SemanticAxis

Exploring Multi-attribute Data by Semantics Construction and Ranking Analysis

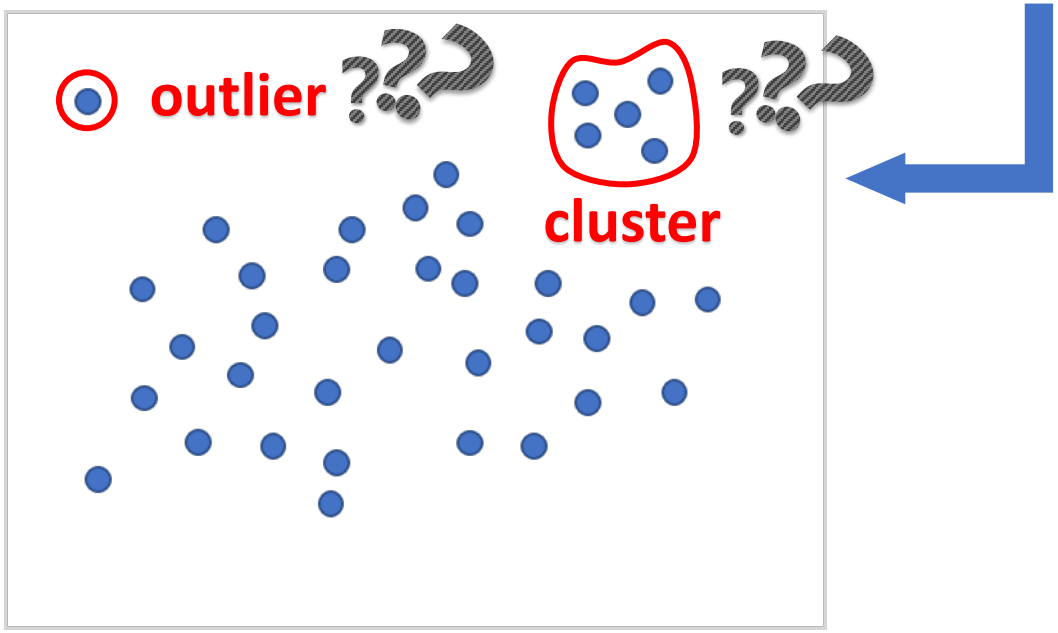
Zeyu Li, Changhong Zhang, Yi Zhang, Jiawan Zhang
Tianjin University

Motivation

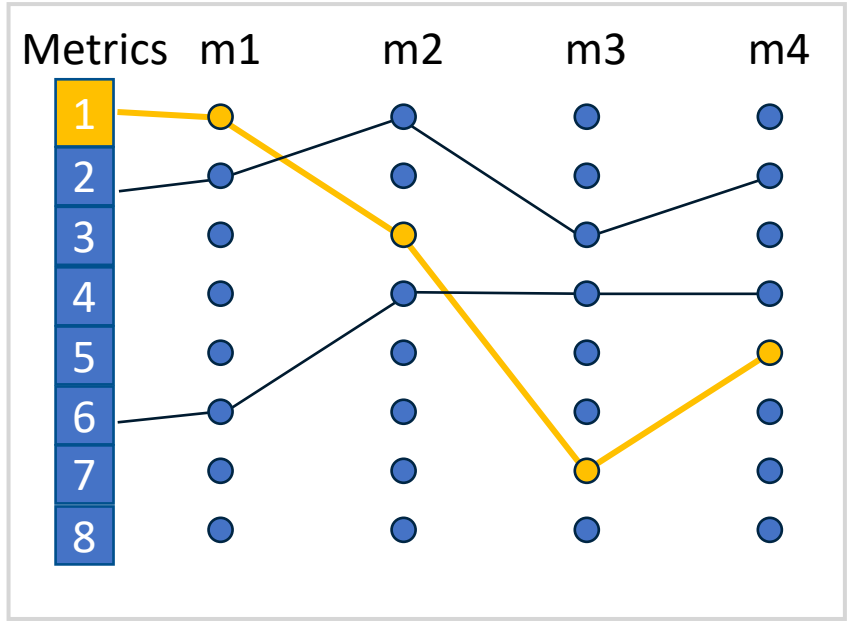
Two common tasks for multi-attribute data analysis:

Pattern recognition

Dimension Reduction (DR)



Multi-attribute rankings

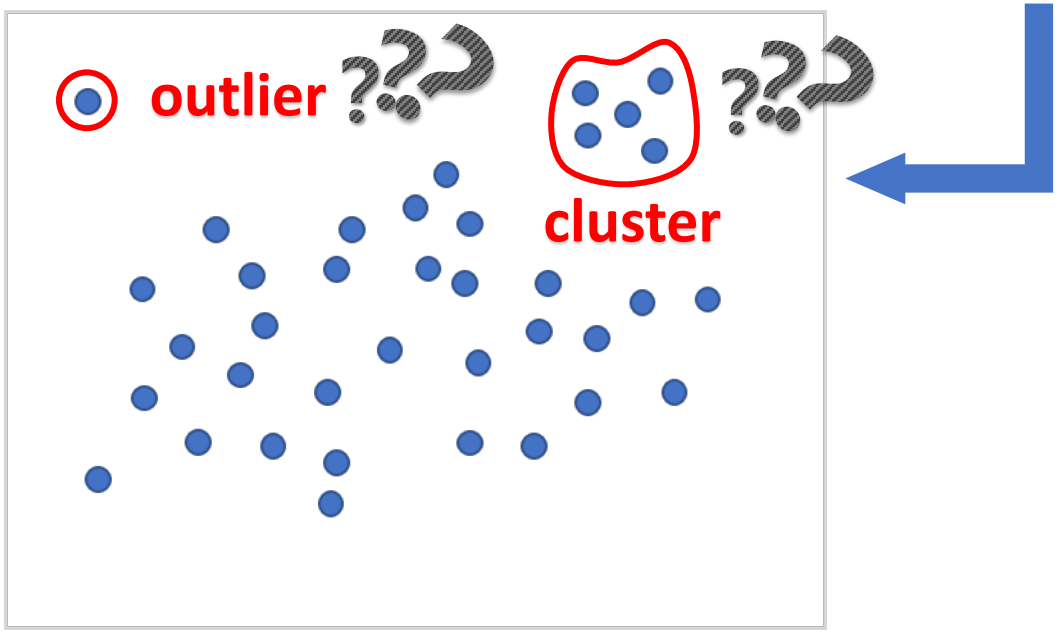


Motivation

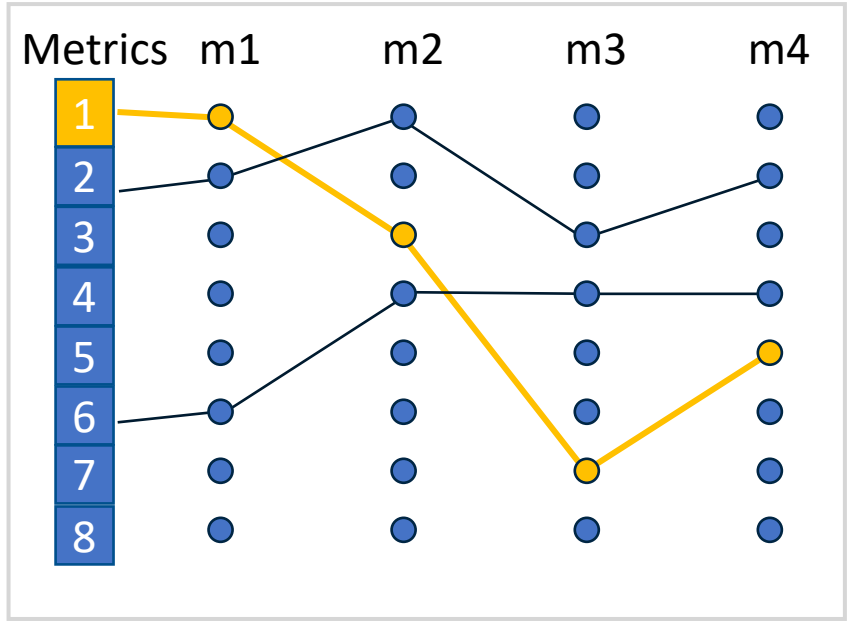
Two common tasks for multi-attribute data analysis:

Pattern finding in DR results

Dimension Reduction (DR)



Multi-attribute rankings

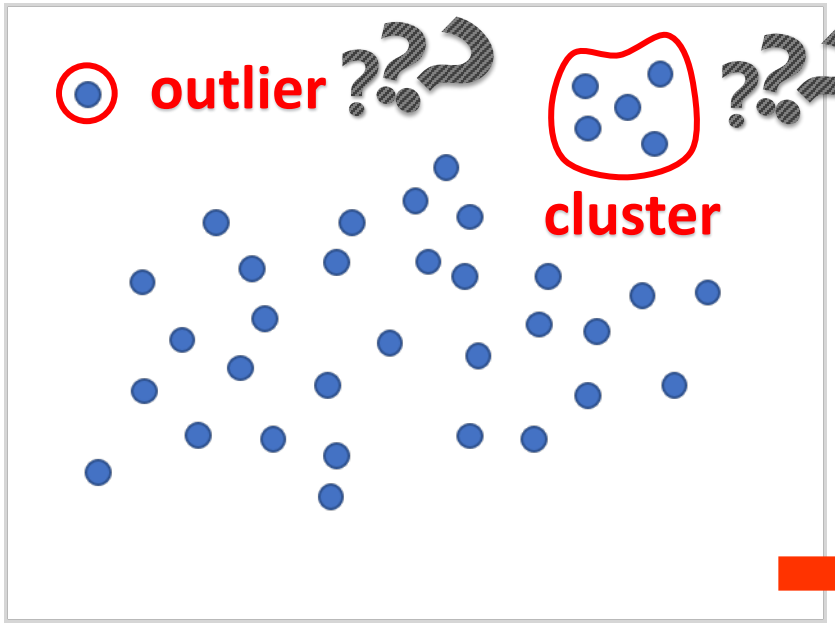


Motivation

Two common tasks for multi-attribute data analysis:

Understanding DR results

Dimension Reduction (DR)



Multi-attribute rankings

Attribute weights

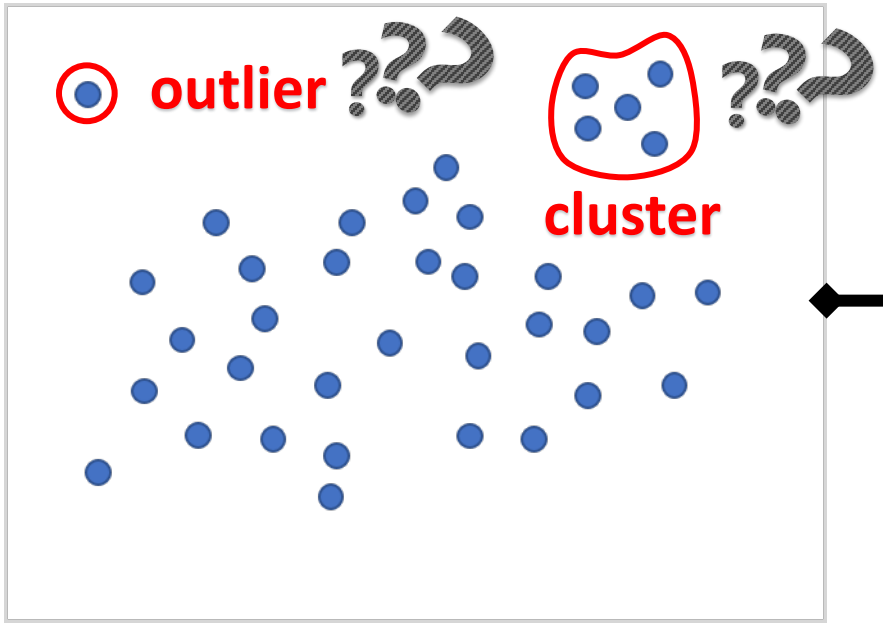
Ranking metrics

Ranking semantics

Motivation

Two common tasks for multi-attribute data analysis:

Understanding DR results



SemanticAxis

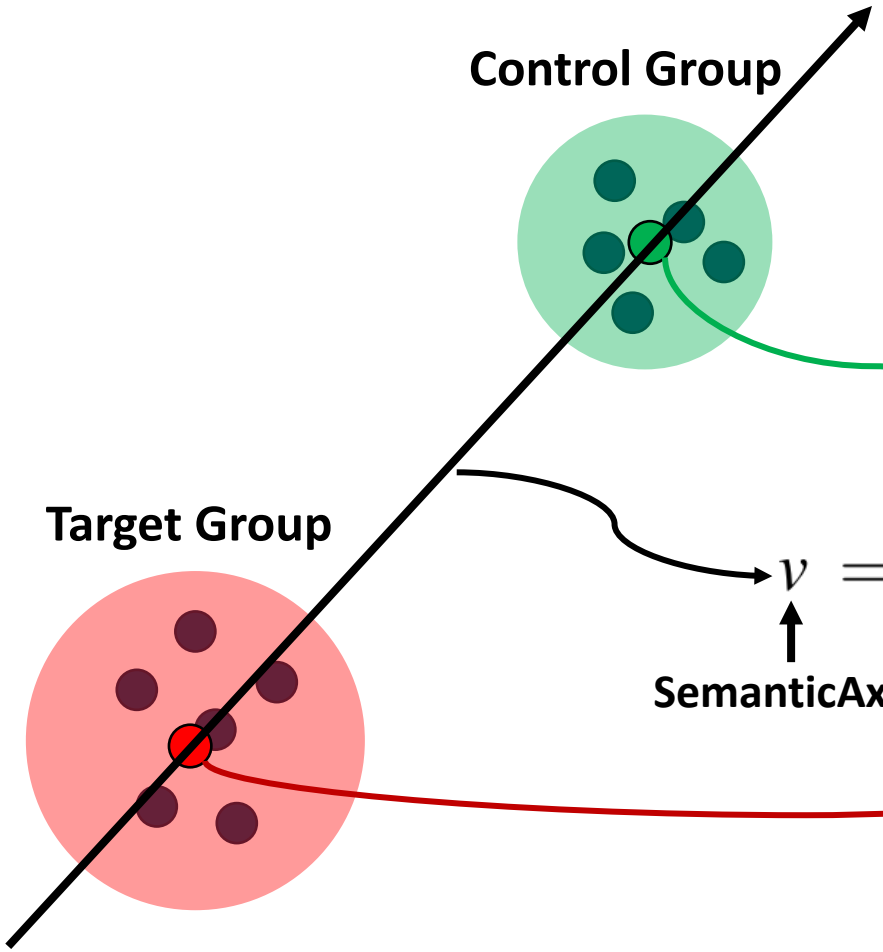
Multi-attribute rankings

Attribute weights

Ranking metrics

Ranking semantics

Creation of SemanticAxis



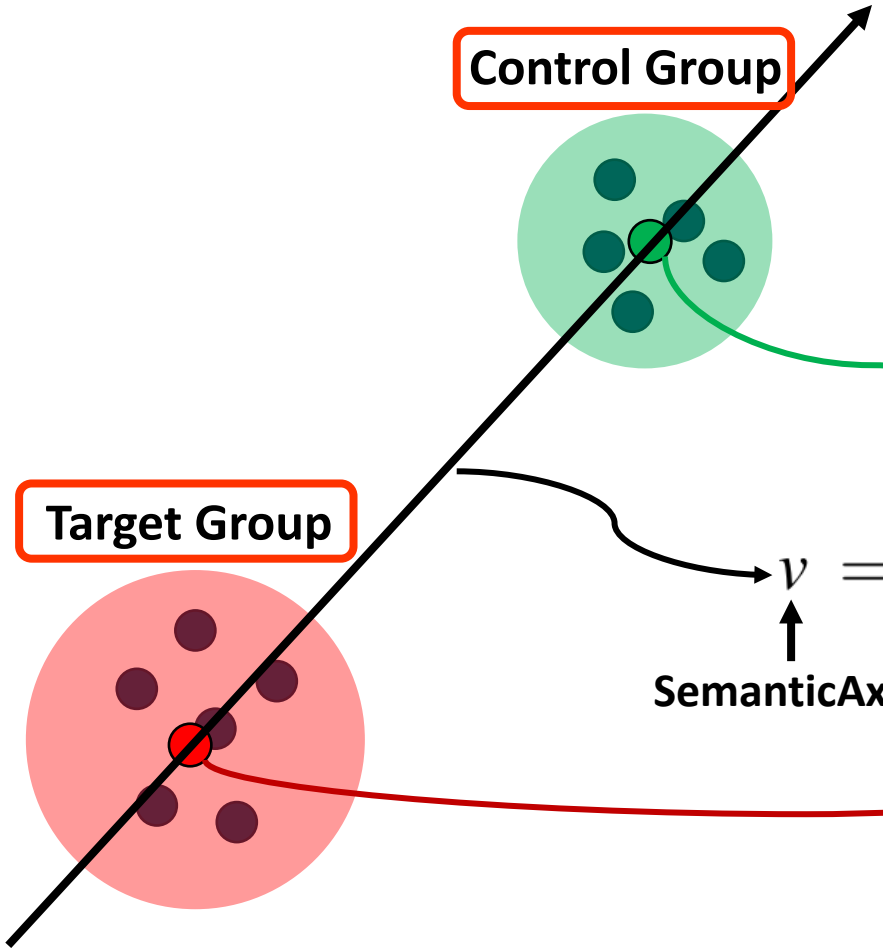
$$v = \left(\frac{1}{C} \sum_{i=1}^{C} p_i - \frac{1}{T} \sum_{j=1}^{T} p_j \right) \odot \omega \qquad p_i^l,\ p_j^r, \omega \in R^{1 \times N}$$

SemanticAxis

element-wise product

attribute weights

Creation of SemanticAxis



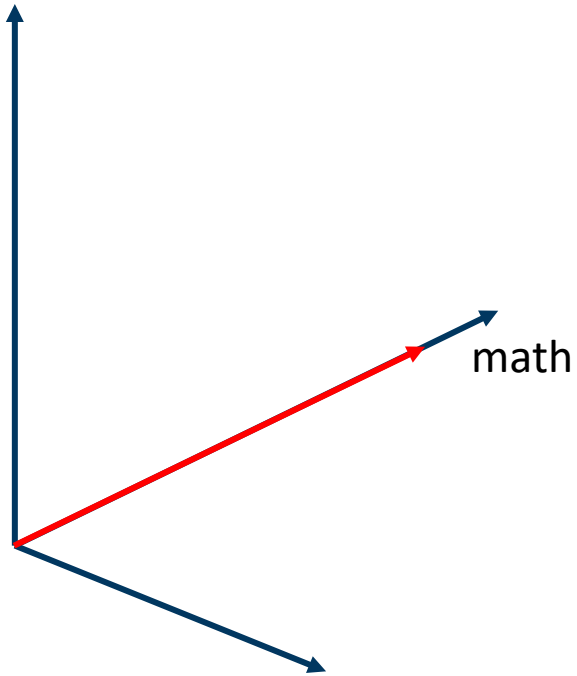
$$v = \left(\frac{1}{C} \sum_{i=1}^{C} p_i - \frac{1}{T} \sum_{j=1}^{T} p_j \right) \odot \omega \qquad p_i^l,\ p_j^r, \omega \in R^{1 \times N}$$

SemanticAxis

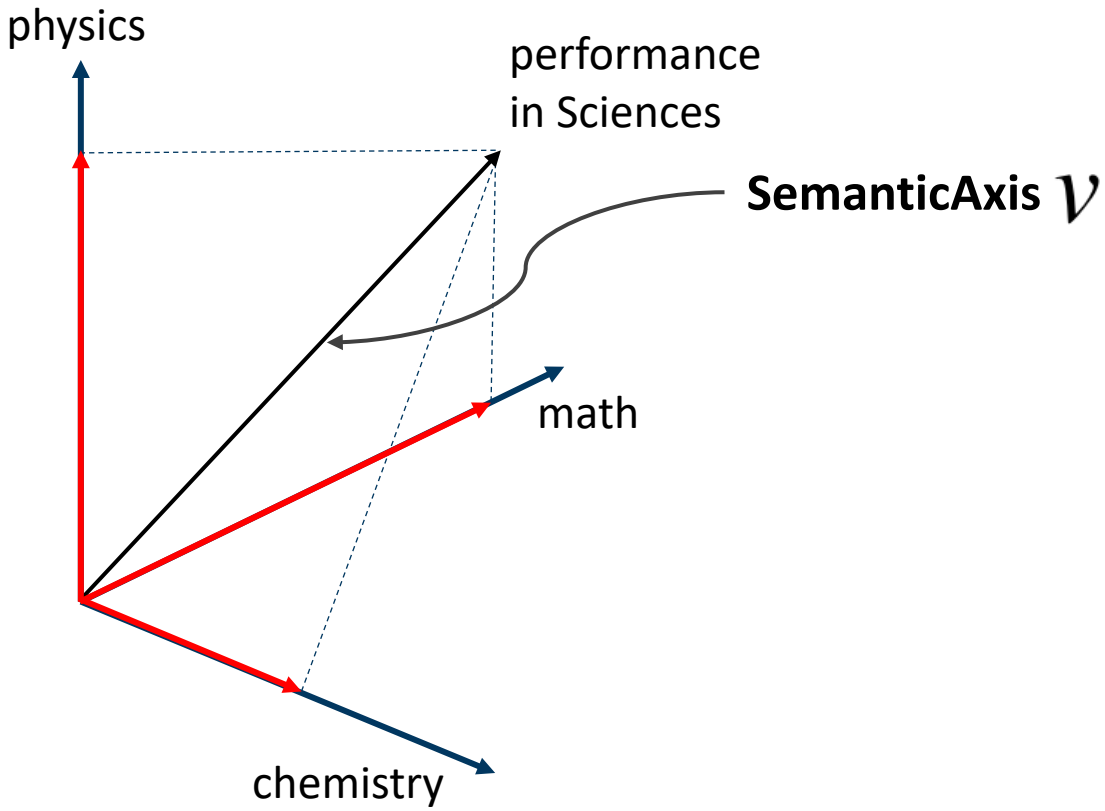
element-wise product

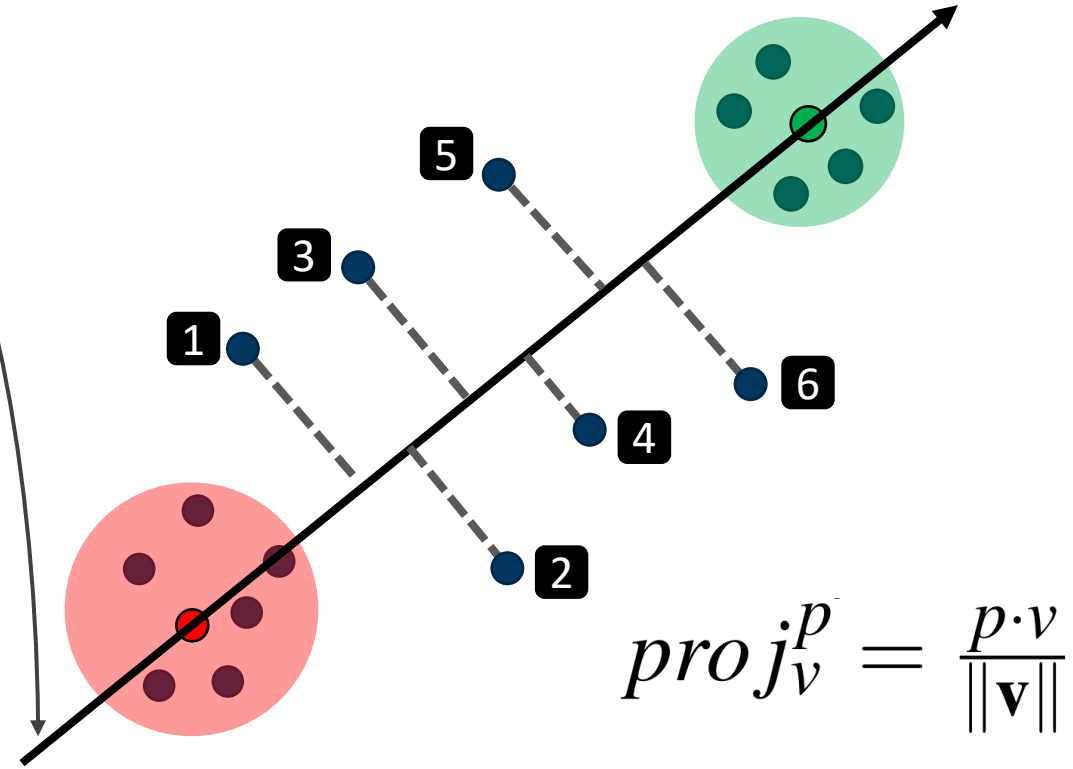
attribute weights

Supporting multi-attribute rankings



Supporting multi-attribute rankings



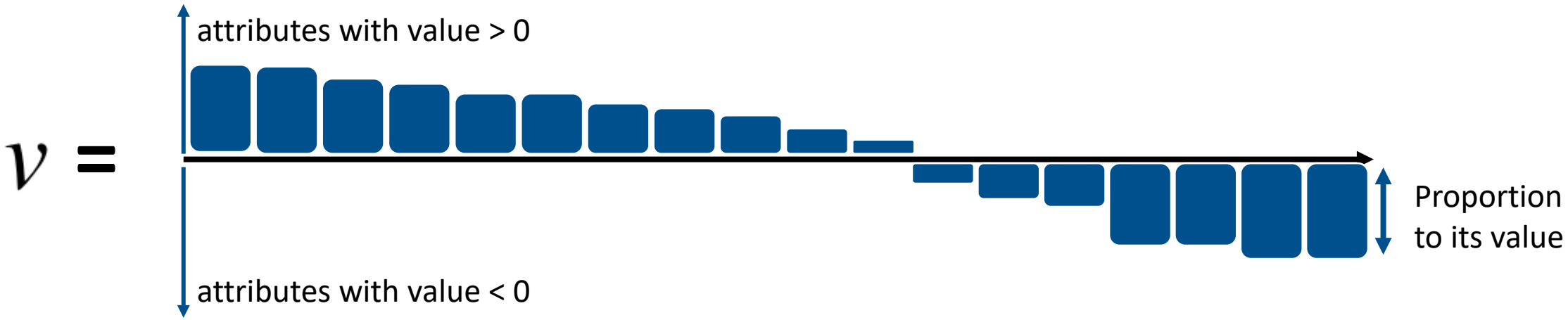


$$proj_v^p = \frac{p \cdot v}{\|\mathbf{v}\|}$$

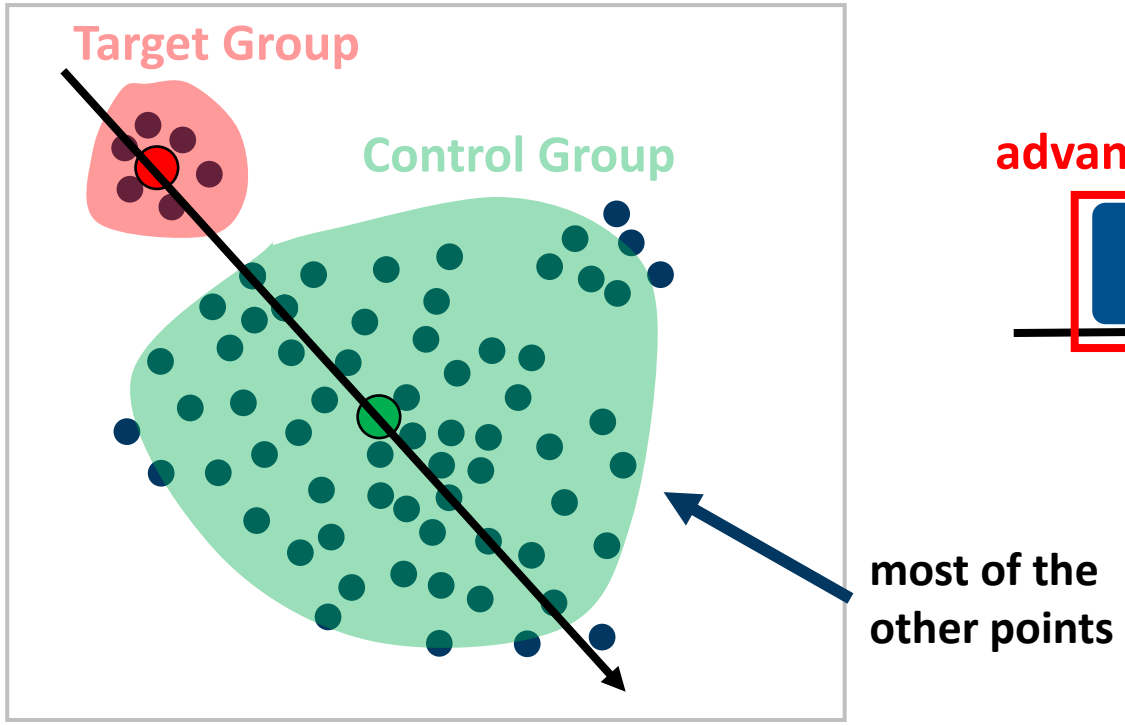
Interpreting the semantics uncovered by a semantic axis

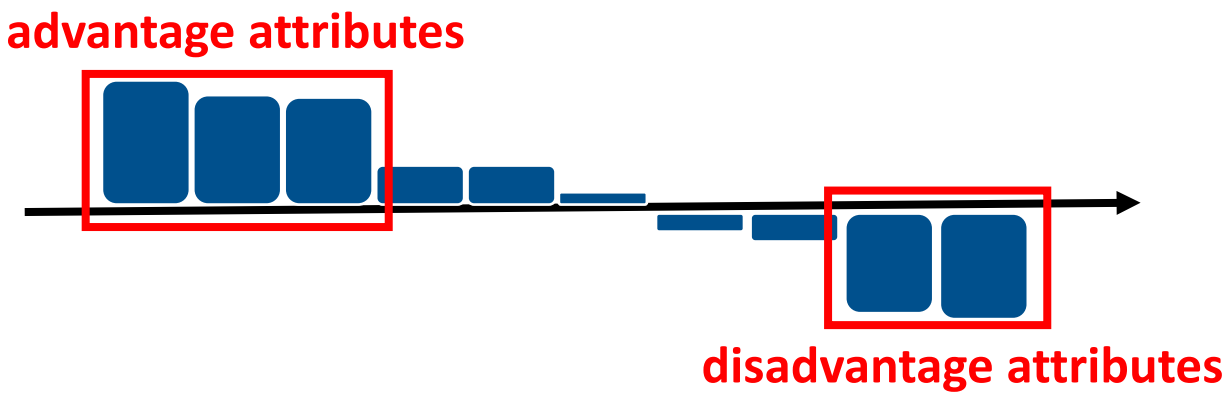
$v =$

Interpreting the semantics uncovered by a semantic axis

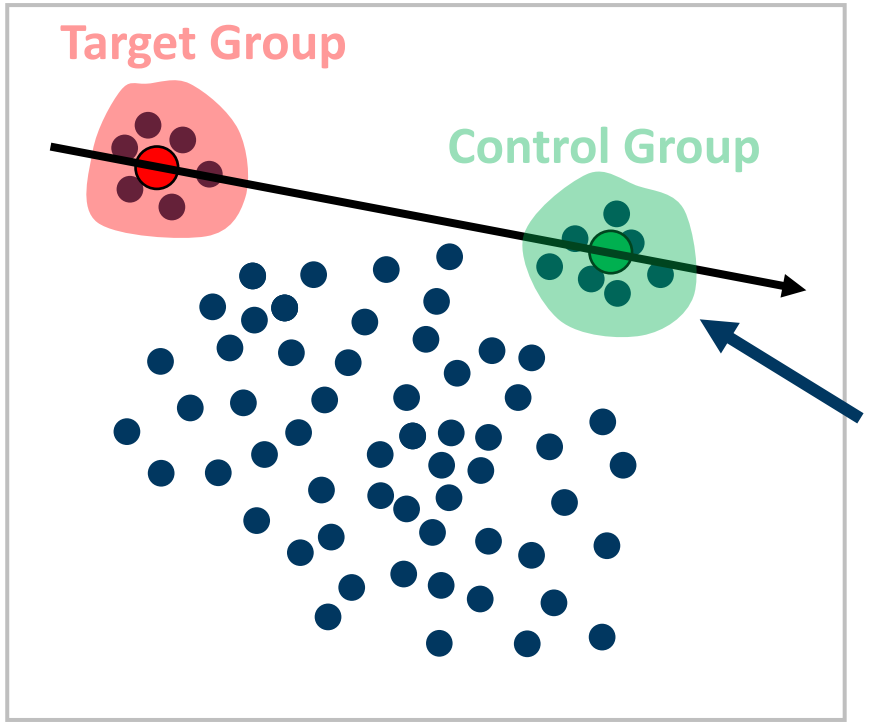


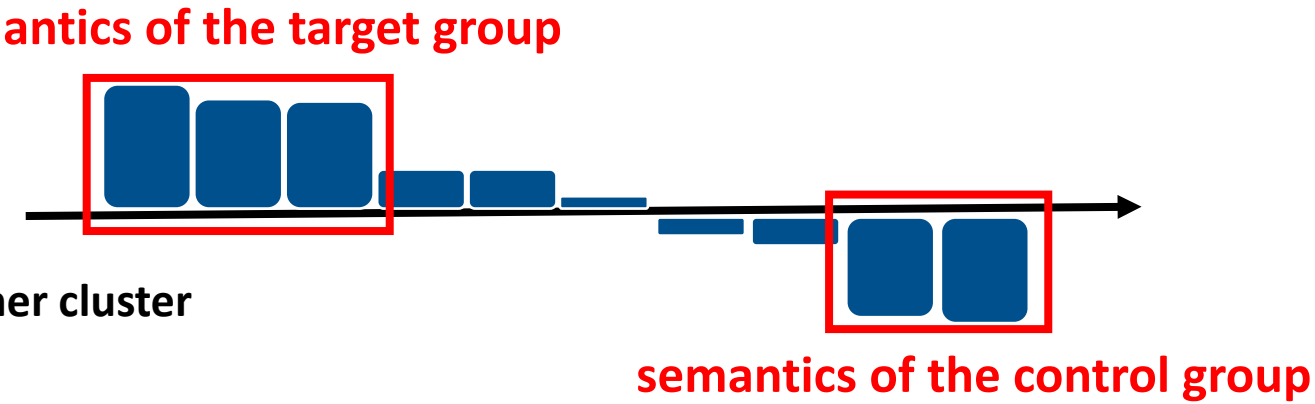
Unipolar semantic axis

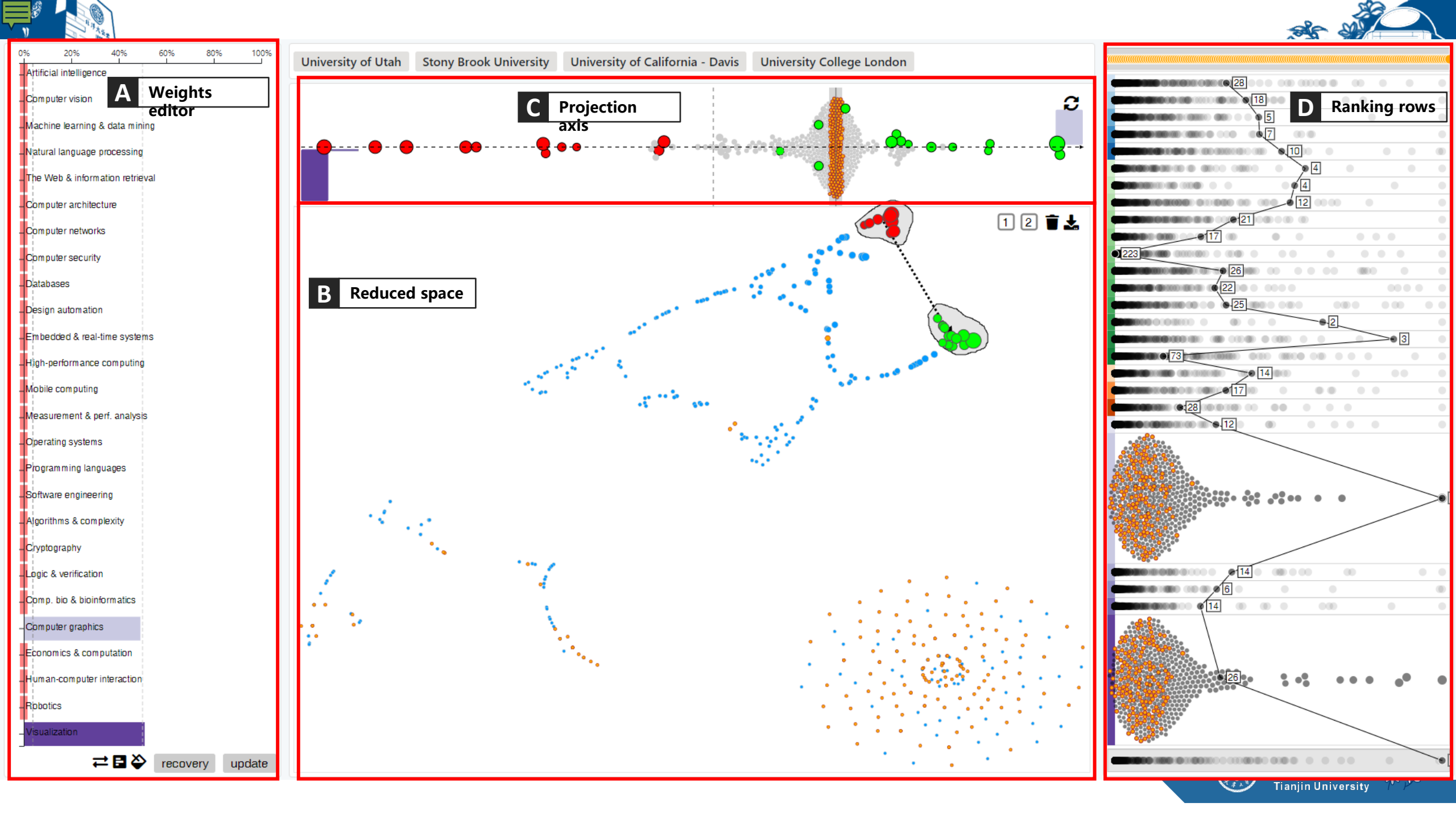


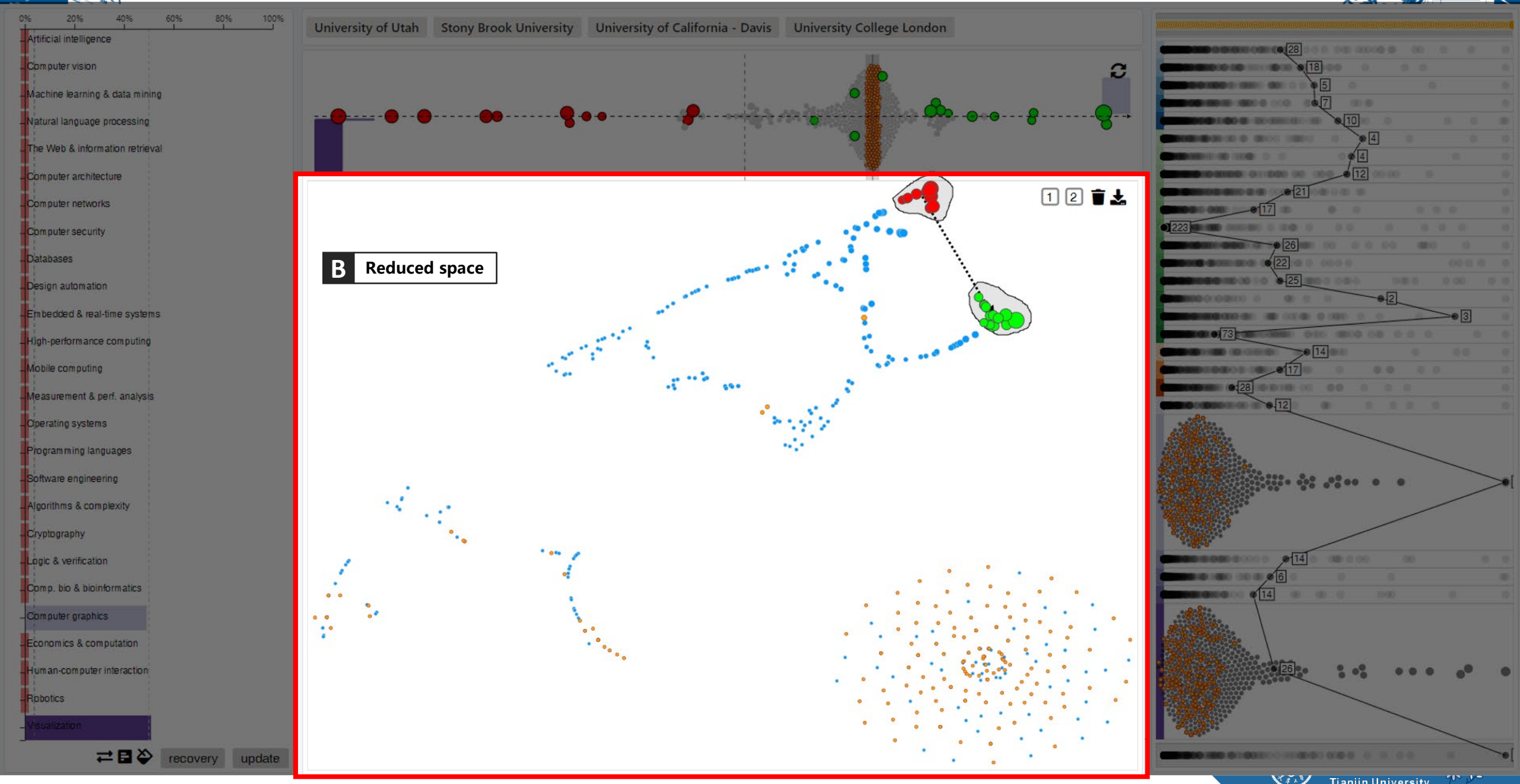


Bipolar semantic axis

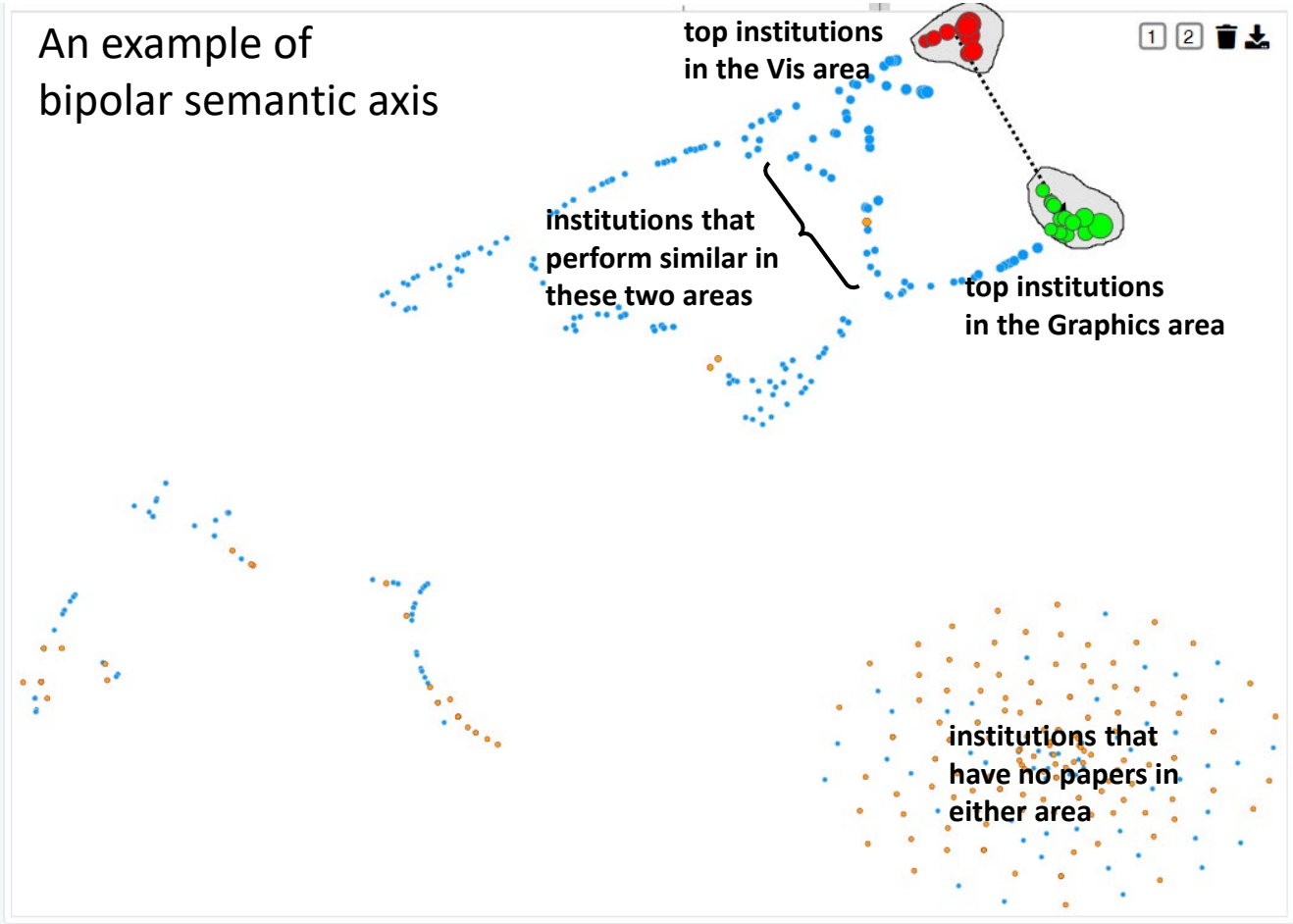





0%
20%
40%
60%
80%
100%
A Weights editor
Artificial intelligence
Computer vision
Machine learning & data mining
Natural language processing
The Web & information retrieval
Computer architecture
Computer networks
Computer security
Databases
Design automation
Embedded & real-time systems
High-performance computing
Mobile computing
Measurement & perf. analysis
Operating systems
Programming languages
Software engineering
Algorithms & complexity
Cryptography
Logic & verification
Comp. bio & bioinformatics
Computer graphics
Economics & computation
Human-computer interaction
Robotics
Visualization
recovery
update
University of Utah
Stony Brook University
University of California - Davis
University College London
C Projection axis
B Reduced space
1
2
D Ranking rows
28
18
5
7
10
4
4
12
21
17
223
26
22
25
2
3
73
14
17
28
12
14
6
14
26
Tianjin University

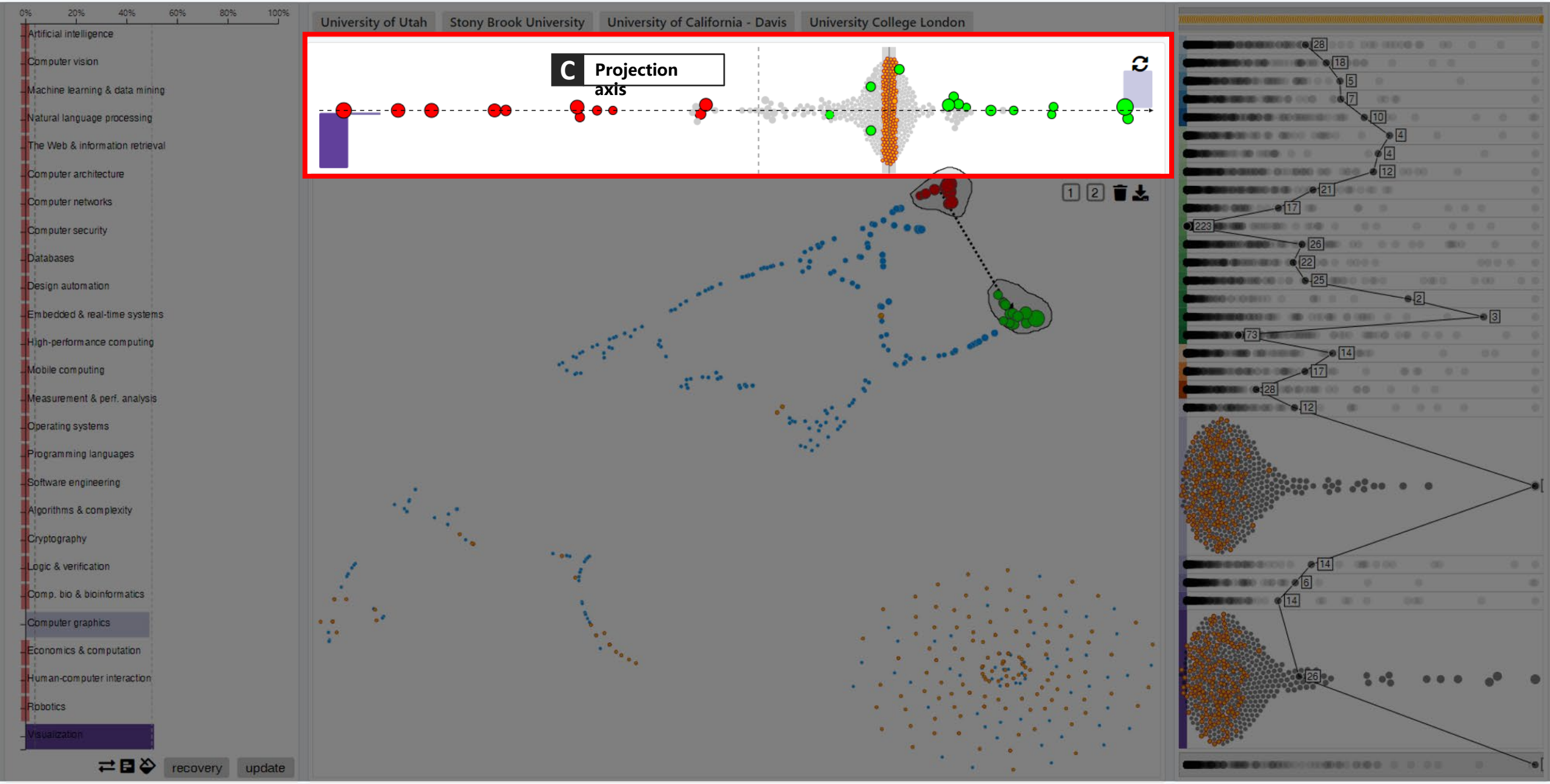

0%
20%
40%
60%
80%
100%
Artificial intelligence
Computer vision
Machine learning & data mining
Natural language processing
The Web & information retrieval
Computer architecture
Computer networks
Computer security
Databases
Design automation
Embedded & real-time systems
High-performance computing
Mobile computing
Measurement & perf. analysis
Operating systems
Programming languages
Software engineering
Algorithms & complexity
Cryptography
Logic & verification
Comp. bio & bioinformatics
Computer graphics
Economics & computation
Human-computer interaction
Robotics
Visualization
recovery
update
University of Utah
Stony Brook University
University of California - Davis
University College London
B
Reduced space
1
2
28
18
5
7
10
4
4
12
21
17
223
26
22
25
2
3
73
14
17
28
12
14
6
14
26

Reduced space

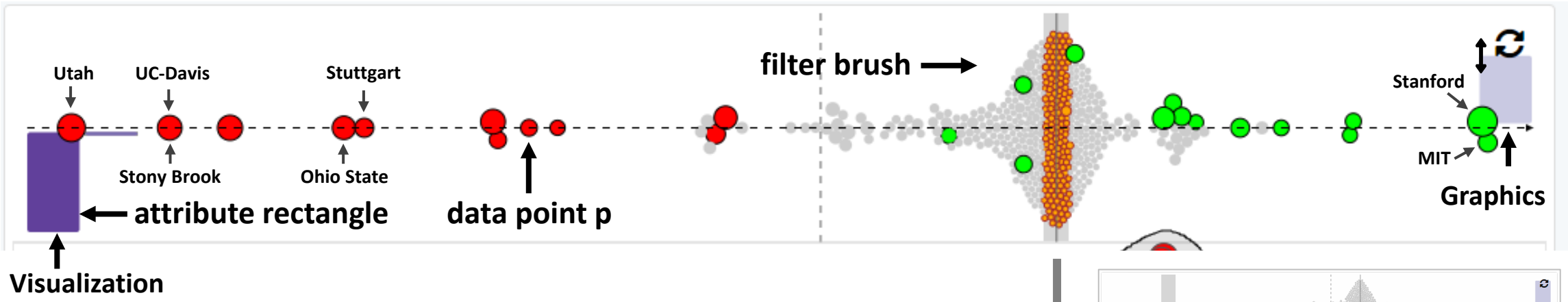


DR algorithm: t-SNE

Radius of points: weighted scores $\sum_{i=1}^{N} \bar{p} \cdot \omega$

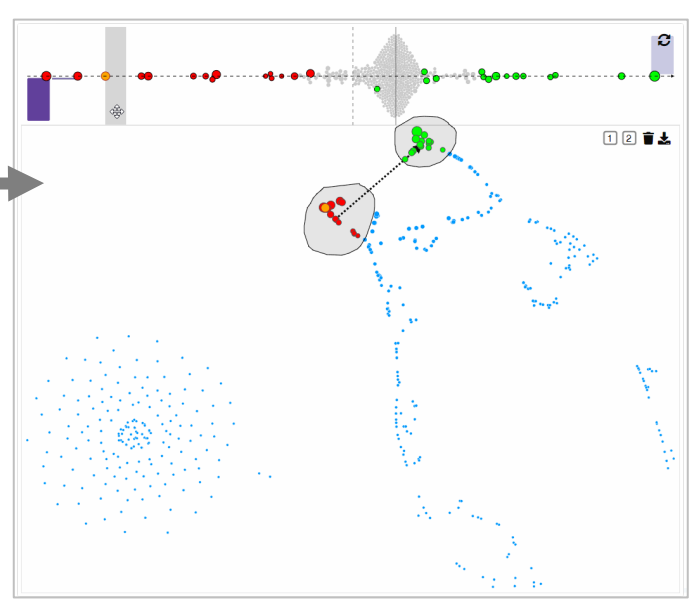

0%
20%
40%
60%
80%
100%
Artificial intelligence
Computer vision
Machine learning & data mining
Natural language processing
The Web & information retrieval
Computer architecture
Computer networks
Computer security
Databases
Design automation
Embedded & real-time systems
High-performance computing
Mobile computing
Measurement & perf. analysis
Operating systems
Programming languages
Software engineering
Algorithms & complexity
Cryptography
Logic & verification
Comp. bio & bioinformatics
Computer graphics
Economics & computation
Human-computer interaction
Robotics
Visualization
recovery
update
University of Utah
Stony Brook University
University of California - Davis
University College London
C
Projection axis
1
2
28
18
5
7
10
4
4
12
21
17
223
26
22
25
2
3
73
14
17
28
12
14
6
14
26

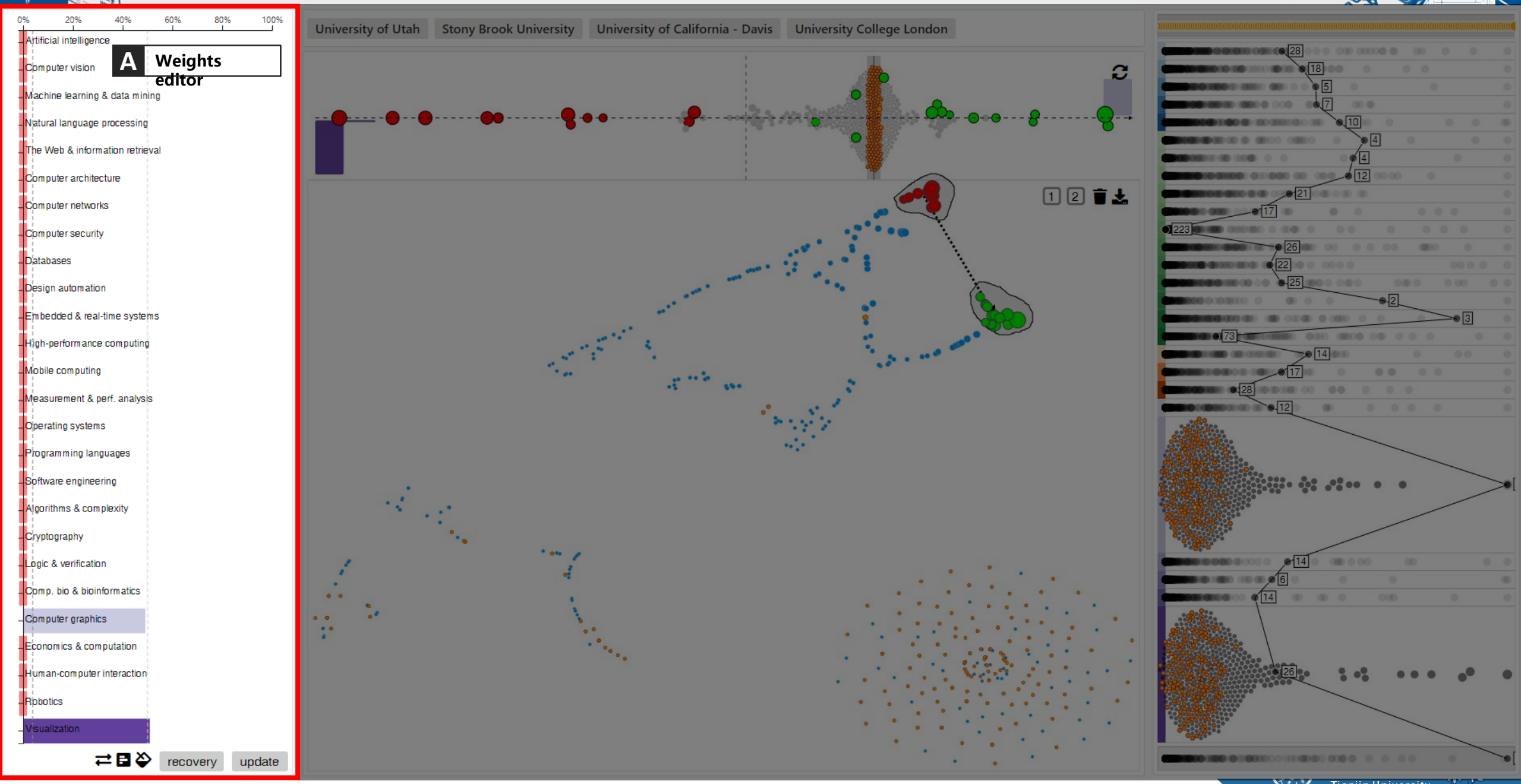
Projection axis



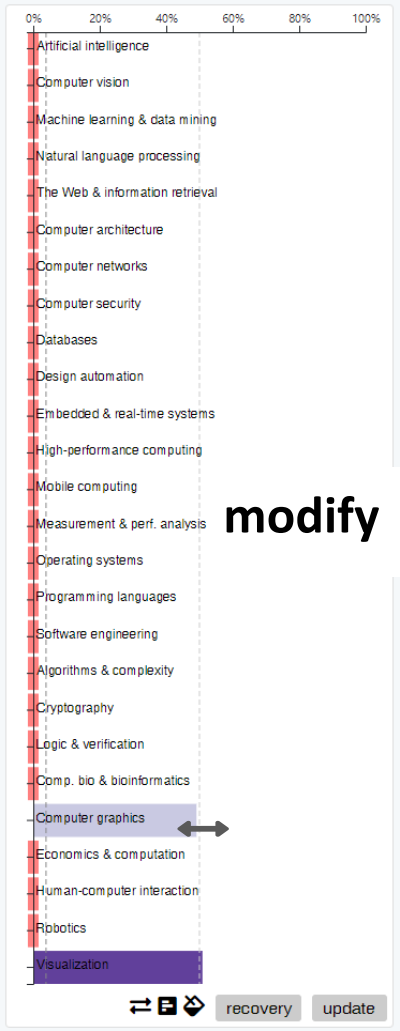
Radius of points: weighted score $\sum_{i=1}^{N} p \cdot \omega$

X-position of points: projection position $proj_v^p = \frac{p \cdot v}{\|\mathbf{v}\|}$




0%
20%
40%
60%
80%
100%
A Weights editor
Artificial intelligence
Computer vision
Machine learning & data mining
Natural language processing
The Web & information retrieval
Computer architecture
Computer networks
Computer security
Databases
Design automation
Embedded & real-time systems
High-performance computing
Mobile computing
Measurement & perf. analysis
Operating systems
Programming languages
Software engineering
Algorithms & complexity
Cryptography
Logic & verification
Comp. bio & bioinformatics
Computer graphics
Economics & computation
Human-computer interaction
Robotics
Visualization
recovery
update
University of Utah
Stony Brook University
University of California - Davis
University College London
1
2
28
18
5
7
10
4
4
12
21
17
223
26
22
25
2
3
73
14
17
28
12
14
6
14
26

Weight editor

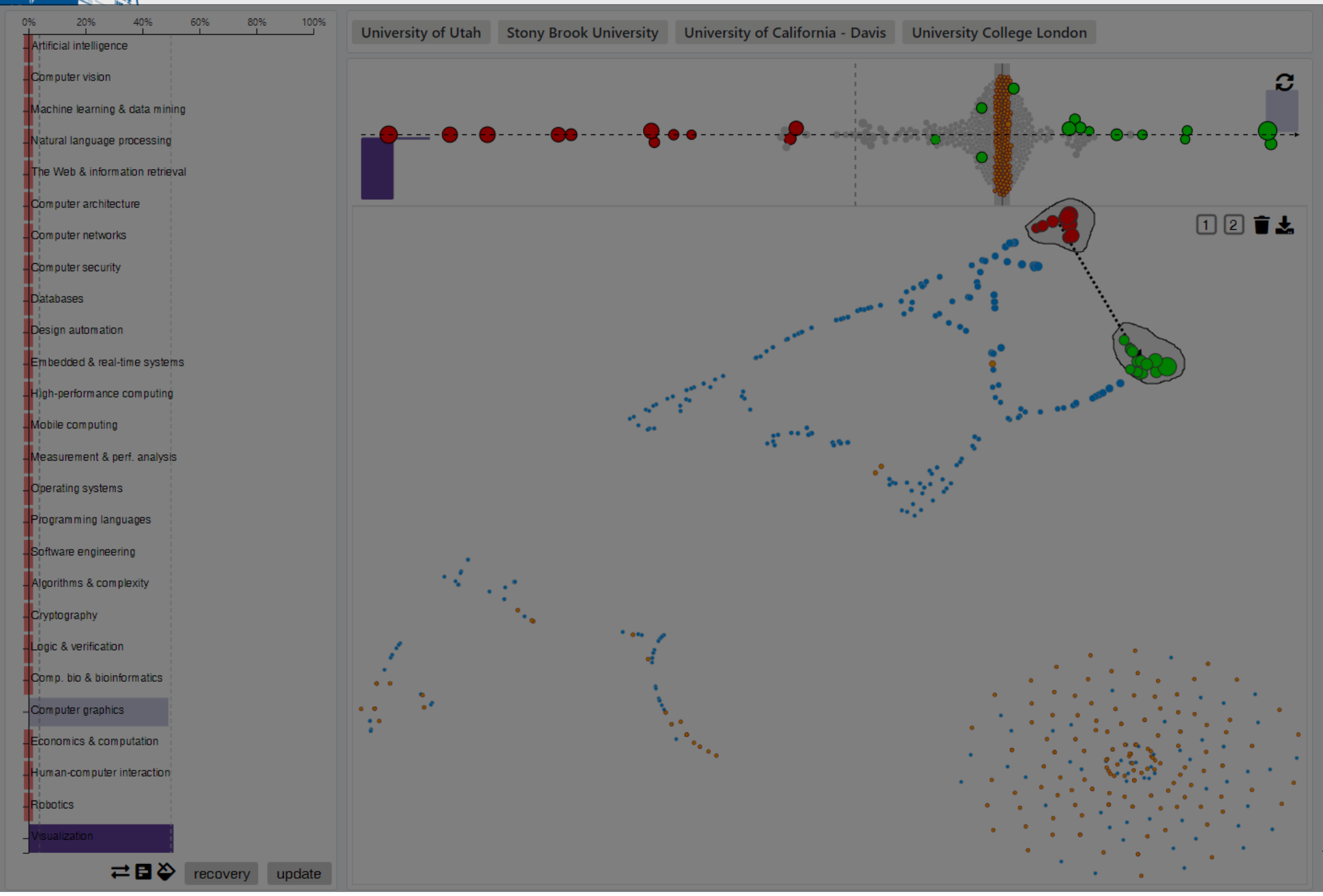


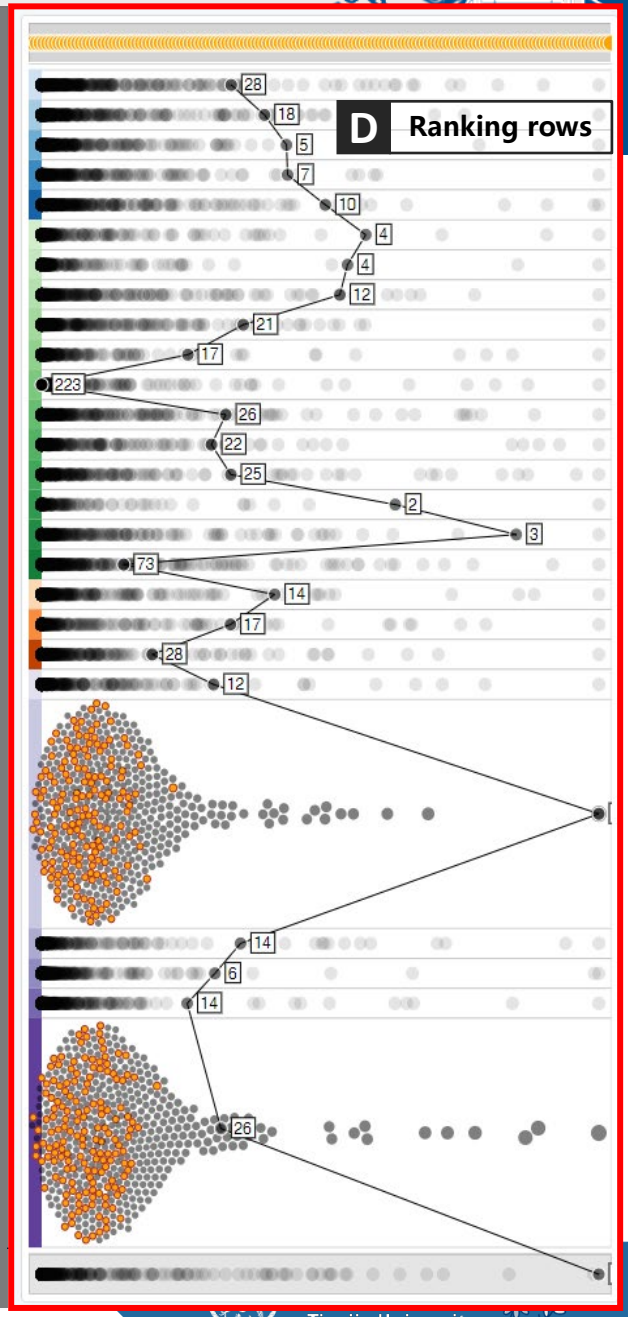
modify ω

Purpose:

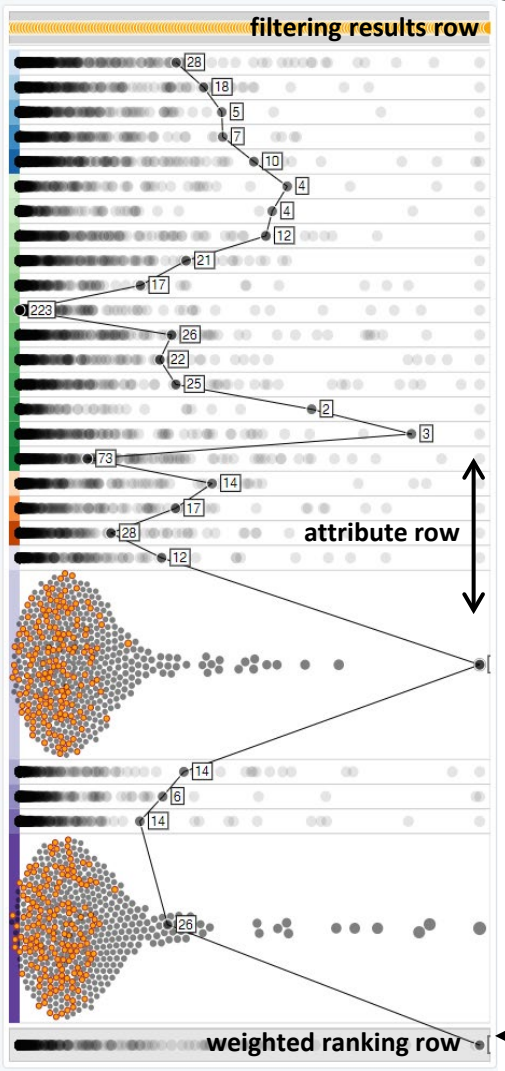
1. Support interactive refinement in multi-attribute rankings, allow shaping ranking criteria according to preferences

2. Rebuild reduced space, in order to gather data points that conform to the target semantics in close proximity


0%
20%
40%
60%
80%
100%
Artificial intelligence
Computer vision
Machine learning & data mining
Natural language processing
The Web & information retrieval
Computer architecture
Computer networks
Computer security
Databases
Design automation
Embedded & real-time systems
High-performance computing
Mobile computing
Measurement & perf. analysis
Operating systems
Programming languages
Software engineering
Algorithms & complexity
Cryptography
Logic & verification
Comp. bio & bioinformatics
Computer graphics
Economics & computation
Human-computer interaction
Robotics
Visualization
recovery
update
University of Utah
Stony Brook University
University of California - Davis
University College London
1
2


D
Ranking rows
28
18
5
7
10
4
4
12
21
17
223
26
22
25
2
3
73
14
17
28
12
14
6
14
26

Ranking rows

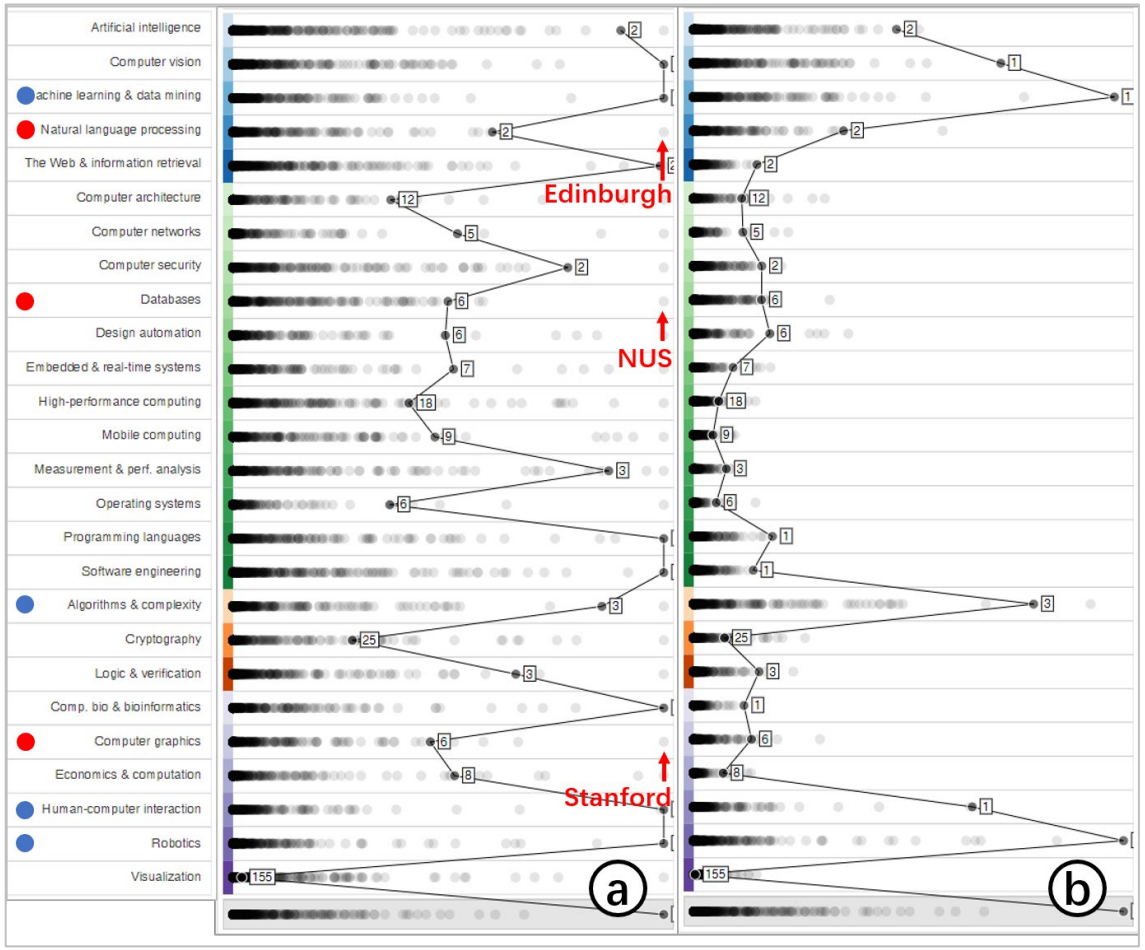


parallel coordinate axis

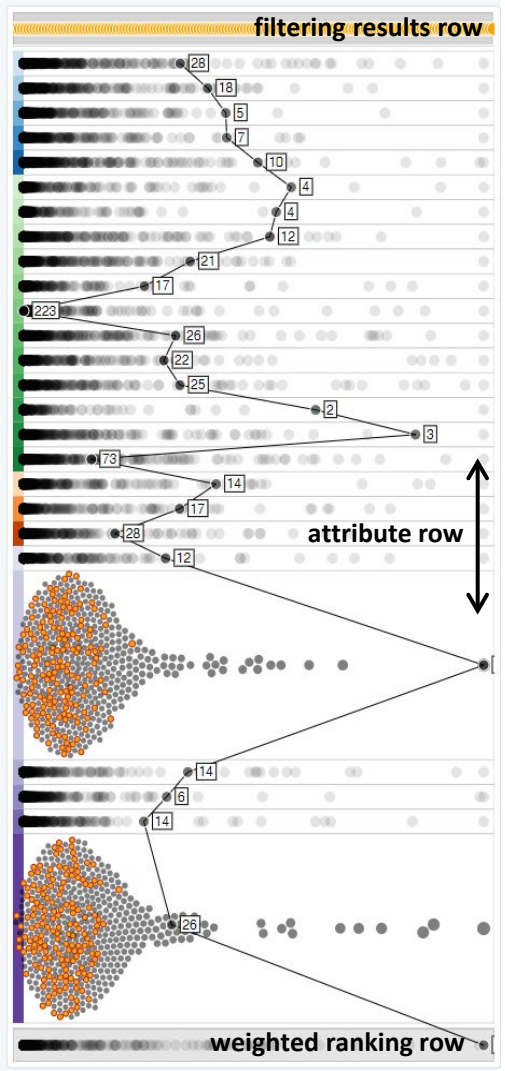
Sorted by weight scores

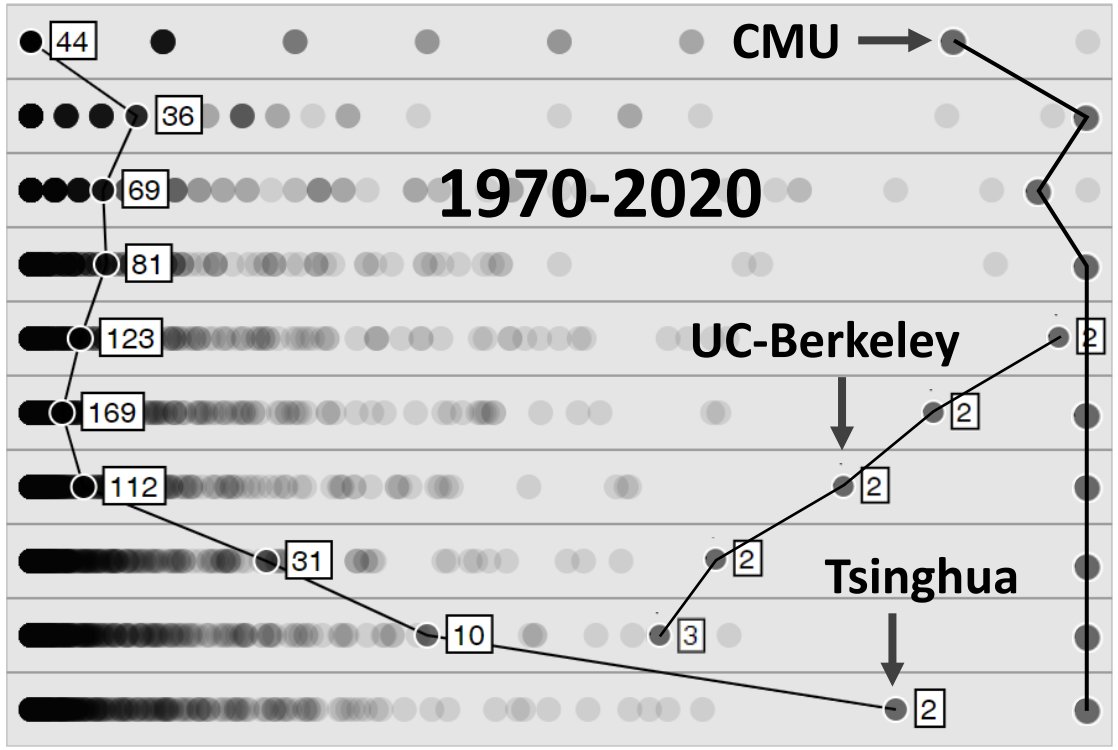
local scale

global scale

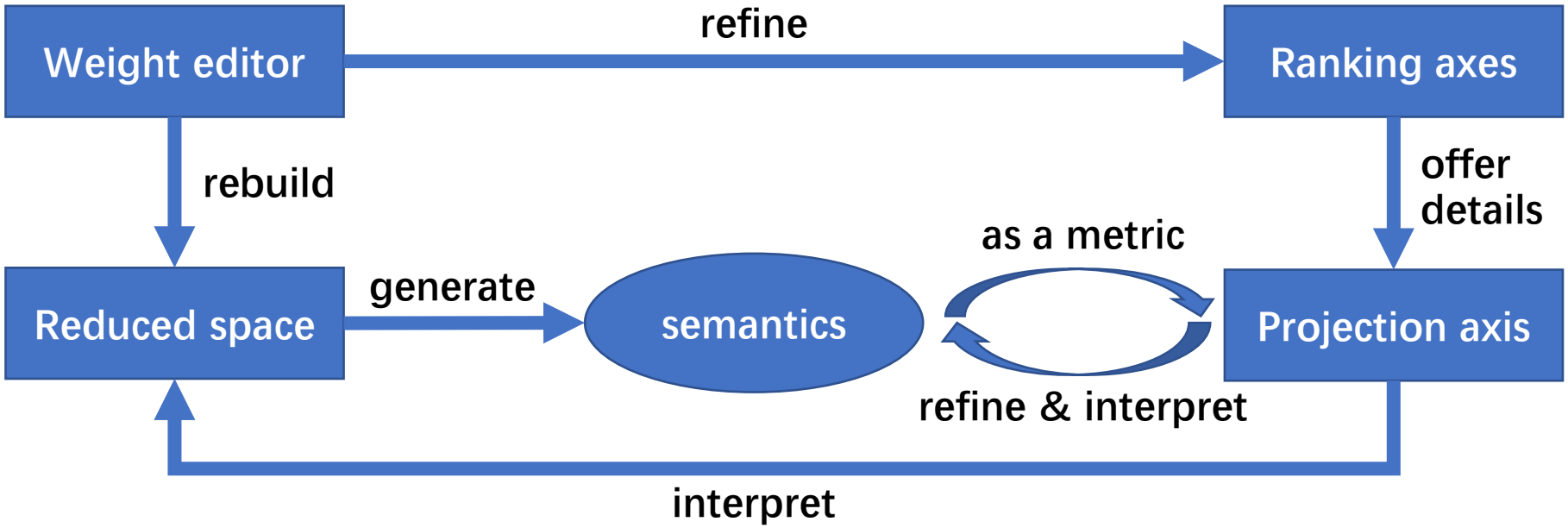


Ranking rows



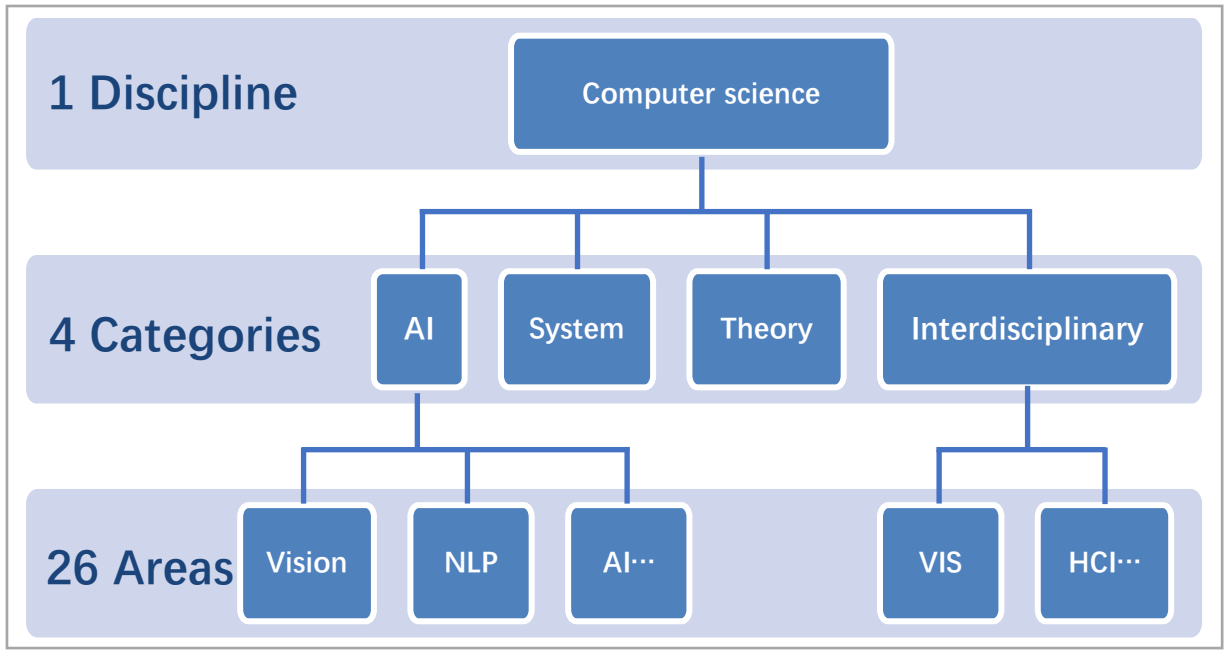


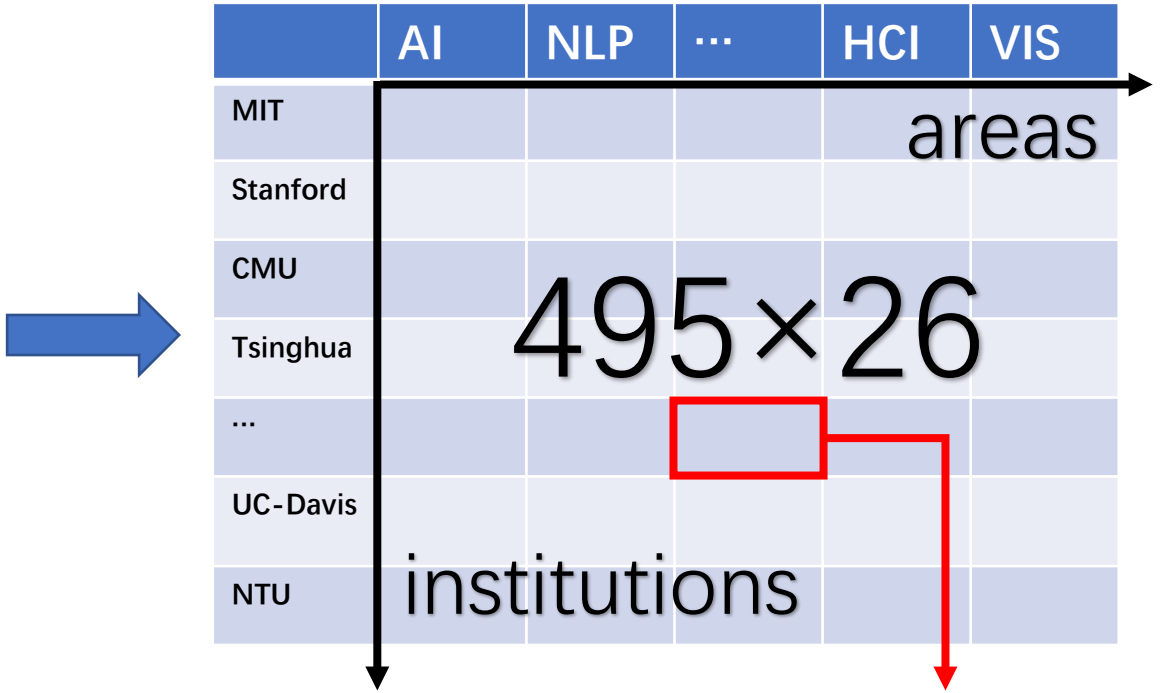
Relationships between the four main components of our system



Case: explore the academic performance and rankings of the institutions in computer science

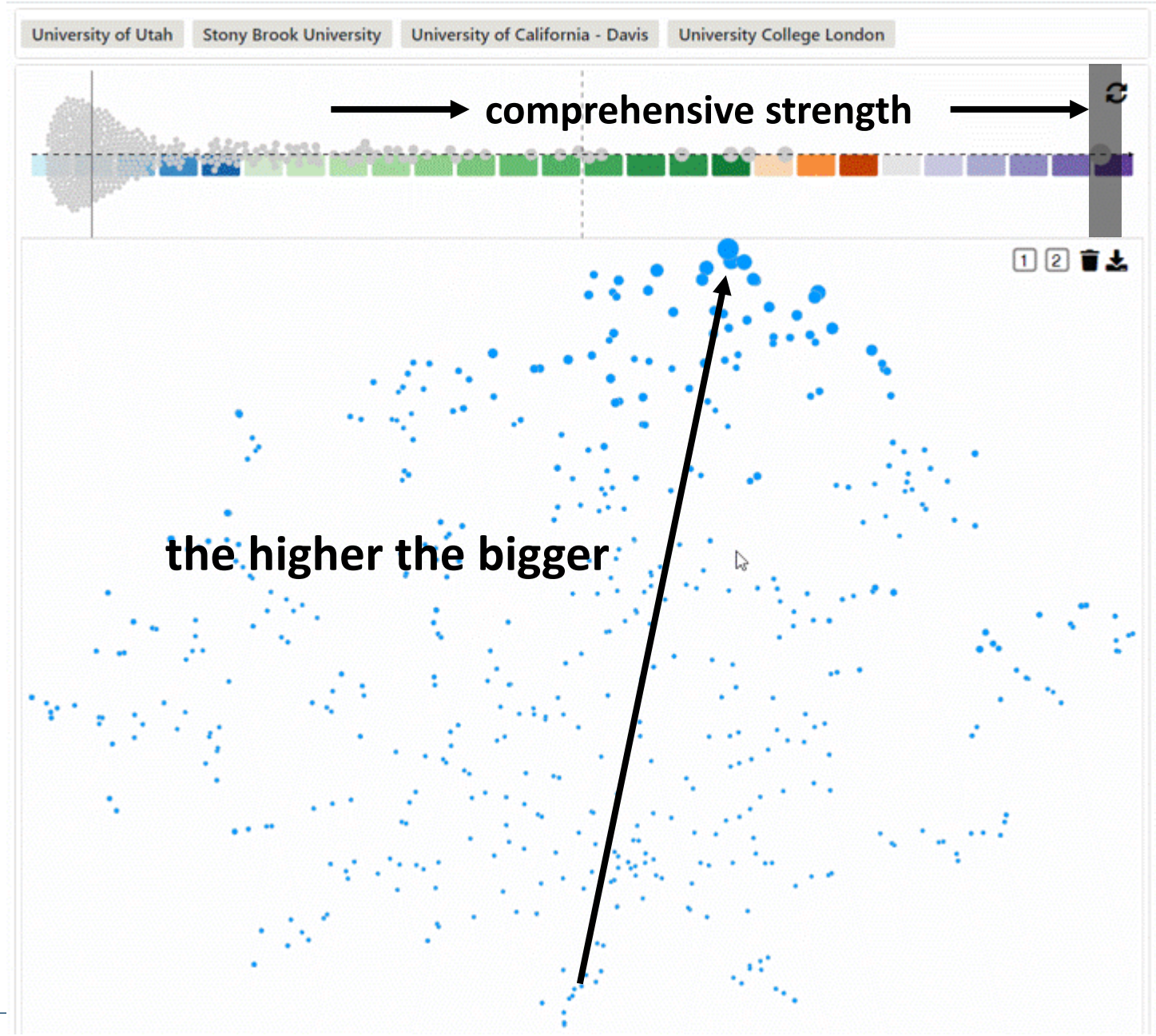
Data description

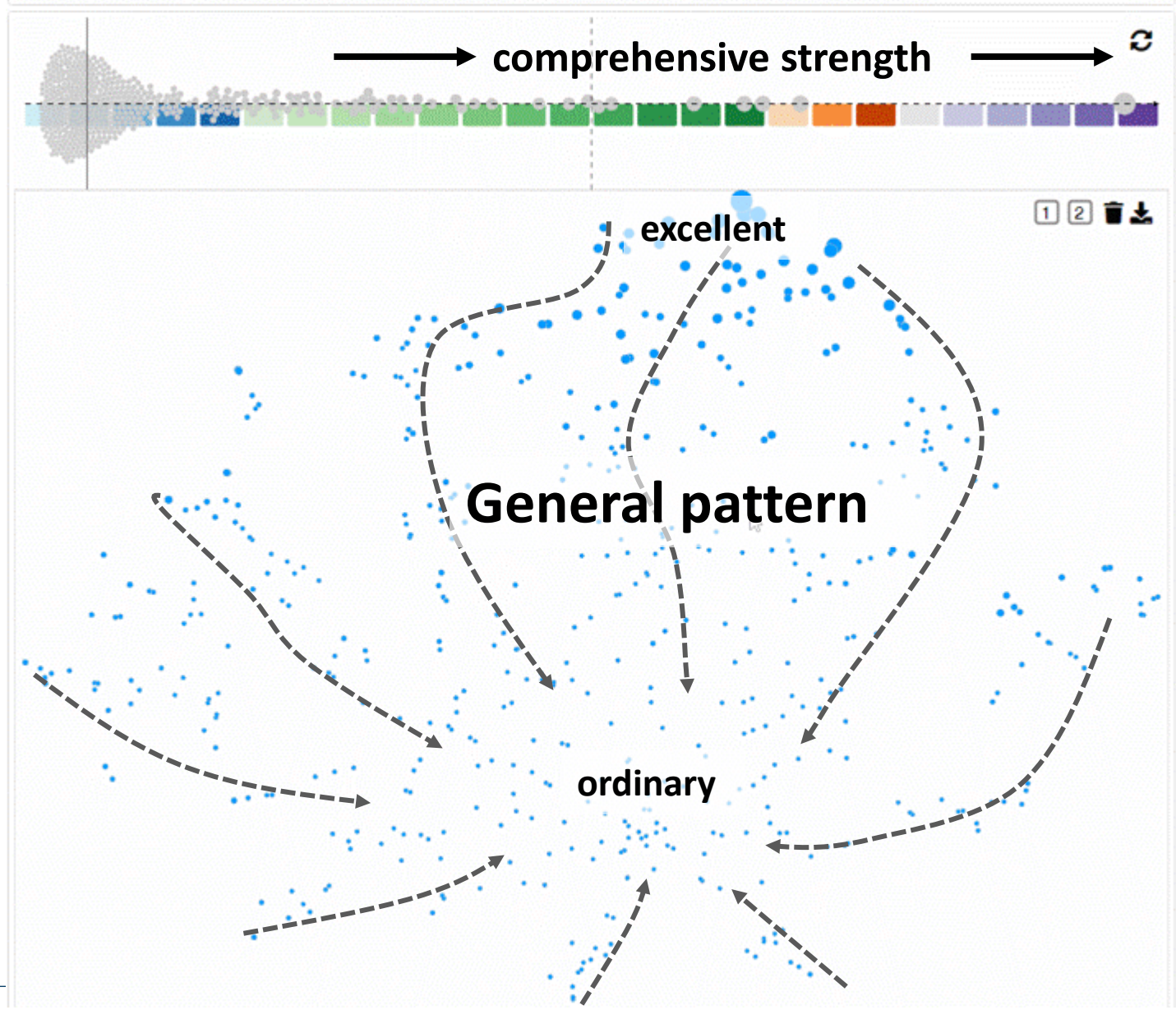


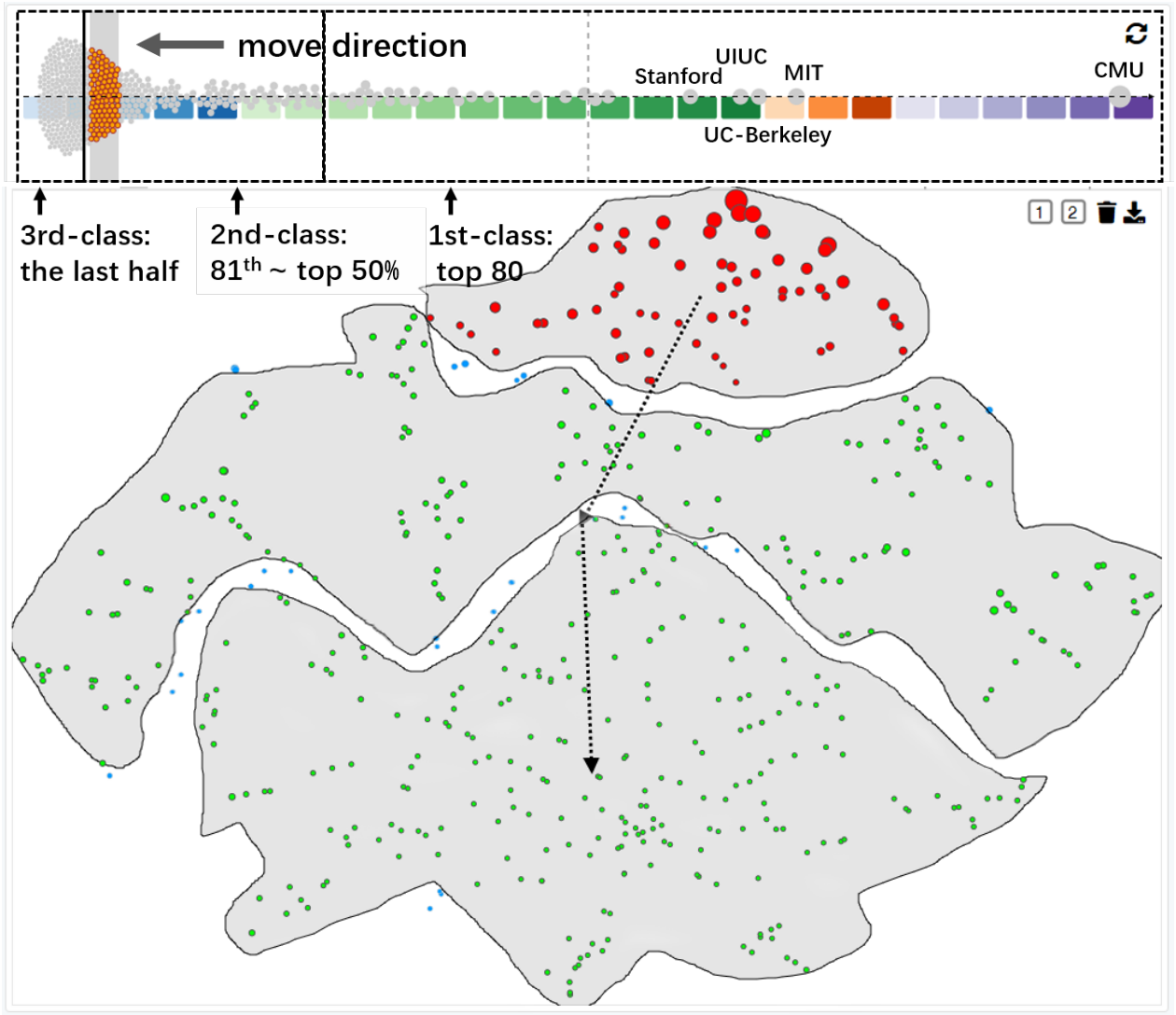


Source: CSRankings

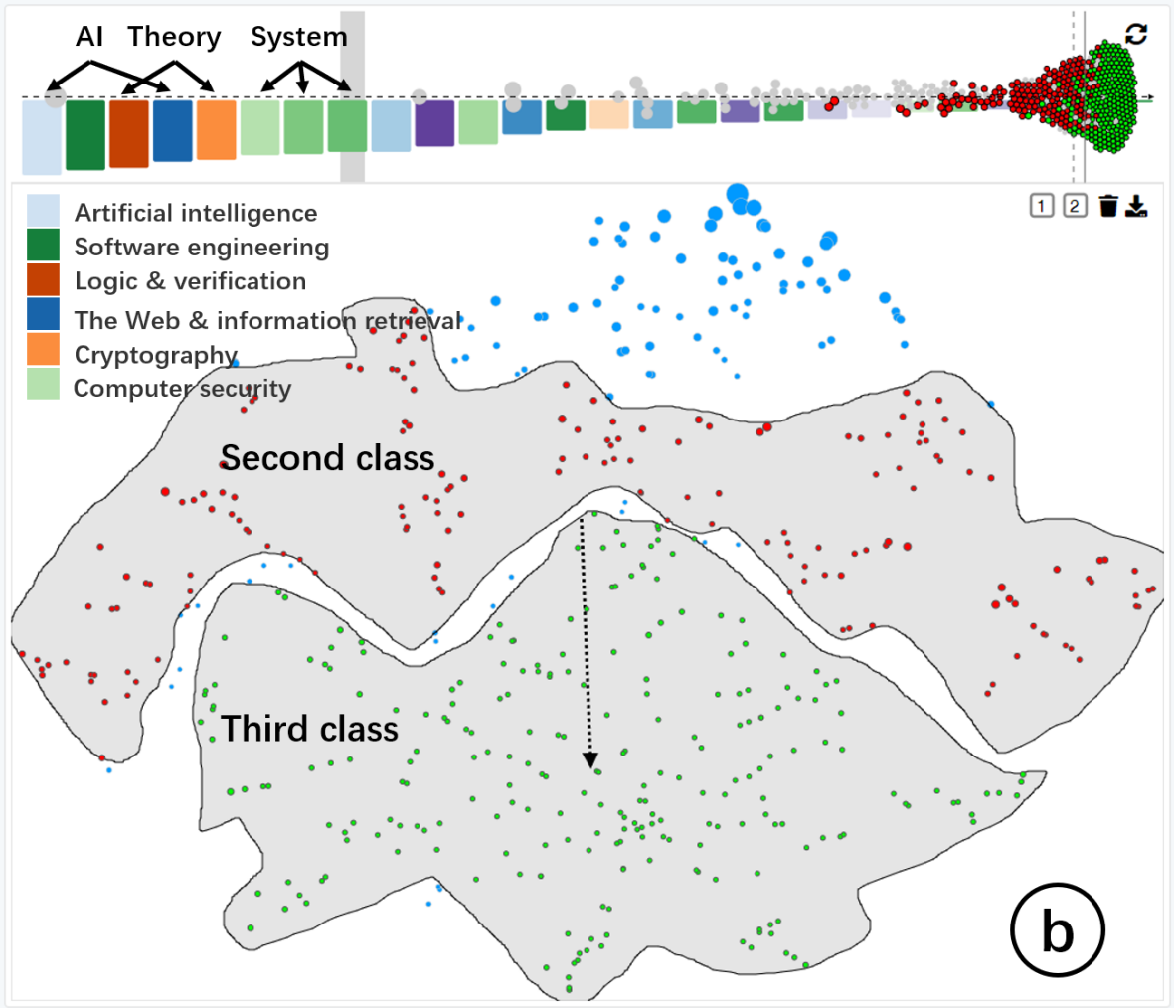
 # published paper

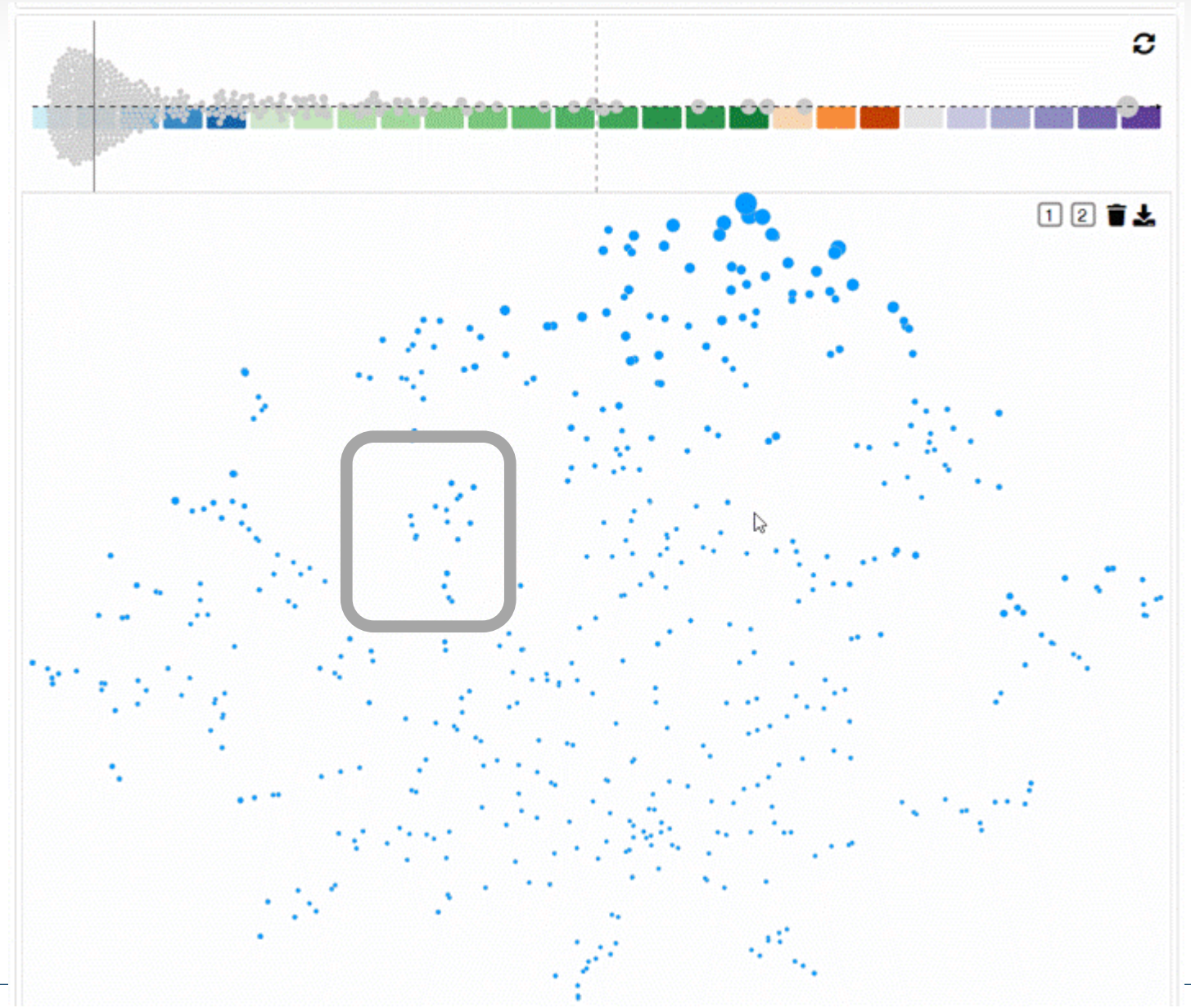

University of Utah
Stony Brook University
University of California - Davis
University College London
comprehensive strength
the higher the bigger

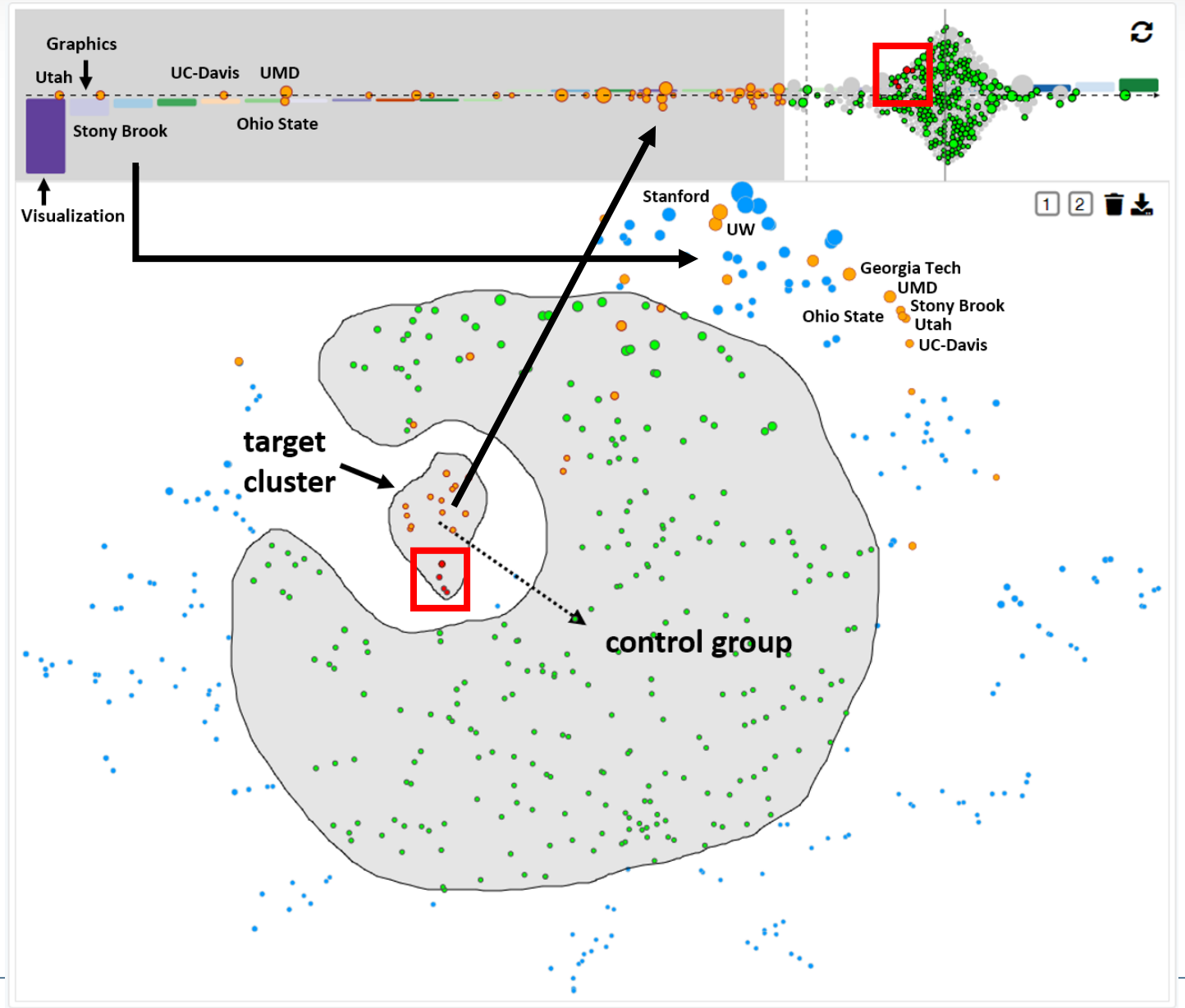

comprehensive strength
excellent
General pattern
ordinary

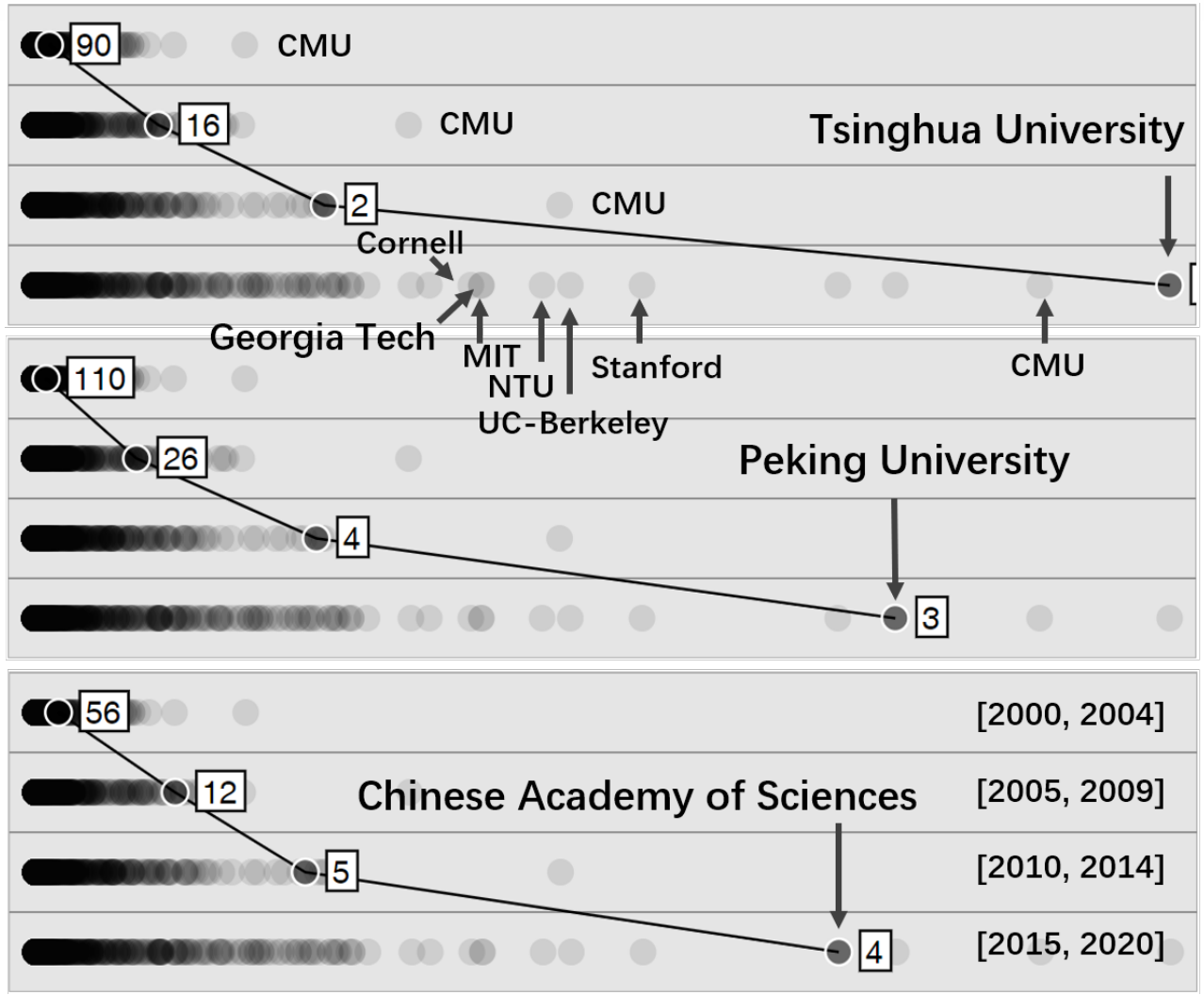

move direction
Stanford
UIUC
MIT
CMU
UC-Berkeley
3rd-class:
the last half
2nd-class:
81th ~ top 50%
1st-class:
top 80


System category
AI category
High-performance computing
Databases
Measurement & perf. analysis
Computer security
Mobile computing
Programming languages
First class
Second class
a


AI
Theory
System
Artificial intelligence
Software engineering
Logic & verification
The Web & information retrieval
Cryptography
Computer security
Second class
Third class
b


1
2


Graphics
Utah
UC-Davis
UMD
Stony Brook
Ohio State
Visualization
Stanford
UW
Georgia Tech
UMD
Ohio State
Stony Brook
Utah
UC-Davis
target
cluster
control group


90
CMU
16
CMU
Tsinghua University
2
CMU
Cornell
Georgia Tech
MIT
NTU
UC-Berkeley
Stanford
CMU
110
26
Peking University
4
3
56
[2000, 2004]
12
Chinese Academy of Sciences
[2005, 2009]
5
[2010, 2014]
4
[2015, 2020]

Discussion

Non-linear semantics

The linearity of the our semantic axis makes it impossible to accurately describe the nonlinear semantics. But it gives a linear approximation of the nonlinear semantics, which tells the analyst what attributes are associated with the target semantics.

The choice of DR algorithm

Any DR algorithm can be used in the reduced space.
For linear DR algorithms, such as PCA, …
For nonlinear DR algorithms, such as t-SNE and UMAP, …

Conclusion

We present **semanticAxis**, a multi-attribute data analysis technique based on the idea of comparative analysis, which combines dimensionality reduction result understanding and multi-attribute data rankings into a unified exploration context.

We also designed a **visual analysis system** with semanticAxis as the core, which demonstrated the effectiveness of semanticAxis, and the rich components and interactions extended its analytical capabilities in complex analysis scenarios.

Due to its linear nature, semanticAxis does not precisely measure and interpret **nonlinear semantics**. But information provided by other components can alleviate this problem.