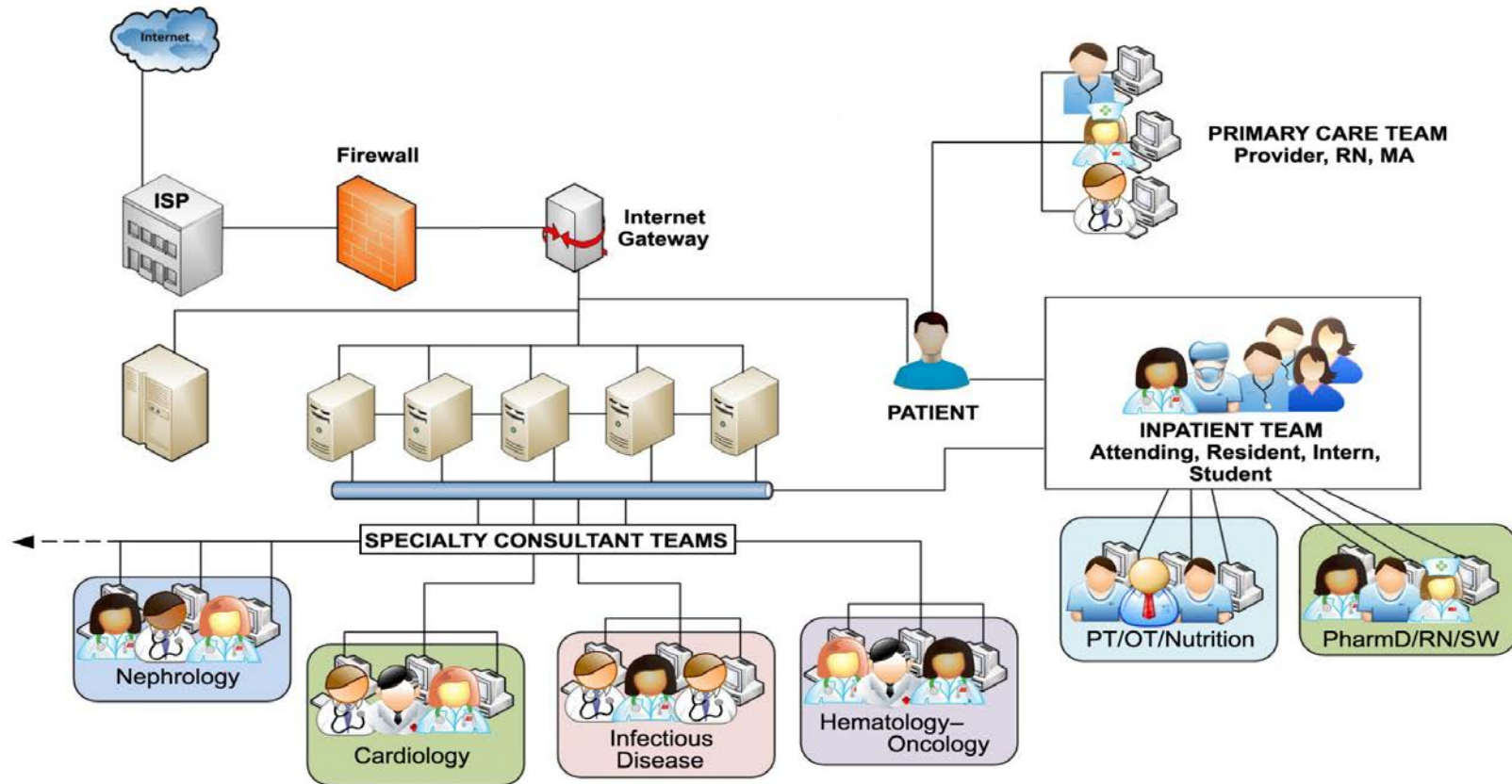
Outline

Visual analytics to enhance:

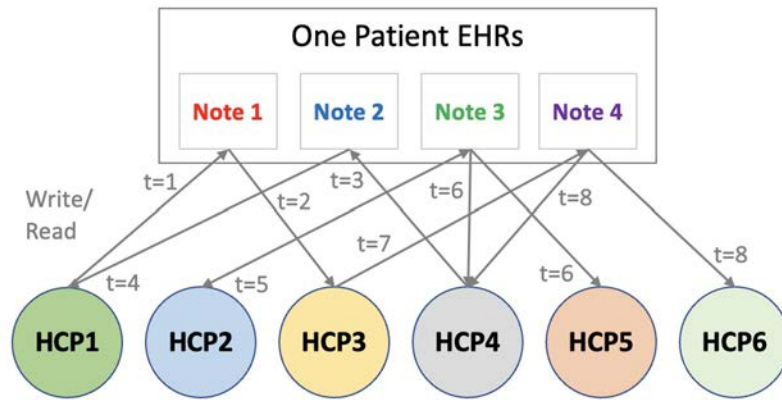
- Teamwork for healthcare
- Operation & performance of extreme-scale systems
- Comprehension of large collection of documents

**EHR Data Analytics
to Enhance Patient Care Multiteam Systems**

Multiteam Systems for Healthcare

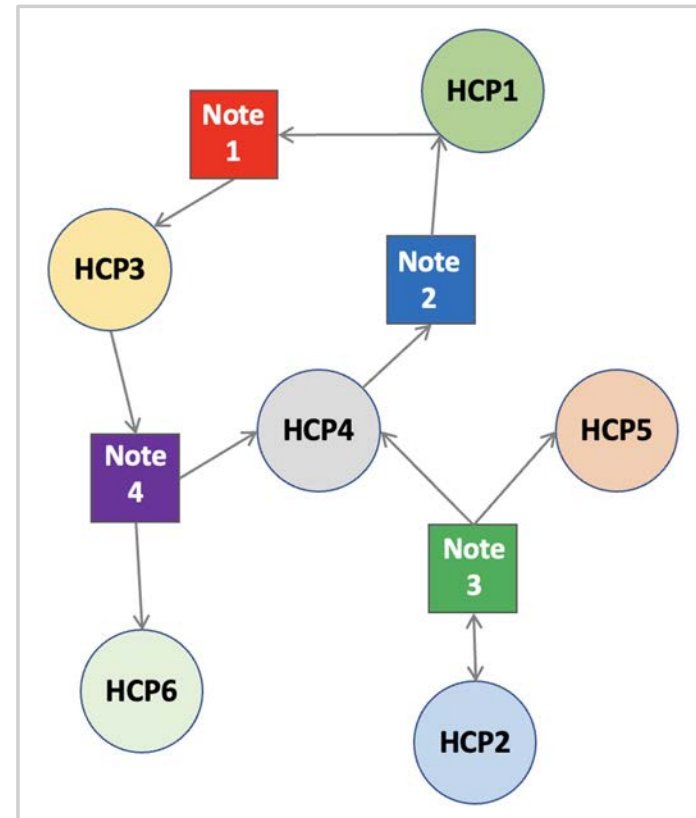


From Access Log to Networks



Stored information:

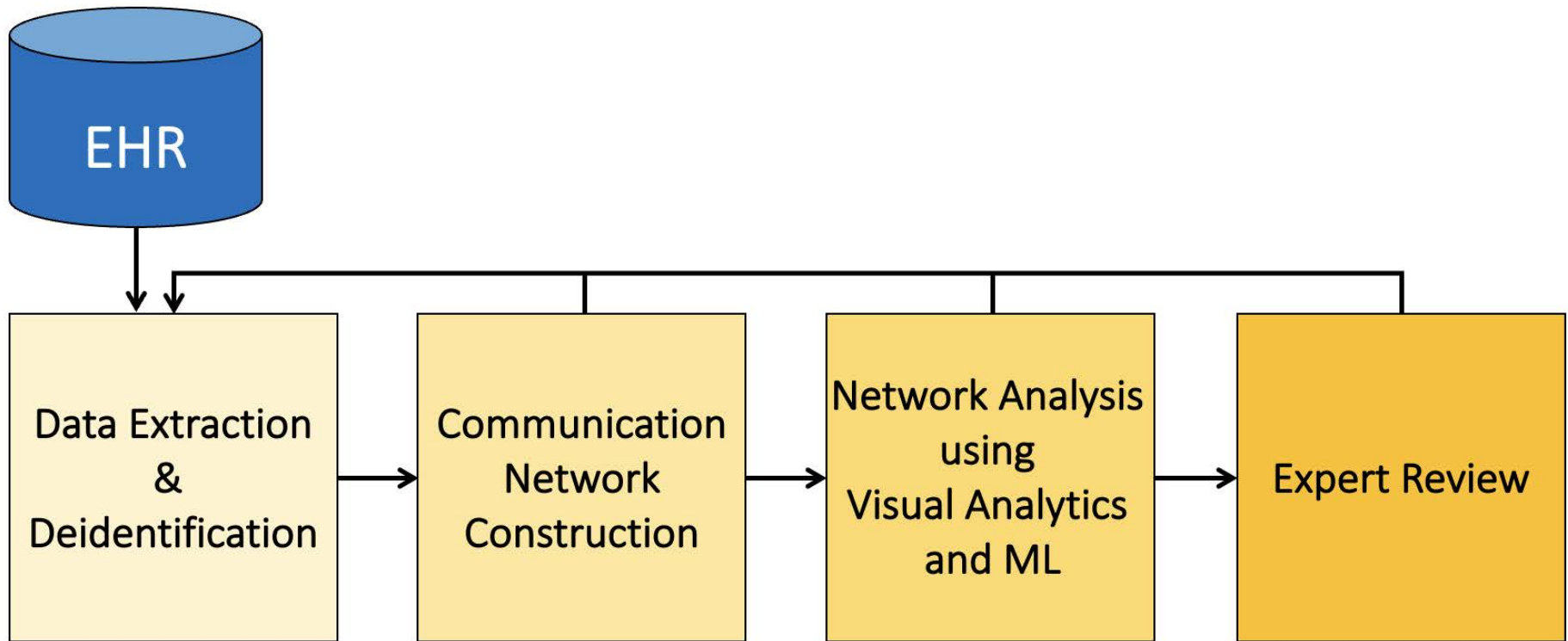
1. Healthcare Professionals (HCPs)
2. Notes
3. Access timestamps
4. Direction of interactions



ML Assisted Visual Analytics

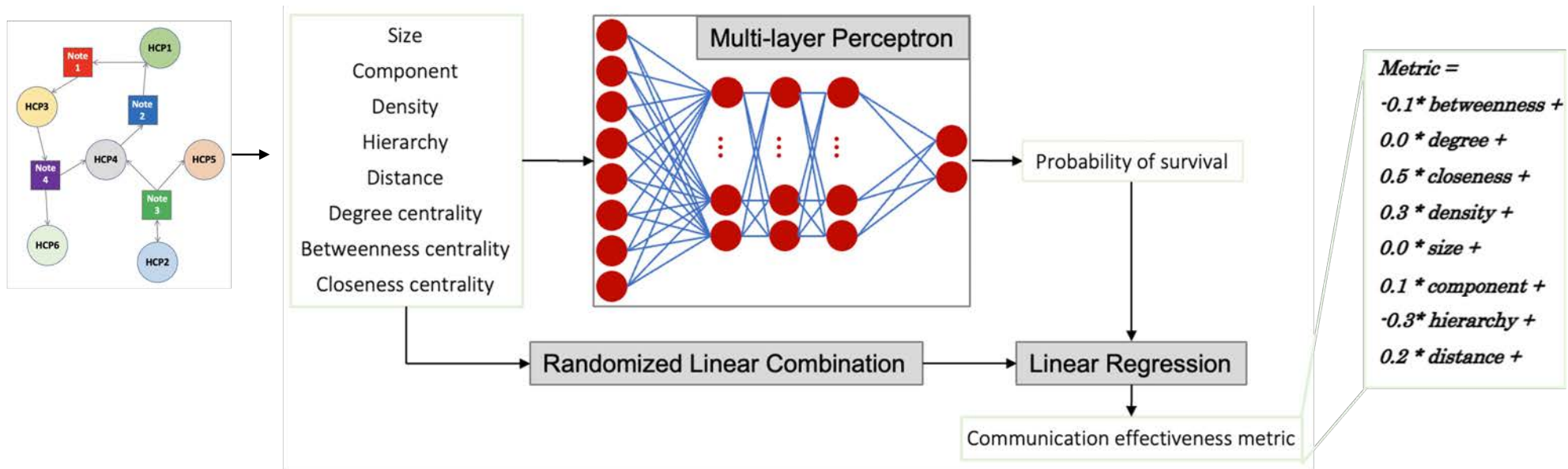
- **Informed by:**
 - Communication network theories: structure, position, and flow
 - Systems engineering principles: workflow, value, waste (time, cognition)
- **Designed to assist HCPs with visualization guided information-processing to:**
 - quickly browse an overview of their patients' medical care
 - identify MTS members involved at specific points and periods in time
 - drill down to discover important details (i.e., notes, messages, reports, etc.)
 - see who has accessed/reviewed specific EHR documentations
 - efficiently rectify gaps in information-sharing by members of the MTS

Workflow



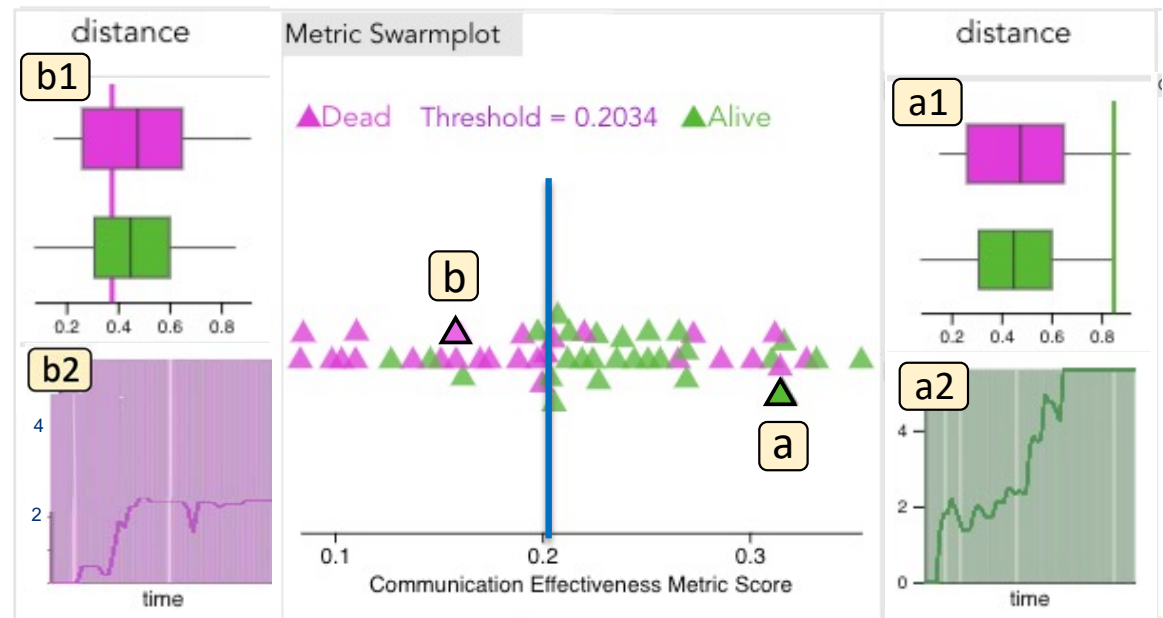
Case Study I

- Patient Group: 53 patients, between age 65-75 with Stage 3 lung cancer, in which 27 survived and 26 did not.
- Communication effectiveness metrics are weighted combinations of multiple network measures (e.g., degree, distance, closeness, betweenness, etc.)

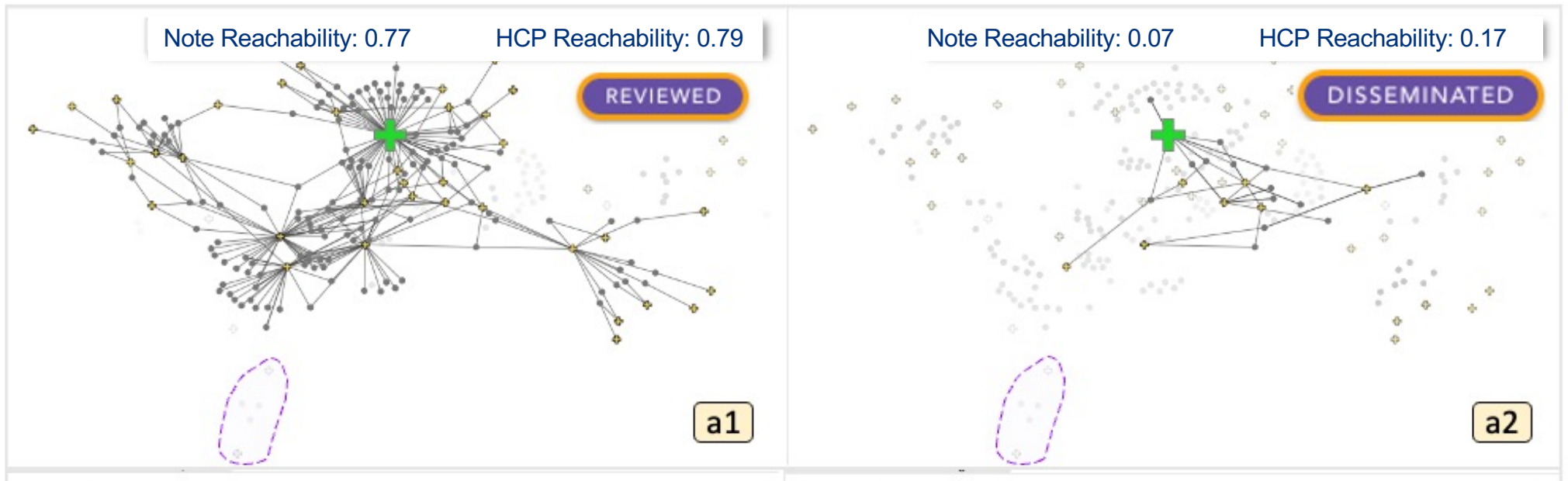


Case Study I

- Patient Group: 53 patients, between age 65-75 with Stage 3 lung cancer, in which 27 survived and 26 did not.
- Communication effectiveness metrics are weighted combinations of multiple network measures (e.g., degree, distance, closeness, betweenness, etc.)
- The distance measure is the average number of directed edges between every *HCP* node and *note* node in a network.
- A network with a higher distance value suggests the *HCPs* in the network are more connected.



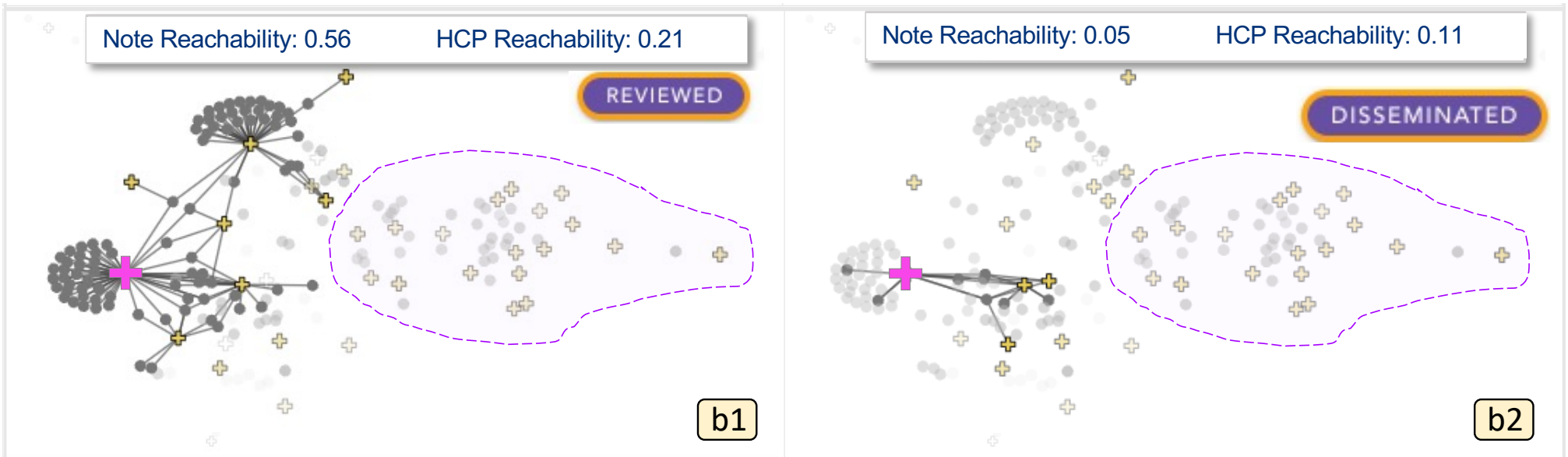
Information Reachability



Information reachability visualization for a patient who survived



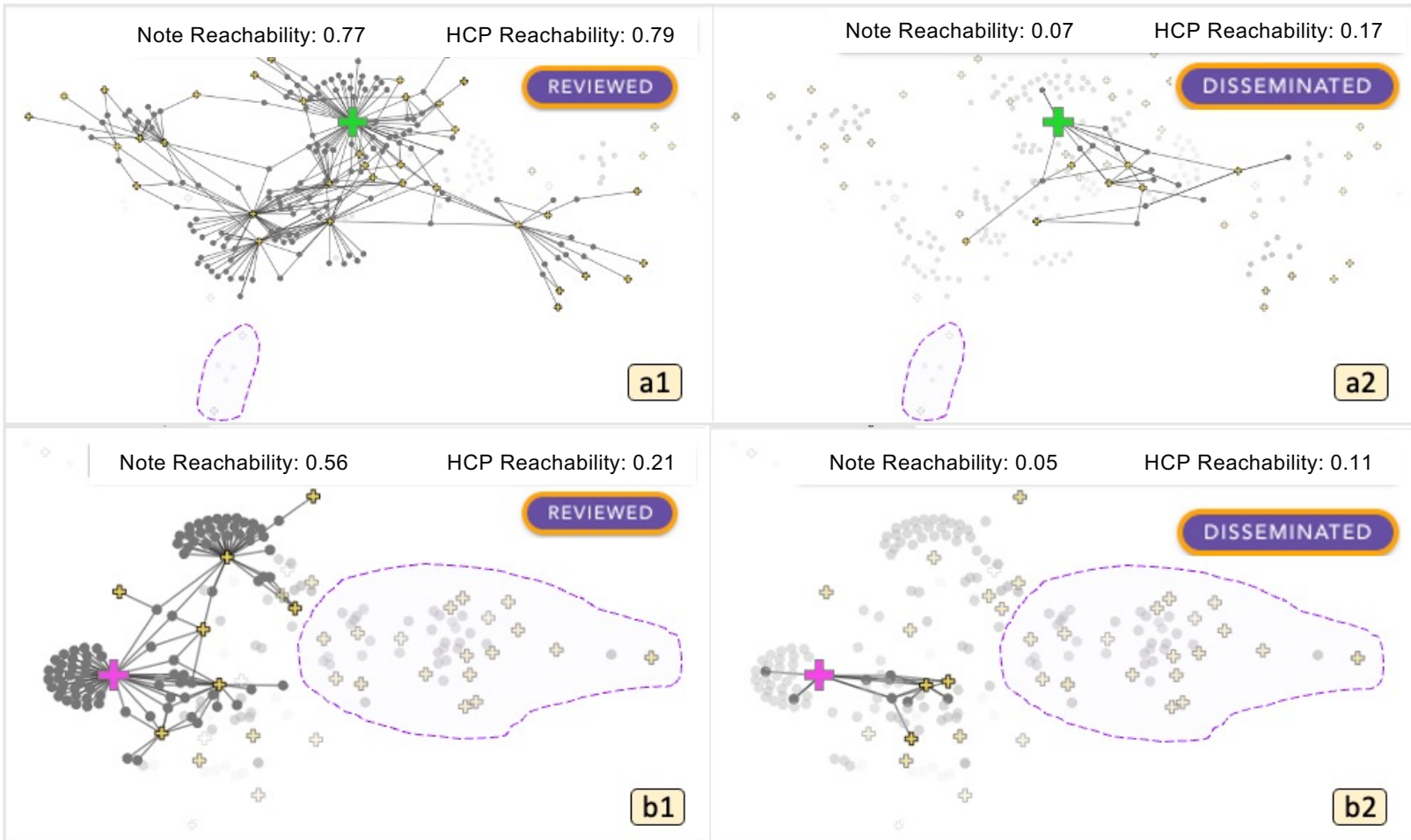
Information Reachability



Information reachability visualization for a patient who did not survive

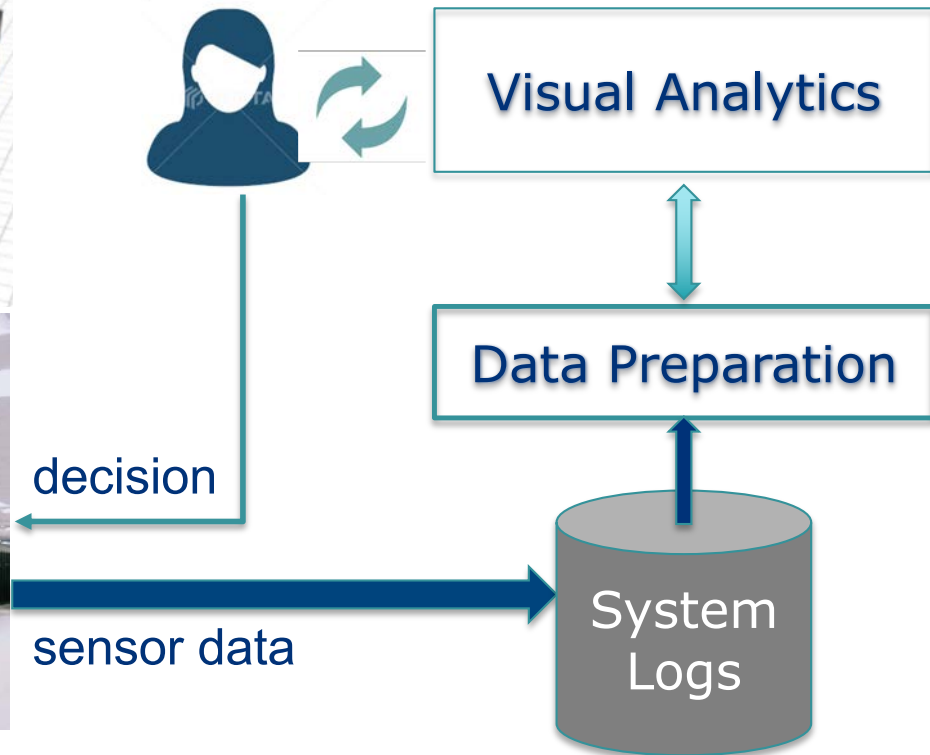


Information Reachability



Streaming Data Analytics for Hardware System Monitoring

Monitoring Supercomputing Systems



Supercomputer Hardware Logs

- Streaming, multivariate, time series
- A representative dataset
 - Extracted from the K Computer, Riken, Japan
 - 864 Racks
 - 390 temperature and 480 voltage readings per rack
 - 1162 measures per rack
 - 288 time points per measure (every 5min)
 - Data volume: ~2GB per day



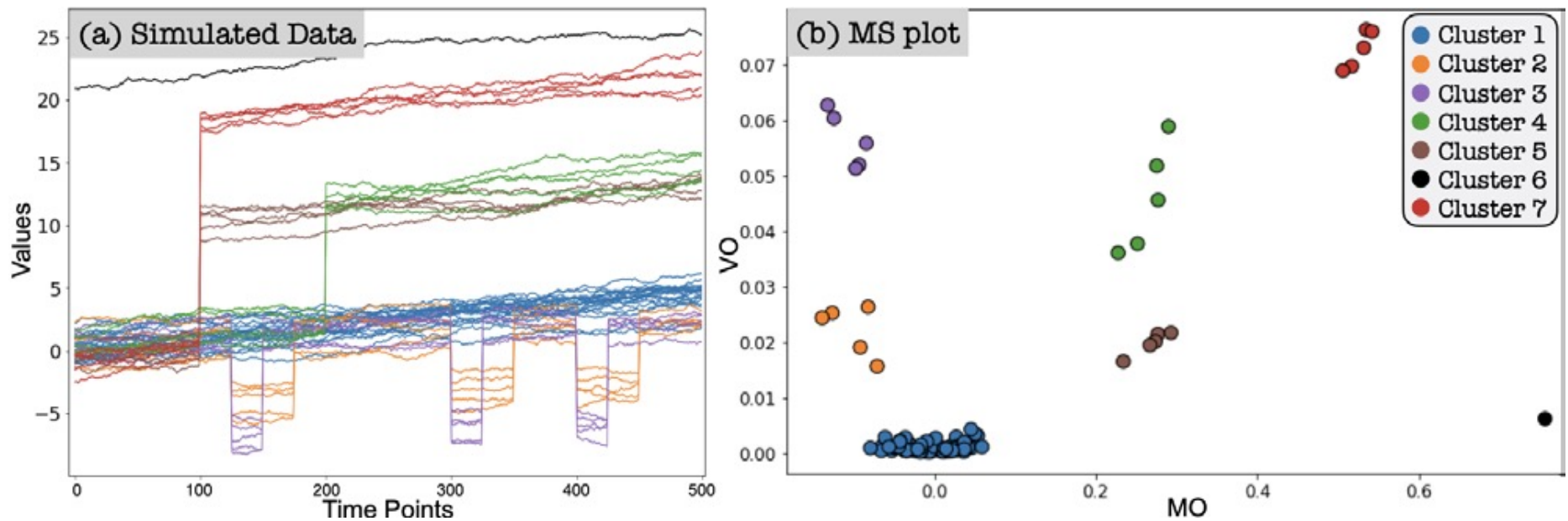
Technical Approach

- Expect to start with select readings and initial time points
- We adopt the magnitude-shape plot (MS plot by [Dai and Genton 2018](#)), which reveals both the functional magnitude and shape outlyingness of time-series data.
- **Incrementally** compute the MS plot with every new time point added to the continuum
- **Progressively** update the MS plot with every new time series added
- Employ FPCA together with MS plots to conduct further analysis of select outliers

Magnitude Shape Plot

- W. Dai and M. Genton, "Multivariate Functional Data Visualization and Outlier Detection", J. Comput. Graph. Stat. 27(4) 2018
- Designed to visualize both the magnitude outlyingness (MO) and shape outlyingness (VO) of multivariate functional data based on measures of directional outlyingness (O)
- The plot depicts how much a time series has a different magnitude and shape with other time series, and thus visually reveals outliers.

Visual Outlier Detection with the Magnitude Shape Plot



- Simulated functional data and the MS plot with magnitude outlyingness (MO) and shape outlyingness (VO) along x-axis and y-axis, respectively.

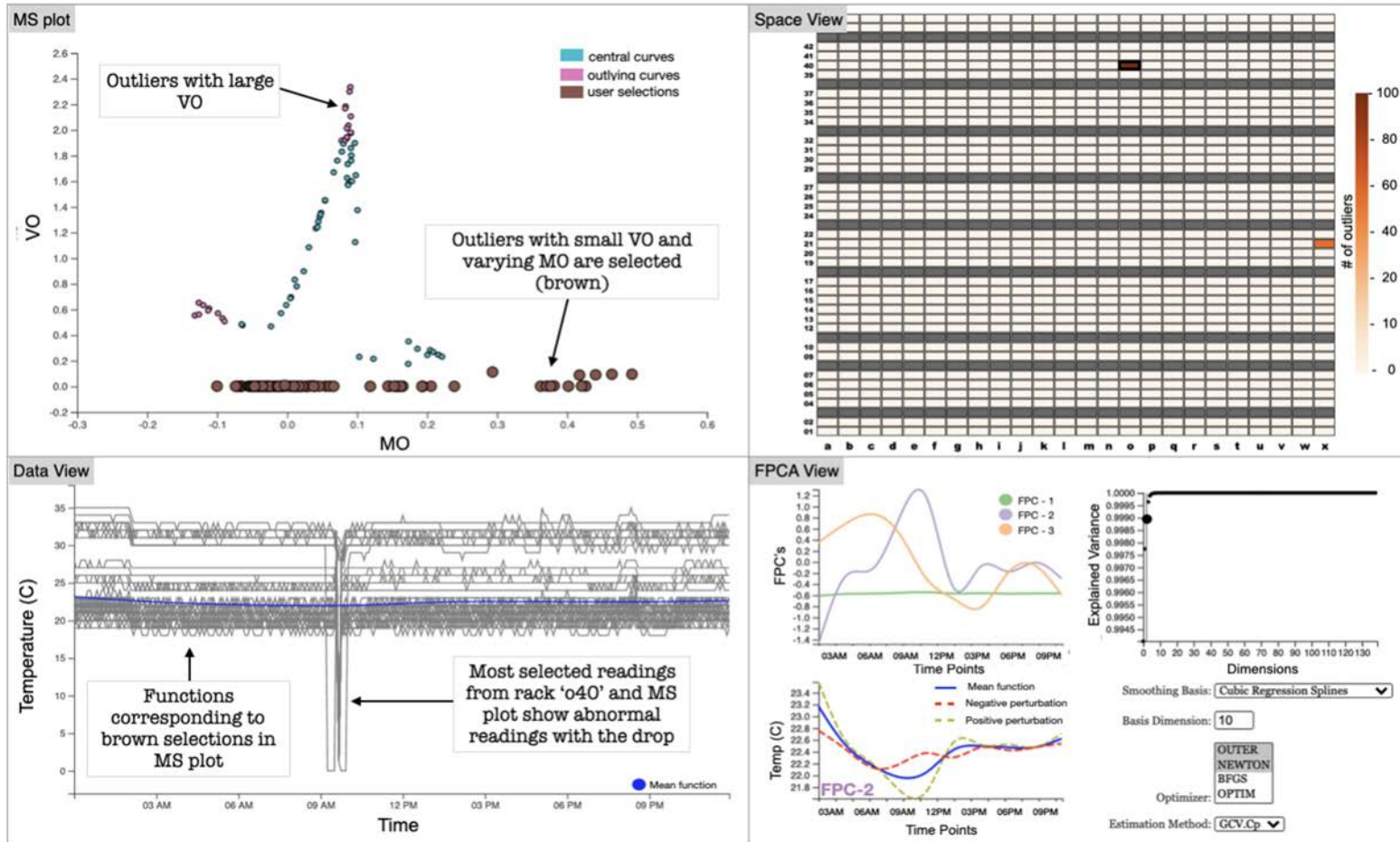
Incremental Update of the MS Plot

- Recomputing the MS Plot when new time points for each series arrive is costly, infeasible for real-time monitoring tasks.
- Incremental update of exact MO and VO!
- $MO^{T+1}(X^{T+1}) = \frac{1}{T+1}(TMO^T(X^T) + O(X^{T+1}[T+1]))$
- The required memory space to save the previous results, MO and FO, for all N time series is $O(N)$.

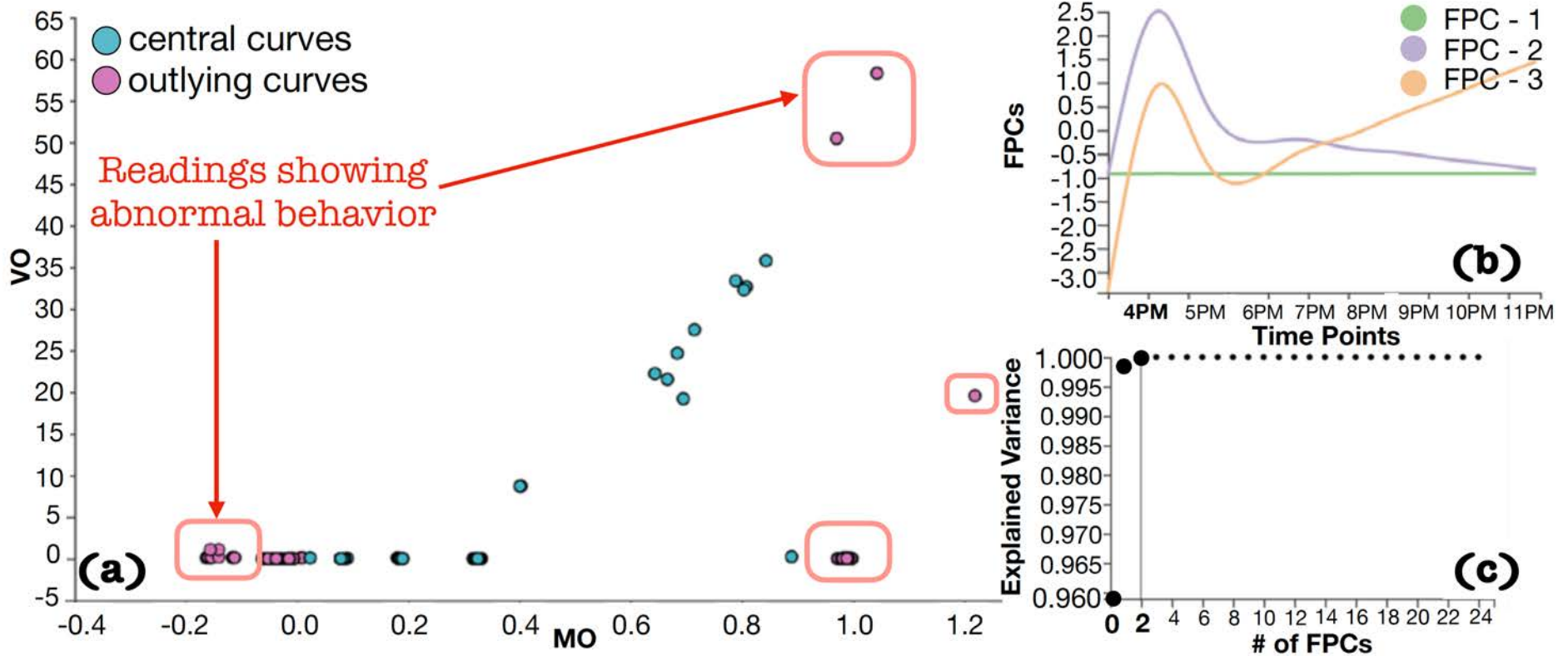
Progressive Update of the MS Plot

- The exact update for a new time series requires recomputation of the measures for all N time series at all time points!
- When the number of time series, N , in a system is large, this recomputation is computationally prohibitive.
- To avoid recomputation, a progressive algorithm is designed to generate the MS plot with **approximated** O .
- Exact update is made only when the KL (Kullback-Leibler) divergence of the mean absolute deviation between the new and original sets of time series becomes larger than a set threshold value.

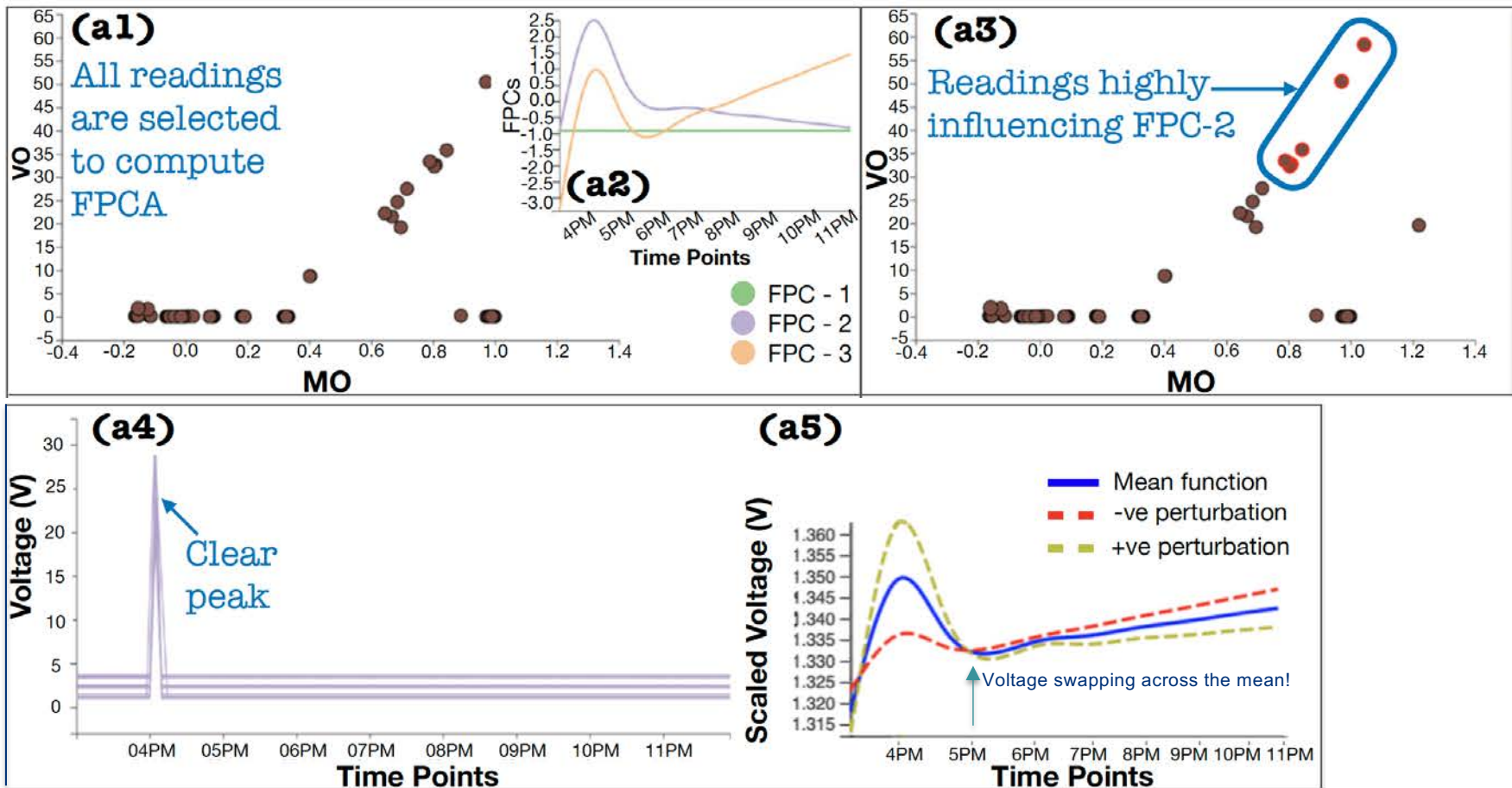
Visual Analytics Interface



Outlier Identification and Validation



Outlier Identification and Validation



**Visual Analytics for
Comprehension of
Large Collections of Documents**

LLM in Document Analysis

- **LLM as Agents**
 - A chatbot that directly answers questions
 - Easy to develop and highly generalizable
 - Lacking transparency & mechanisms to convey uncertainty/confidence
 - No robust, established ways of evaluating the quality of responses
- **LLM as NLP (Natural Language Processing) Task Solvers**
 - NLP Tasks: Topic analysis, sentiment analysis, relation extraction, etc.
 - Transparent processes, interpretable results, and customizable tasks
 - Prompt engineering is non-trivial and using LLMs is still a costly approach

A Combined Approach

- Sensemaking of large collections of unstructured text
- Design considerations:
 - Overview of article topics, keywords, and their connections
 - Support of drill-down, progressive disclosure
 - Transparency, interpretability, and direct manipulation
 - Detailed analysis of a specific target of interest
- A combined approach:
 - Processing data with LLMs followed by interactive visualization to find and keep track of *targets of interest*
 - Sensemaking on the targets of interest with LLM Chatbots

Example I: HyperMap

a

b

c

d

1. Vladimir Putin
2. US Presidents
3. Benjamin Netanyahu

Mentioned Entities:
Donald Trump, Dilma Rousseff, Luiz Inácio Lula da Silva, Michel Temer, Confirmations of Barack Obama's Cabinet, Barack Obama, John F. Kennedy, Naoto Kan, Viktor Orbán, First impeachment of Donald Trump, Václav Klaus, Barbara Bush, Gerald Ford, Impending impeachment of the president, Stephen Harper, Miloš Zeman, Malcolm Turnbull, Upcoming presidential debates

Controversies and Challenges

Ride-hailing and Technology

Political and Social Issues

Security Threats in the Middle East

Political Developments in the Modern World

Contemporary American Society and Culture

Show Entity Label Off
Show Article Label On
Search Off

input search text

Relevance >= 0.80 Filter Search

Political Developments in the Modern World

Dilma Rousseff of Brazil Says She Won't Be Silenced in Impeachment Trial - The New York Times

The article discussed **Dilma Rousseff's** impeachment trial in the Senate, where she defended herself against charges of manipulating the federal budget. She compared her situation to past Brazilian presidents who were ousted by their opponents and argued that her removal would increase the risks of investing in the country. If convicted, **Michel Temer**, the interim president, will serve until 2018.
Article ID: #21479 Topic: Political Developments in the Modern World Relevance:

New Zealand Struck by Powerful Earthquake - The New York Times

The article discussed a powerful **earthquake** that hit the east coast of **New Zealand's** South Island, causing tsunami waves and resulting in at least two deaths. The **earthquake** was followed by several aftershocks, and officials warned people living near the coast to move to higher ground. Prime Minister John Key confirmed the deaths and reported damage to buildings in Wellington, where he had met with Secretary of State John Kerry.
Article ID: #22321 Topic: Political Developments in the Modern World Relevance:

Five Illuminating Takeaways From Your Family Money Letters - The New York Times

The article discussed the value of writing a "money letter" to impart financial wisdom to loved ones. **Joe Olivier** wrote a letter to his niece, asking her to read books on personal finance, take an online accounting course, and write reports. Other letter writers emphasized the importance of recognizing emotions around money, the connection between money and sex, the need for risk

Given the selected documents, what was the reason for Dilma Rousseff to be impeached?

According to the selected article, Dilma Rousseff was impeached because of allegations that she manipulated the federal budget to hide Brazil's economic problems.

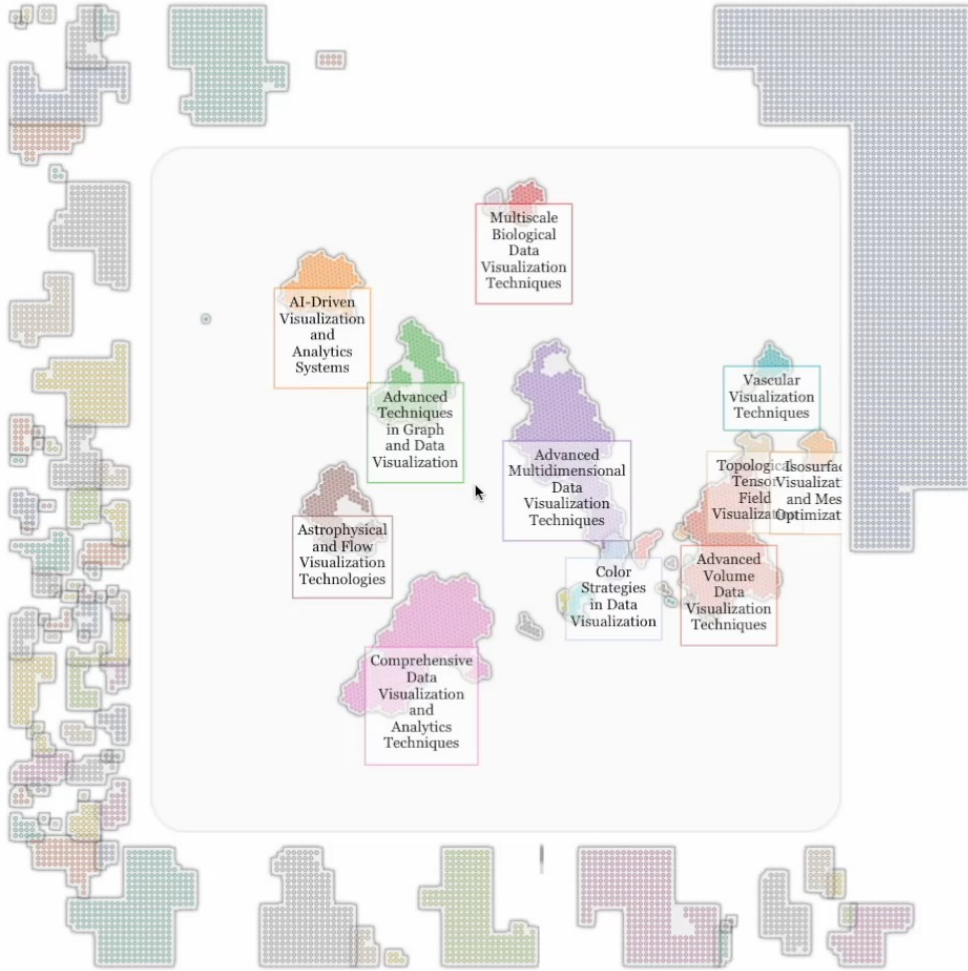
Could you elaborate more on these allegations?

Certainly! The allegations against Dilma Rousseff revolved around her alleged manipulation of the federal budget. It was claimed that Rousseff used accounting tricks, known as "fiscal pedaling" or "creative accounting," to hide the true level of Brazil's budget deficit during her re-election campaign in 2014. Specifically, Rousseff was accused of taking unauthorized loans from state-owned banks to cover budget shortfalls and make government finances appear better than they actually were. By doing so, she was allegedly violating Brazil's fiscal responsibility laws, which prohibit such practices. The impeachment process accused Rousseff of committing a "crime of responsibility" by manipulating the budget and violating fiscal laws. The charges were based on an

Given the selected documents,

Selected Documents:
#21479

Use Summary Use Full Content Clear Chat



Show Entity Label

Show Article Label

Search

Search...

Relevance >= 0.8



Document List

- Article ID: #1 Topic: undefined Relevance: undefined

- Article ID: #2 Topic: undefined Relevance: undefined

Given the selected documents,

Selected Documents:

Example I: HyperMap

- User Study on literature review using HyperMap
 - Chatbot group vs HyperMap group
- Findings
 - Chatbot is good at explaining unknown keywords and summarizing articles or the course of conversation.
 - Users have less trust over the Chatbot for critical questions (e.g., if a statement/claim is true/justified)
 - The performance of the Chatbot group is less consistent (e.g., the user may get contradictory answers or be unsure what to ask next.)
 - Visualizations can help direct the process and track the analysis targets.

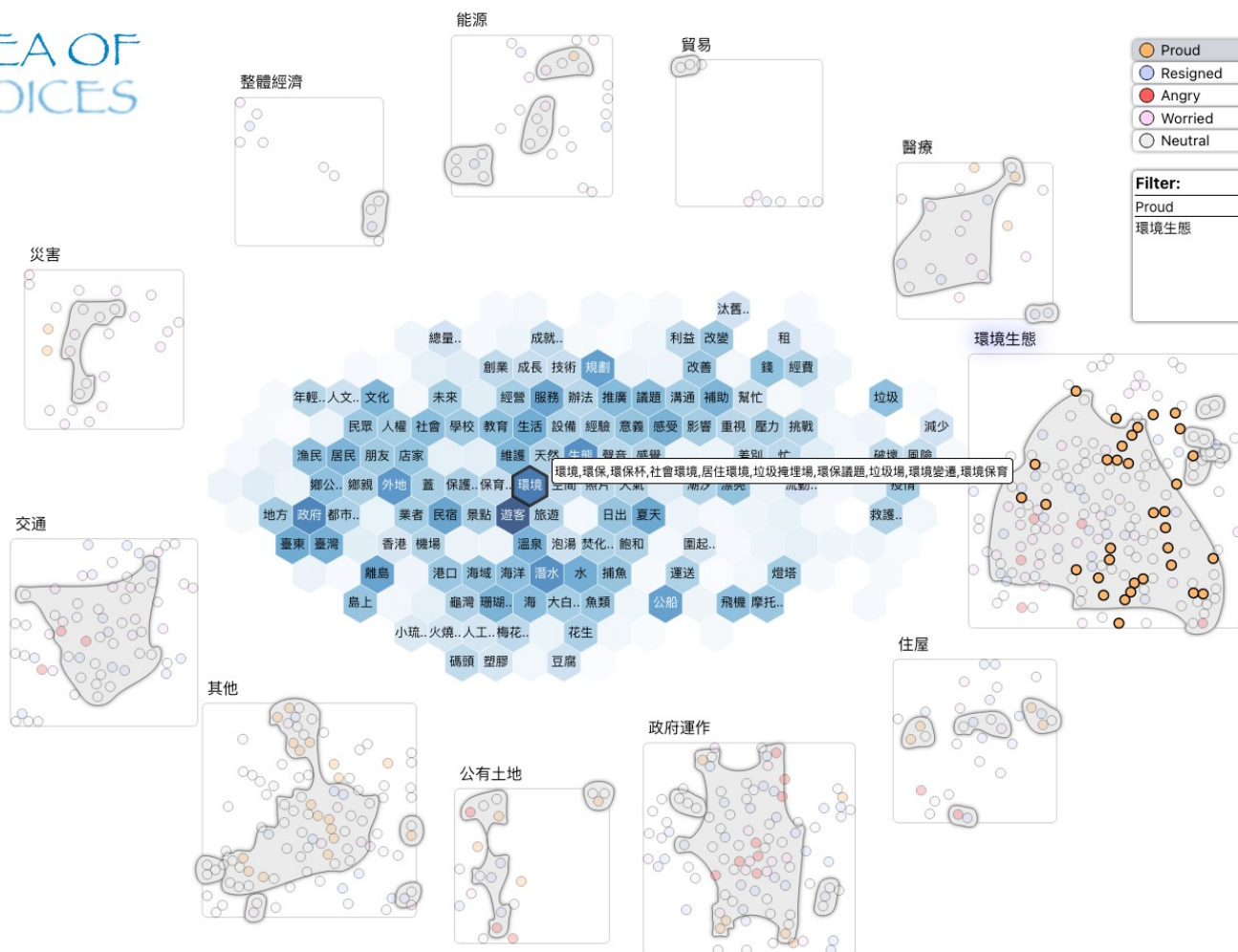
Example II: Sea of Voices

SEA OF
VOICES

Surfacing the Unheard Voices
Lyudao Resident Interview Exploration

Example II: Sea of Voices

SEA OF VOICES



Proud
 Resigned
 Angry
 Worried
 Neutral

Filter:
 Proud
 環境生態

N2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
N3	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
N4	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
N5	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
N6	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

31. 傳統文化保存與當地女性角色的重要性

Host:
你剛剛說的是那個技能的部分，我也想問一下，為什麼你會覺得這裏需要這樣子幫助？為什麼需要聽上那些課程？

Guest:
 嗯，因為像在地的一些阿姨，他們其實都會自己做一些像我們很有名的叫艾粿，對然後跟他們也會曬章魚乾，然後後面也會炸飛魚或者是什麼，但他們通常都只會拿來給自己的吃，但他們比這樣，沒有人去協助他們去做一個對外做產銷的這個東西，因為他們不會行銷，他們就只會自己吃，所以如果有人去幫助他們做這些行銷的話，我會覺得很好，然後像很傳統都這樣，艾粿啊，花生豆腐。花生系列像我以前都還會吃到那個花生豆腐，但現在我完全吃不到花生豆腐啊，因為阿姨，他們都受傷了，不做了。雖然說花生都不好做，沒有錯，可是我覺得這些傳統是該被留下來的，就是我覺得政府可以留一些經費或一些管道，或者是收集一個什麼樣子的地方可以讓他們更就是保存這些的存在可能是一個博物館，它博物館裡面可能保留的是如果是咕咾石，然後綠島為什麼會為什麼會有綠島嘛，綠島第一批人，所有的歷史，我覺得所有的綠島不一定是只有人權這件事情，我覺得人權這件事情是很重要的，沒有錯，但我覺得它不是一個一定是應該是它可以有別的存在，他別的存在就是綠島的歷史是什麼？因為大家來的只會說我要去冷泉博物館，可是完全沒有了解到綠島歷史是什麼，我覺得歷史這件事情很重要。對然後，題外話，我自己很喜歡這裏的一個東西叫野百合，那野百合對我來講啊，它就是一其實代表的就是人權；第二，它代表的其實就是這裡的女性，我覺得這裡的女性就跟野百合一樣，它是非常的堅強的，它可以度過寒風啊，或者是什麼，對我來講，這裡的女性就是阿姨他們那些人都是非常厲害的，因為他們從小就直接嫁過來這裡，然後在這裡完全沒有任何的資源的時候，他們就

Visual Analytics for High Dimensional Data

- Hyeon Jeon, Yun-Hsin Kuo, Michaël Aupetit, Kwan-Liu Ma, Jinwook Seo: Classes are not Clusters: Improving Label-based Evaluation of Dimensionality Reduction. **IEEE Transactions on Visualization and Computer Graphics** 30(1) (2024) (presented at **IEEE VIS** 2023)
- Takanori Fujiwara, Yun-Hsin Kuo, Anders Ynnerman, Kwan-Liu Ma: Feature Learning for Nonlinear Dimensionality Reduction toward Maximal Extraction of Hidden Patterns. **IEEE PacificVis** 2023: 122-131
- Shilpika, Takanori Fujiwara, Naohisa Sakamoto, Jorji Nonaka, Kwan-Liu Ma: A Visual Analytics Approach for Hardware System Monitoring with Streaming Functional Data Analysis. **IEEE Transactions on Visualization and Computer Graphics** 28(6): 2338-2349 (2022)
- Takanori Fujiwara, Xinhai Wei, Jian Zhao, Kwan-Liu Ma: Interactive Dimensionality Reduction for Comparative Analysis. **IEEE Transactions on Visualization and Computer Graphics** 28(1): 758-768 (2022) (presented at **IEEE VIS** 2021)
- Takanori Fujiwara, Shilpika, Naohisa Sakamoto, Jorji Nonaka, Keiji Yamamoto, Kwan-Liu Ma: A Visual Analytics Framework for Reviewing Multivariate Time-Series Data with Dimensionality Reduction. **IEEE Transactions on Visualization and Computer Graphics** 27(2): 1601-1611 (2021) (presented at **IEEE VIS** 2020)
- Takanori Fujiwara, Jia-Kai Chou, Shilpika, Panpan Xu, Liu Ren, Kwan-Liu Ma: An Incremental Dimensionality Reduction Method for Visualizing Streaming Multidimensional Data. **IEEE Transactions on Visualization and Computer Graphics** 26(1): 418-428 (2020)

Acknowledgments

- My students and collaborators
- National Institute of Health
- Department of Energy
- National Science Foundation
- University of California Climate Action Initiative
- Bosch Research
- Intel Corporation
- Institute of Statistical Science, Academia Sinica