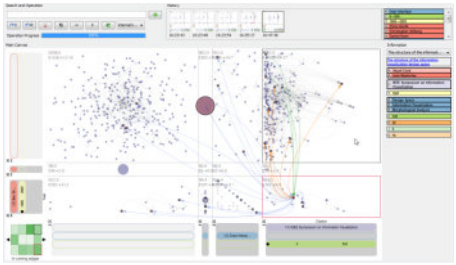
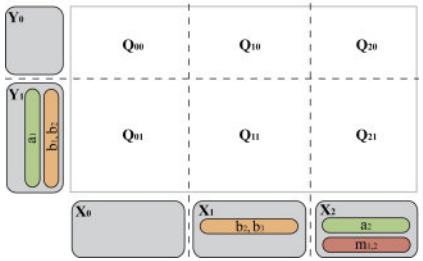
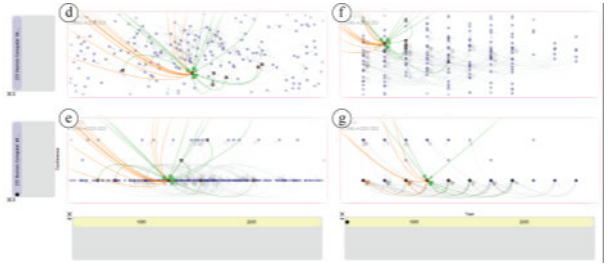






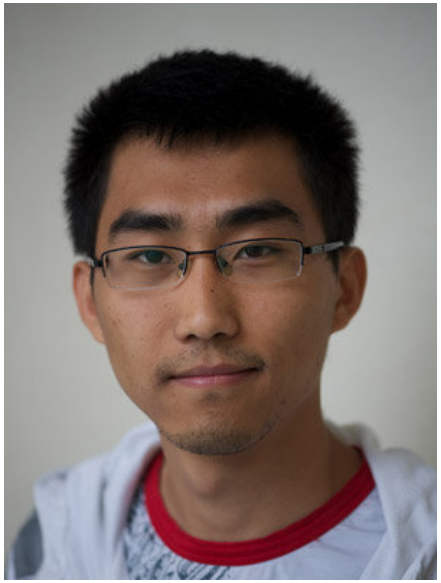
Interactive Exploration of Implicit and Explicit Relations in Faceted Datasets

Jian Zhao, Christopher Collins, Fanny Chevalier, and Ravin Balakrishnan

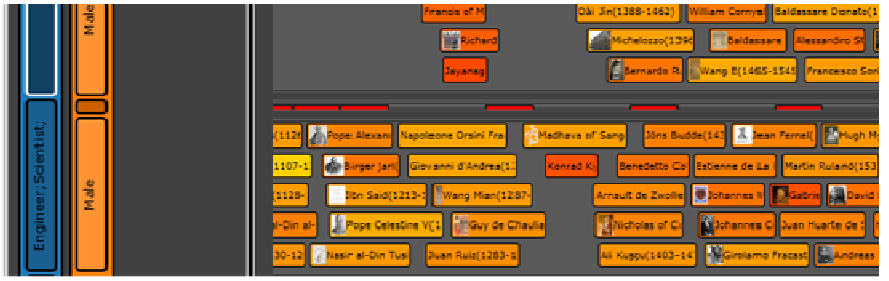
UNIVERSITY OF TORONTO

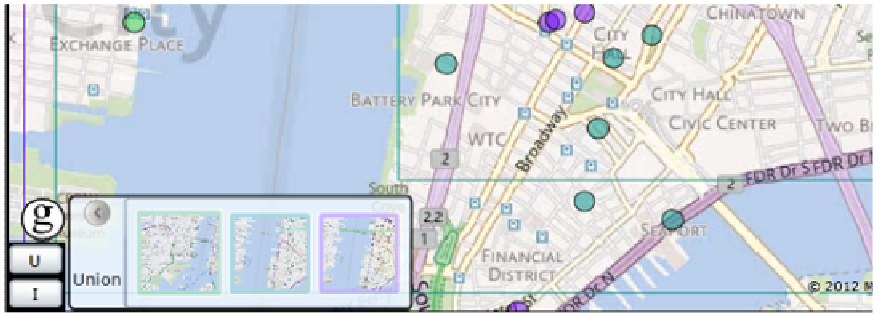
UOIT
CHALLENGE INNOVATE CONNECT

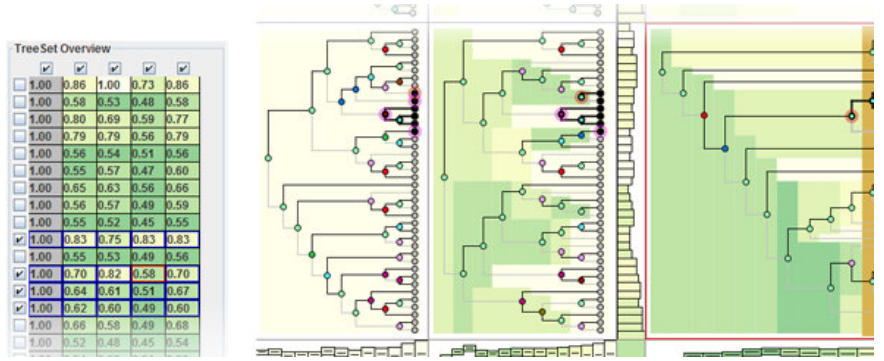
Jian Zhao



www.cs.toronto.edu/~jianzhao/

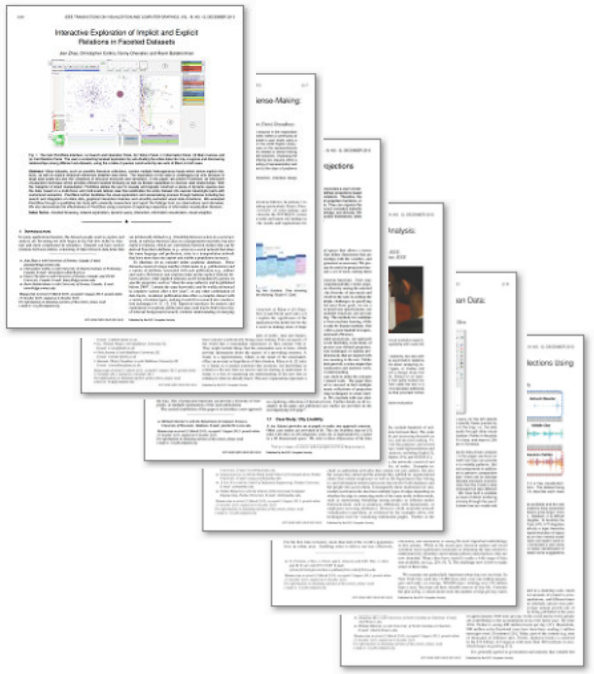






Faceted Document Data

- Large in *Size*



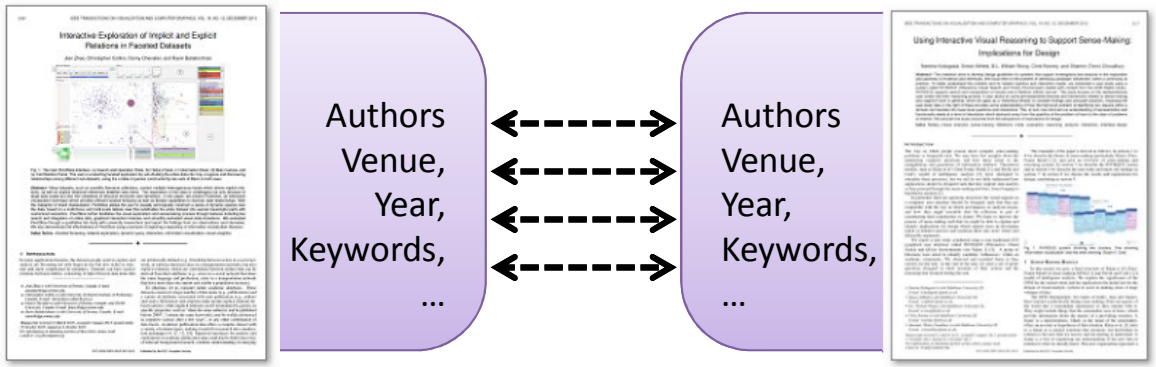
Faceted Document Data

- Large in *Size*
- Rich in *Structure*
  - Heterogeneous facets (attributes)
  - Numerical, categorical, and ordinal

Authors
Venue,
Year,
Keywords,
…

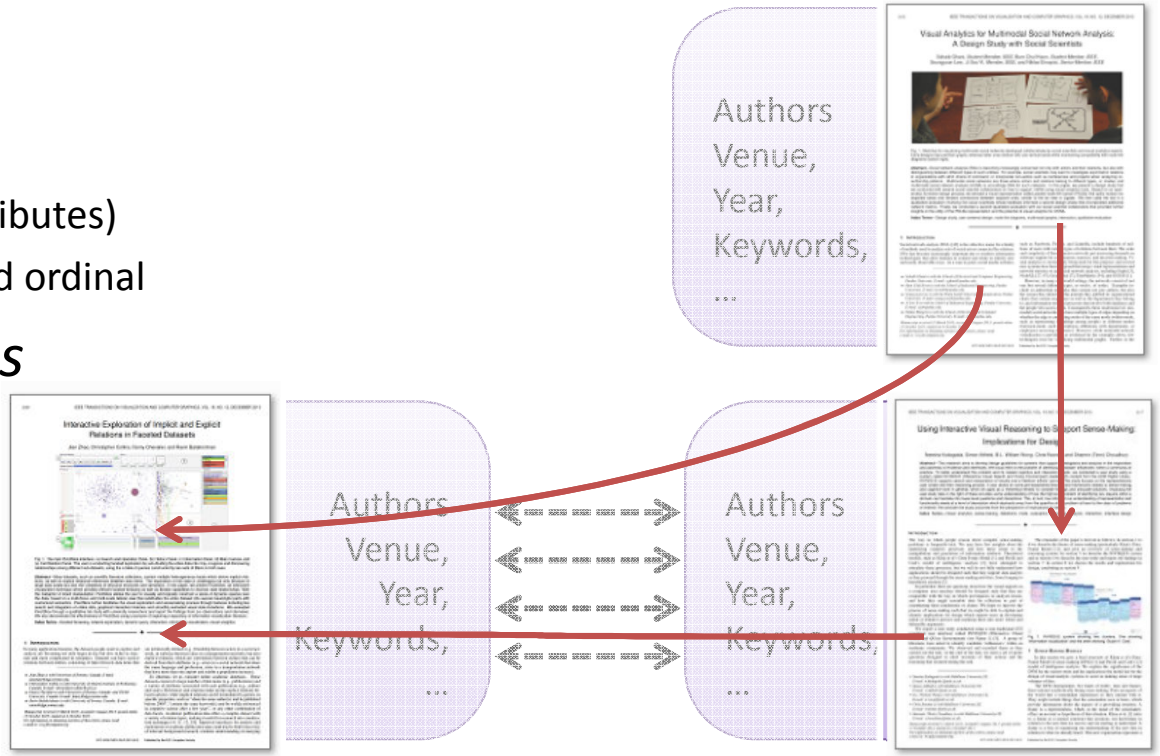
Faceted Document Data

- Large in *Size*
- Rich in *Structure*
  - Heterogeneous facets (attributes)
  - Numerical, categorical, and ordinal
- Complex in *Semantics*
  - Implicit relationships

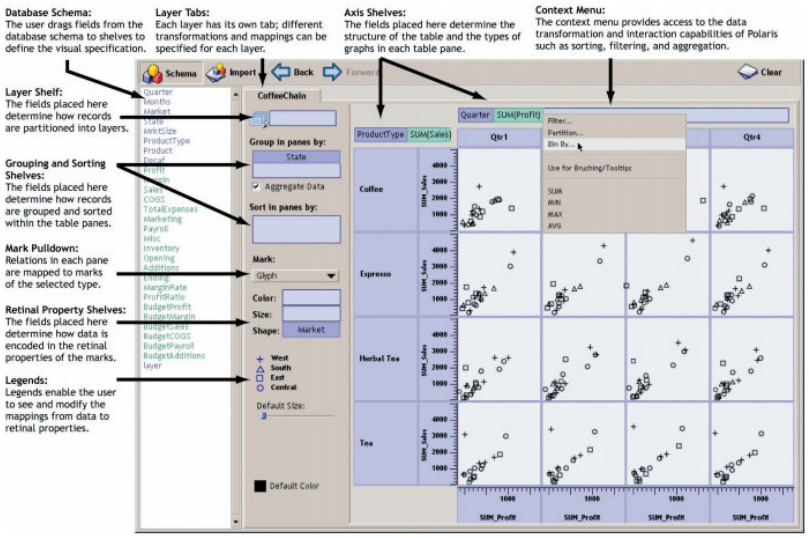


Faceted Document Data

- Large in *Size*
- Rich in *Structure*
  - Heterogeneous facets (attributes)
  - Numerical, categorical, and ordinal
- Complex in *Semantics*
  - Implicit relationships
  - Explicit references



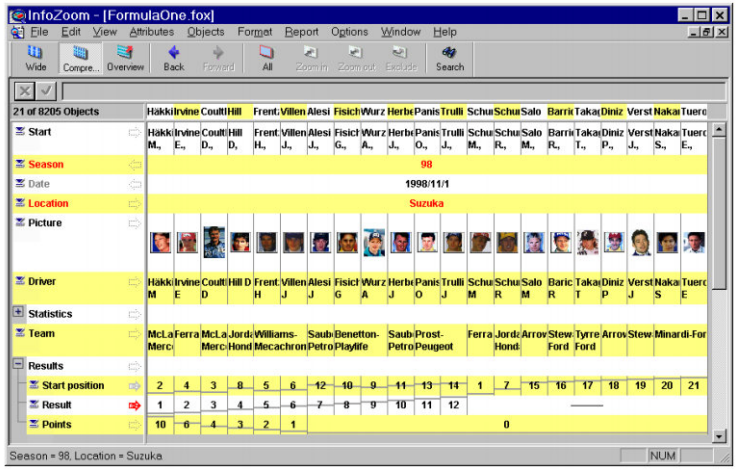
Related Work - Polaris



- Visual query language
- Direct manipulation

C. Stolte, D. Tang, and P. Hanrahan. Polaris: A system for query, analysis, and visualization of multidimensional relational databases. *IEEE Trans. on Visualization and Computer Graphics*, 8(1):52–65, 2002.

Related Work - InfoZoom



- Filter on attribute values
- Compressed views

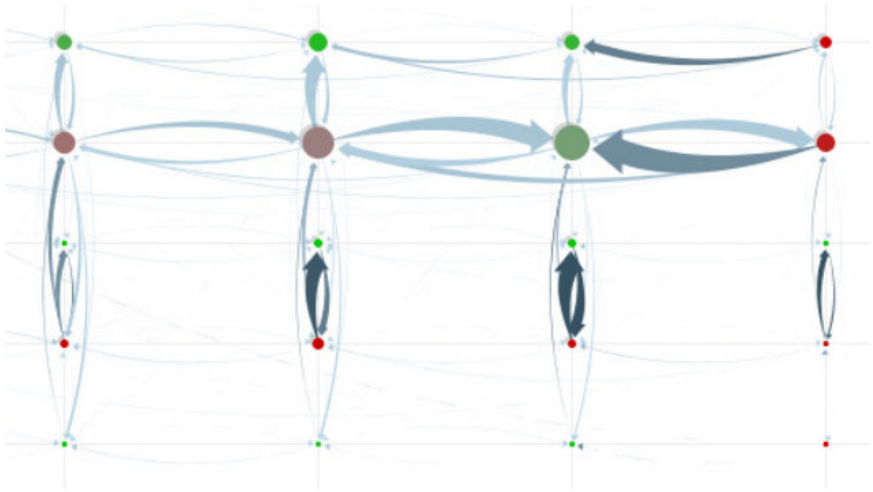
M. Spenke and C. Beilken. Infozoom - Analysing formula one racing results with an interactive data mining and visualisation tool. In *Proceedings of Data Mining*, pages 455–464, 2000.

Related Work – Semantic Substrates

- Partitioned dataset
- Links within and across canvas areas

B. Shneiderman and A. Aris. Network visualization by semantic substrates. *IEEE Trans. on Visualization and Computer Graphics*, 12(5):733–740, 2006.

Related Work – Pivot Graphs

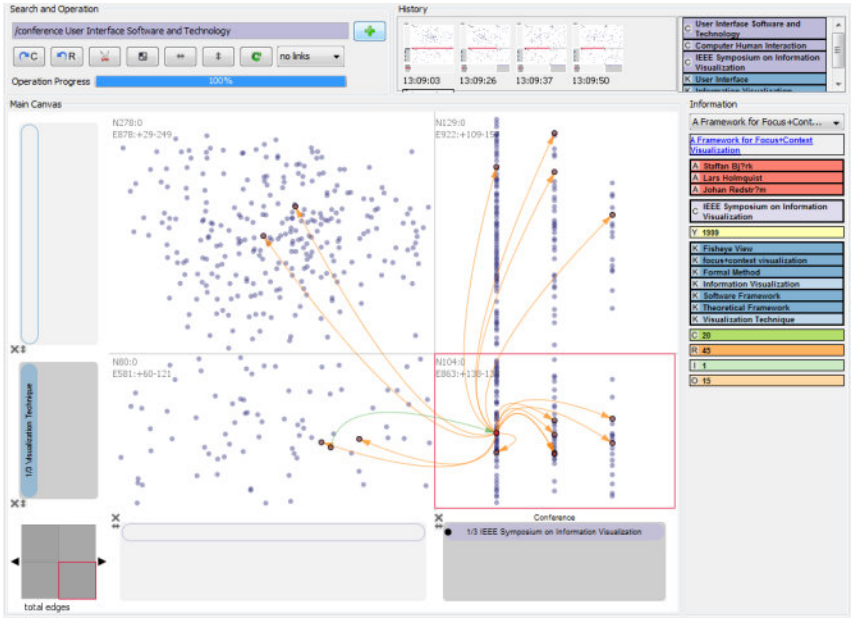


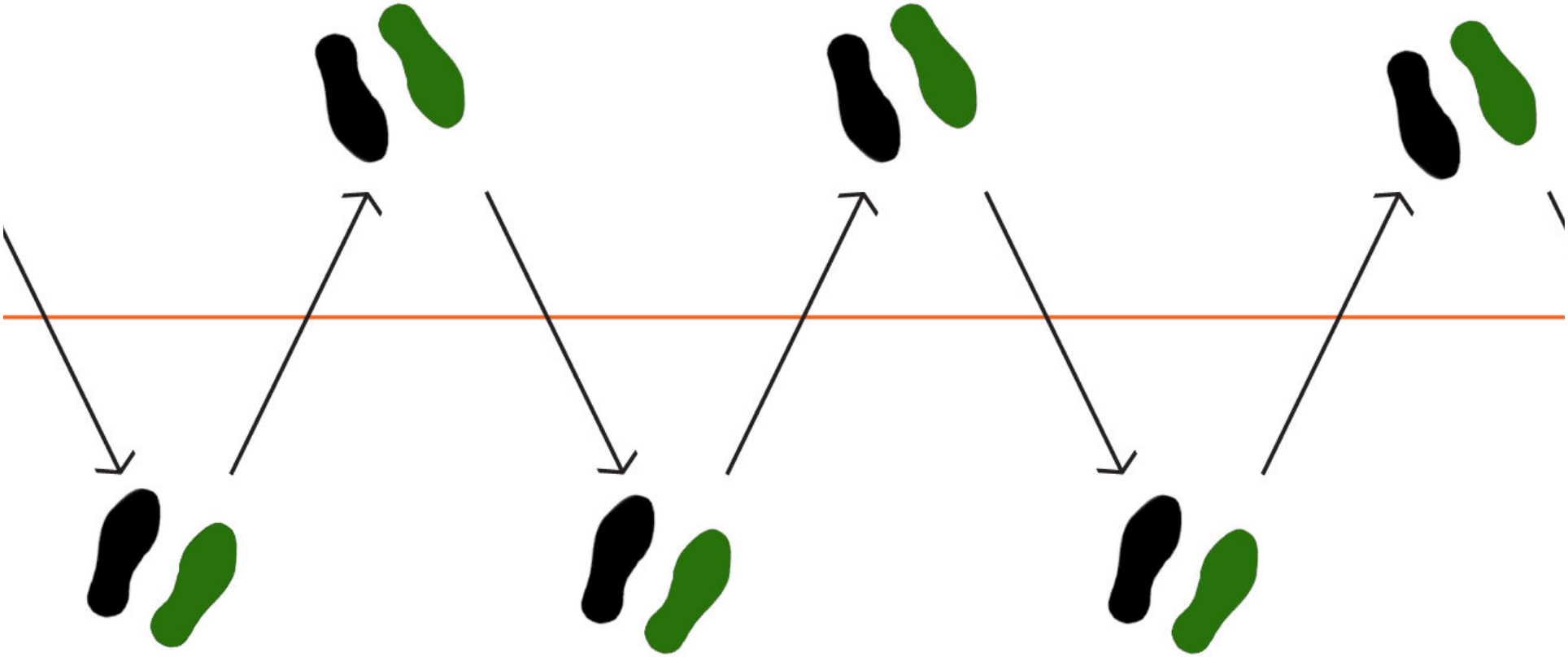
- Attribute-based organization
- Roll-up of nodes

M. Wattenberg. Visual exploration of multivariate graphs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 811–819, 2006.

PivotSlice

- **Data:** scholarly publications from an online database
- **Concept:** documents arranged as graph nodes in a sliceable workspace
- **Queries:** visual queries based on implicit relations (document meta-data)
- **Relations:** explicit relations revealed (citations and references)

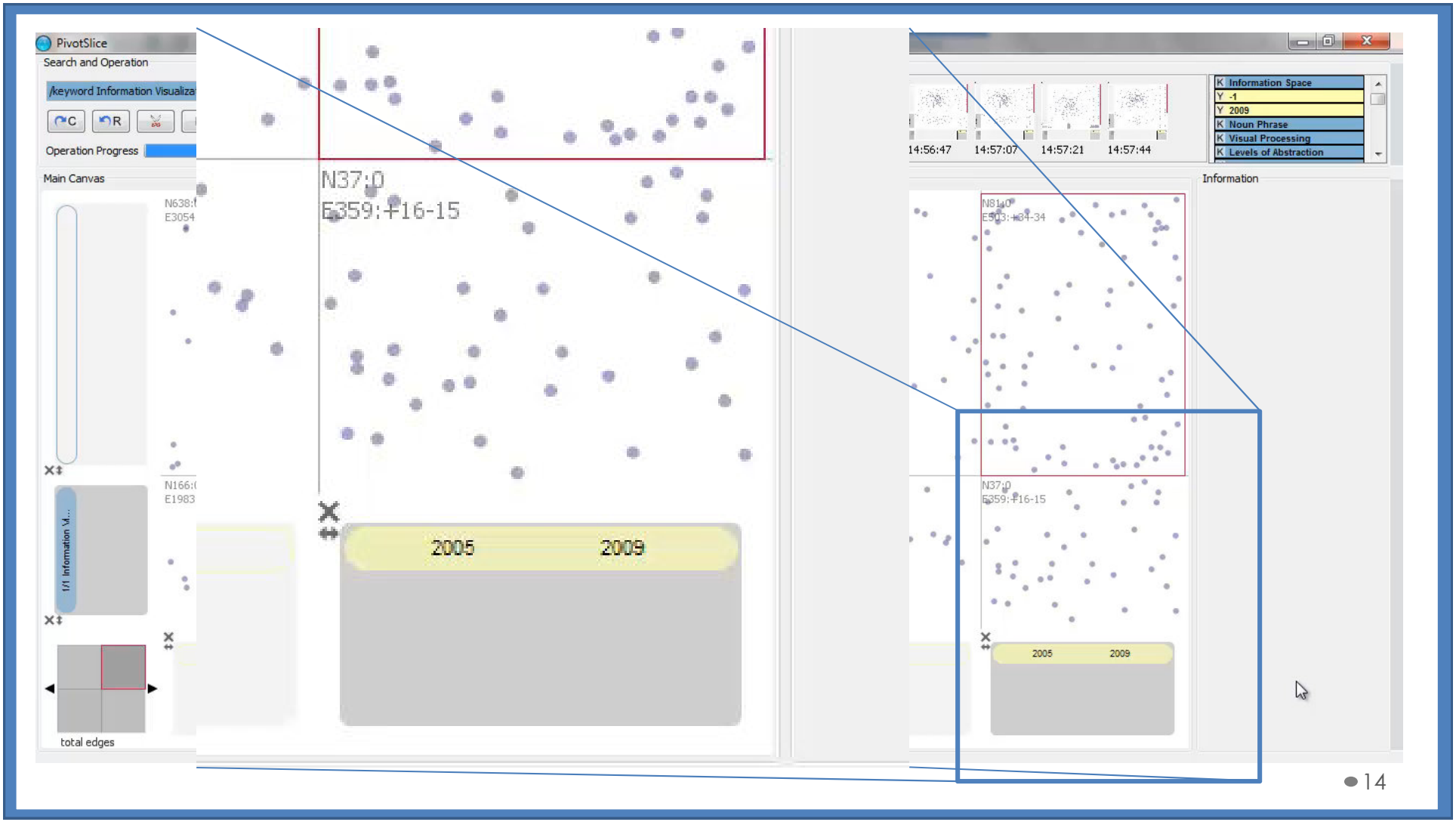




PivotSlice

Pivot

- Aggregate, sort, or align data items with any faceted attributes on the two axes


PivotSlice
Search and Operation
/keyword Information Visualiza
C
R
Operation Progress
Main Canvas
N638:
E3054
N166:
E1983
1/1 Information Vi...
total edges
N37:0
E359:+16-15
2005
2009
14:56:47
14:57:07
14:57:21
14:57:44
Information Space
-1
2009
Noun Phrase
Visual Processing
Levels of Abstraction
Information
N814:0
E503:+84-34
N37:0
E359:+16-15
2005
2009

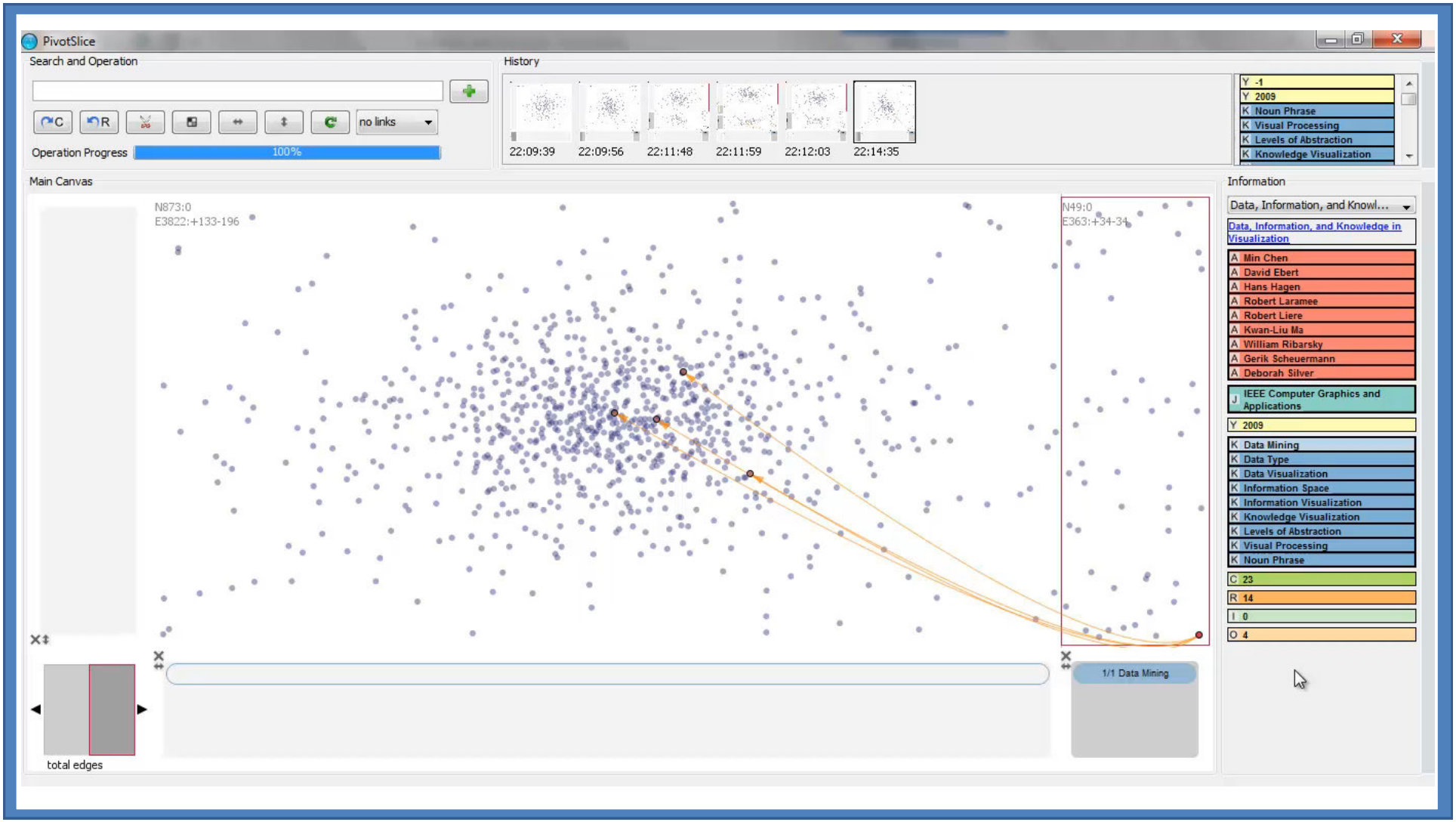
PivotSlice

Pivot

- Aggregate, sort, or align data items with any faceted attributes on the two axes

Slice

- Distribute the information space into a tabular view of sub-regions based on logical queries


PivotSlice
Search and Operation
C
R
no links
Operation Progress
100%
History
22:09:39
22:09:56
22:11:48
22:11:59
22:12:03
22:14:35
Y -1
Y 2009
K Noun Phrase
K Visual Processing
K Levels of Abstraction
K Knowledge Visualization
Main Canvas
N873:0
E3822:+133-196
N49:0
E363:+34-34
1/1 Data Mining
total edges
Information
Data, Information, and Knowl...
Data, Information, and Knowledge in Visualization
A Min Chen
A David Ebert
A Hans Hagen
A Robert Laramee
A Robert Liere
A Kwan-Liu Ma
A William Ribarsky
A Gerik Scheuermann
A Deborah Silver
J IEEE Computer Graphics and Applications
Y 2009
K Data Mining
K Data Type
K Data Visualization
K Information Space
K Information Visualization
K Knowledge Visualization
K Levels of Abstraction
K Visual Processing
K Noun Phrase
C 23
R 14
I 0
O 4

PivotSlice

Pivot

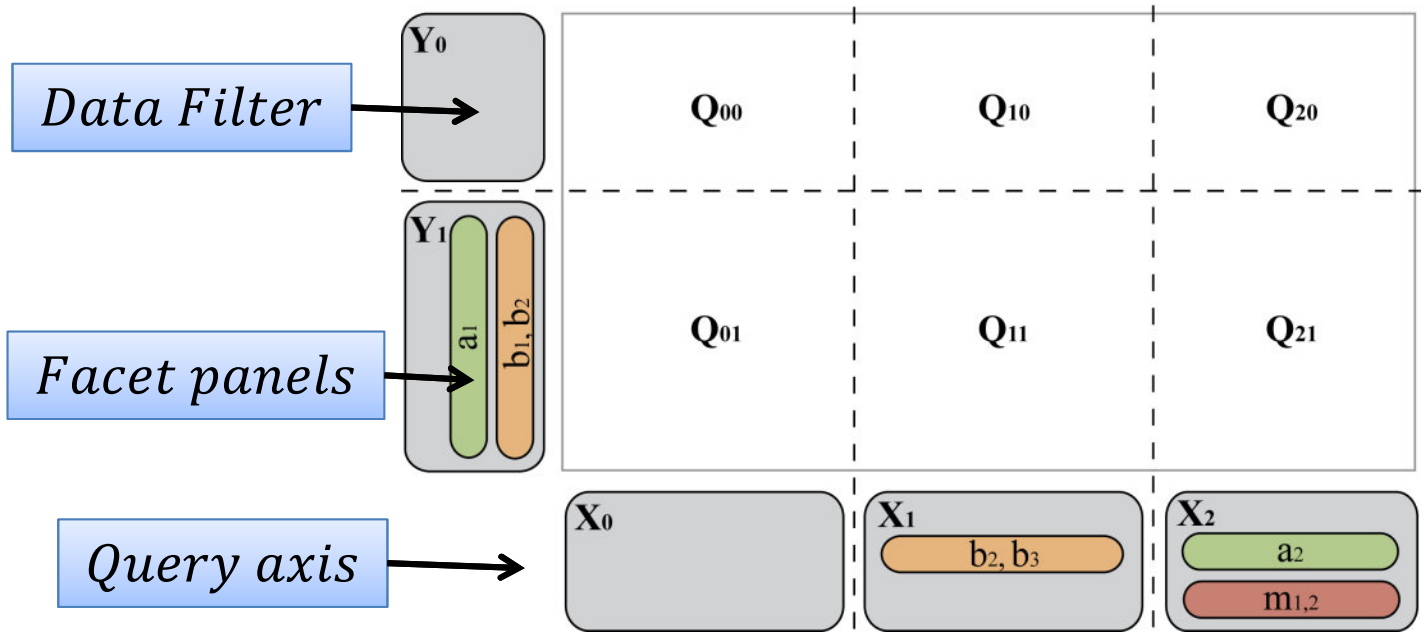
- Aggregate, sort, or align data items with any faceted attributes on the two axes

Slice

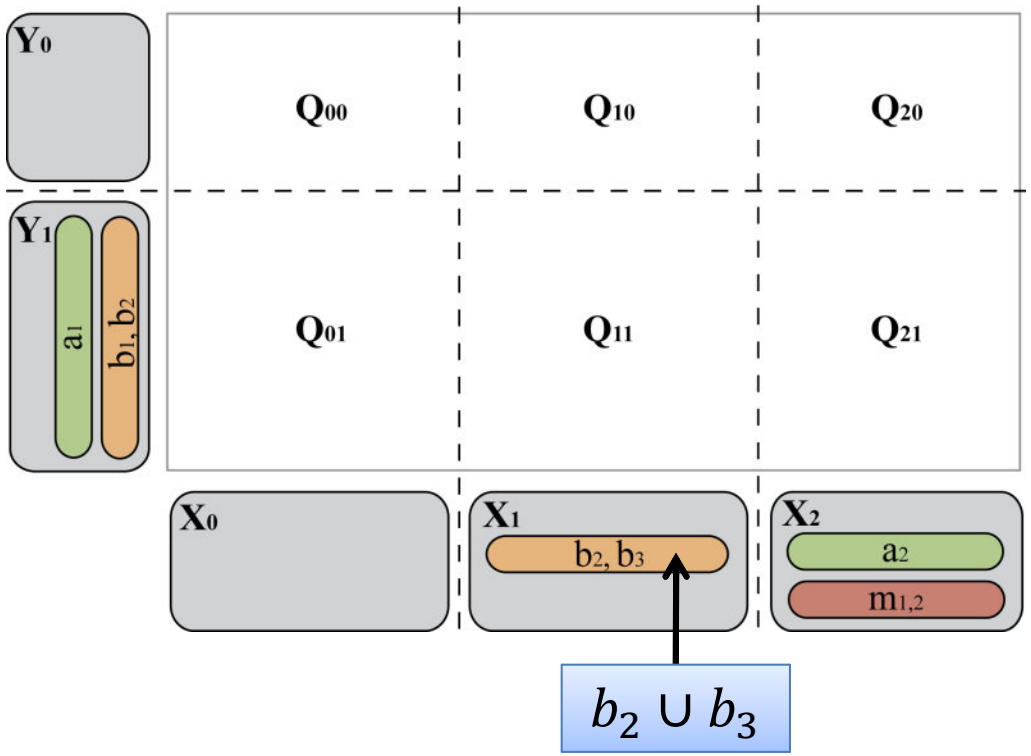
- Distribute the information space into a tabular view of sub-regions based on logical queries

- Multi-scale representation
- Direct manipulation
- Smooth animation

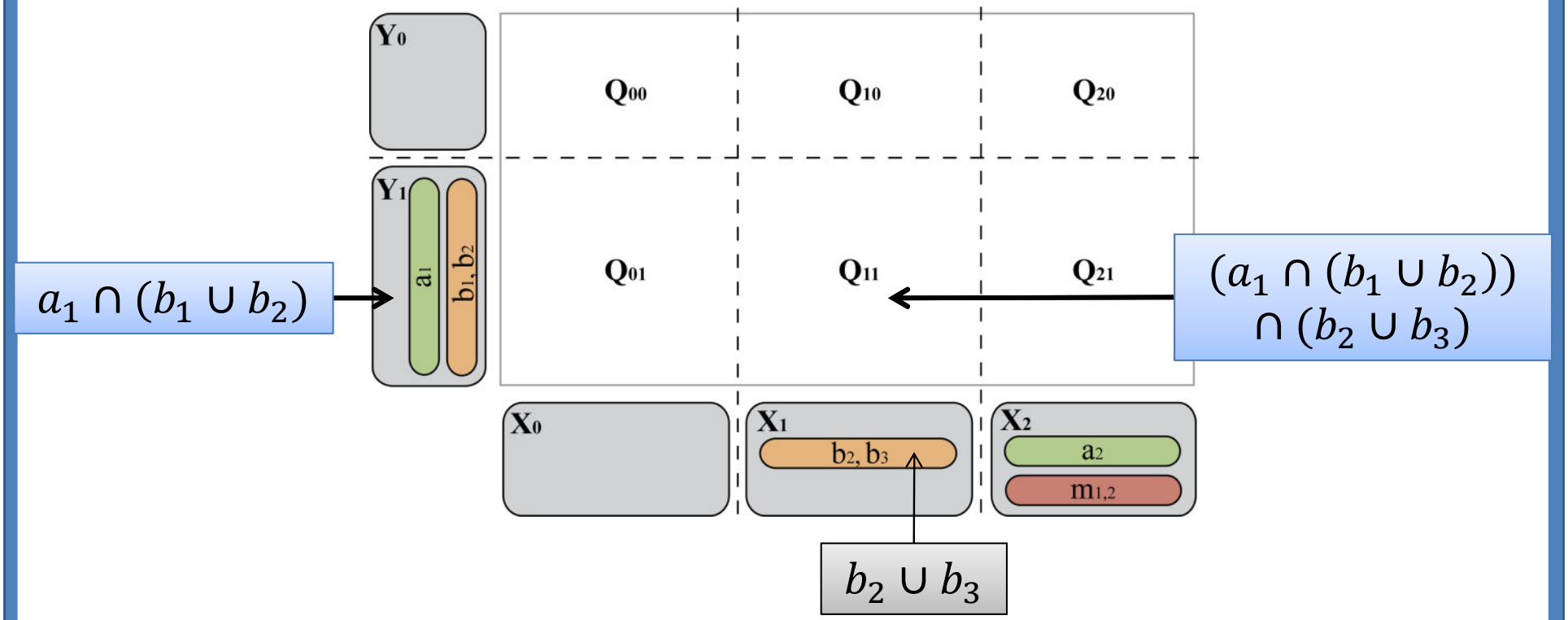
Visual Query Logic



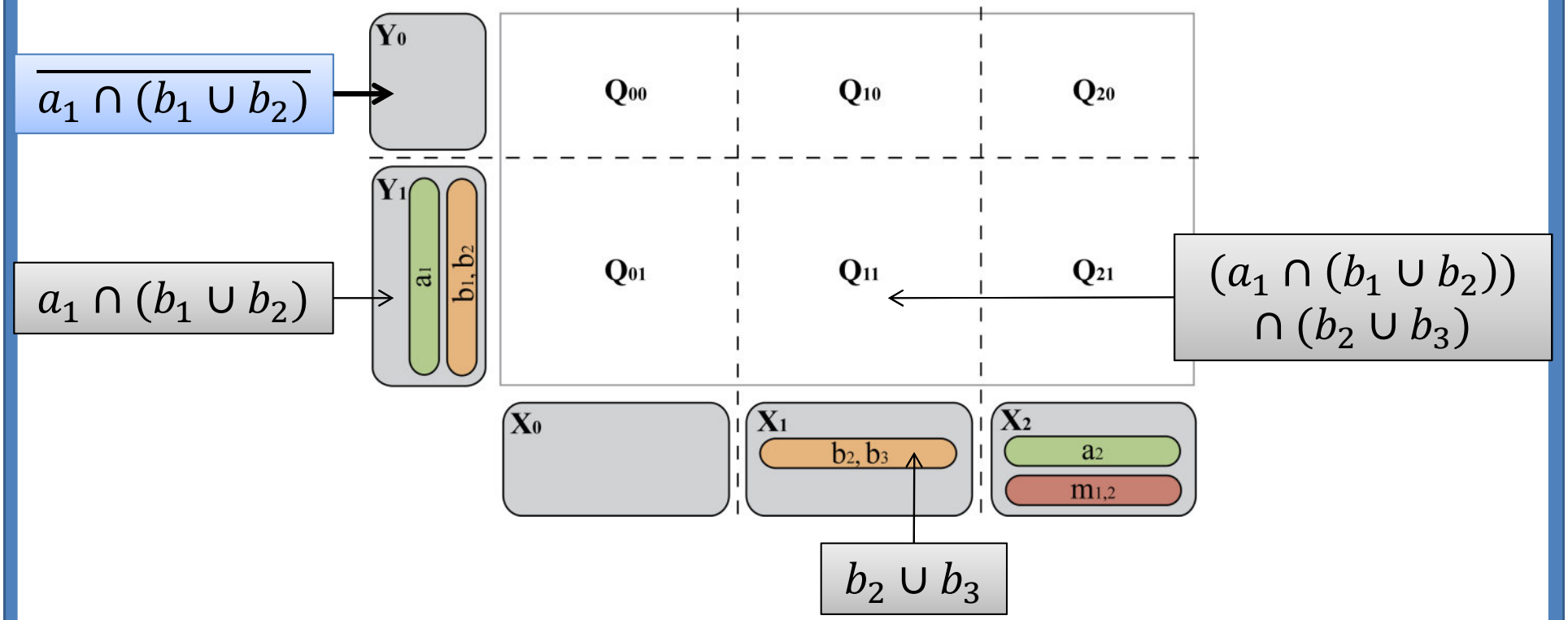
Visual Query Logic

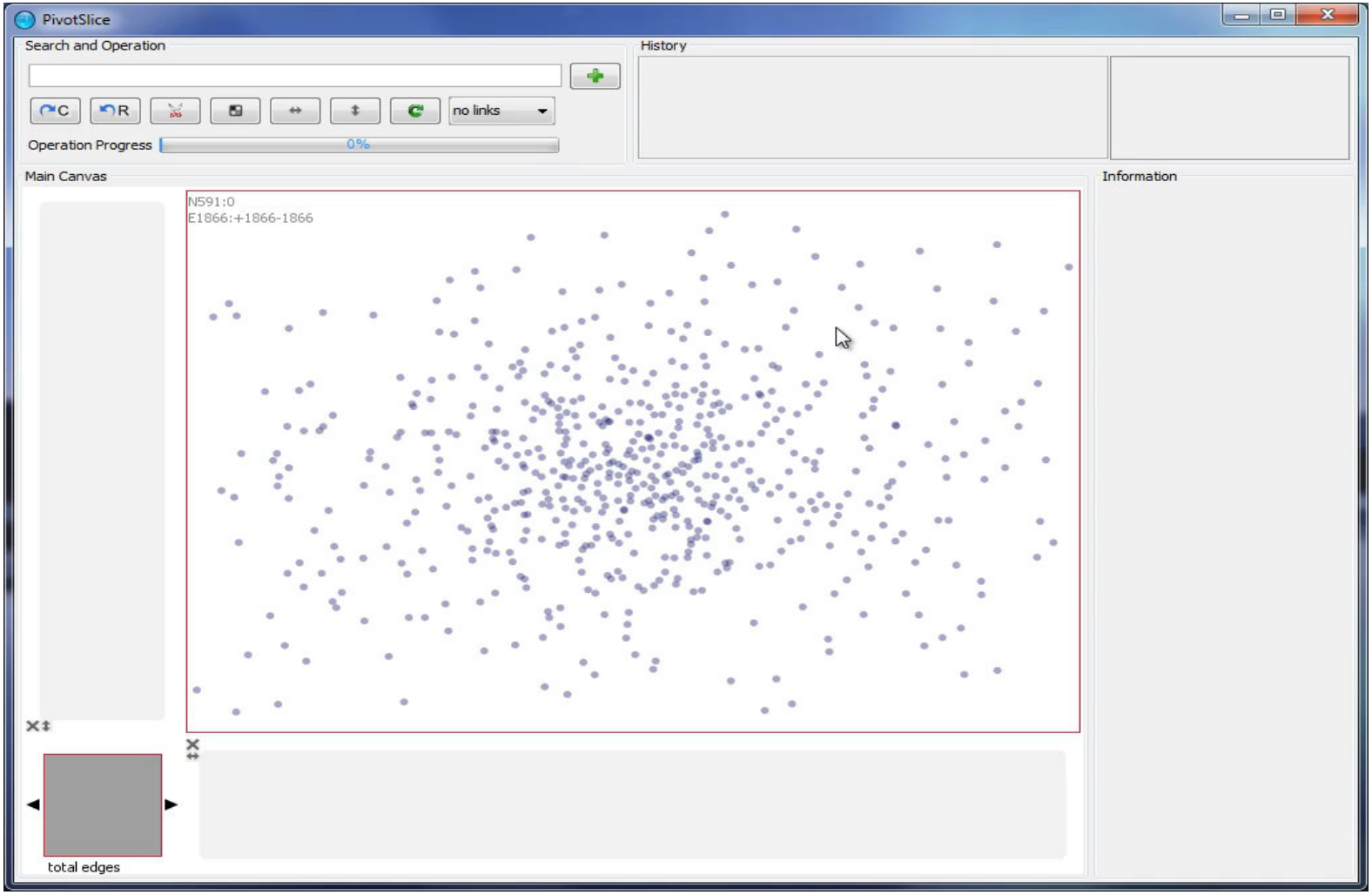


Visual Query Logic



Visual Query Logic




PivotSlice
Search and Operation
C
R
no links
Operation Progress
0%
History
Main Canvas
N591:0
E1866:+1866-1866
Information
total edges

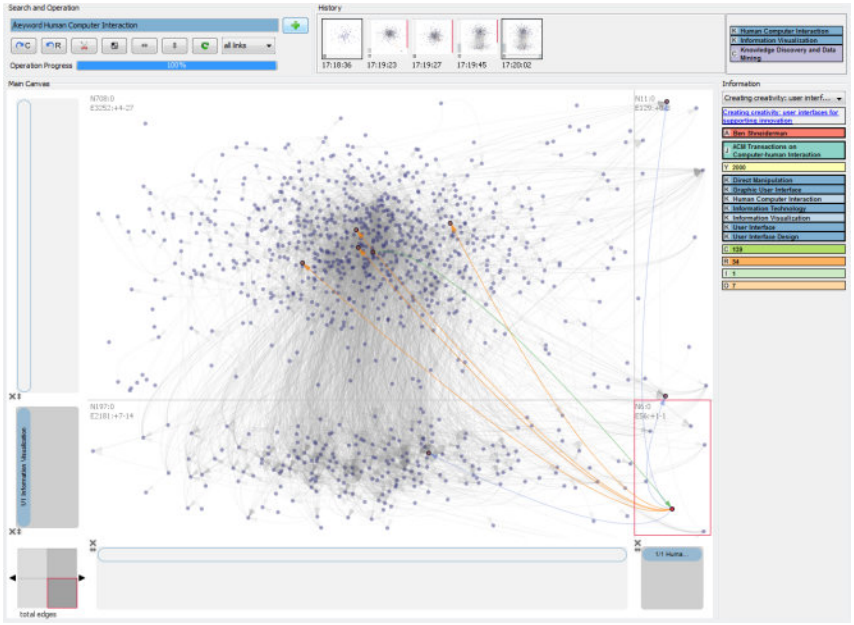
Information Seeking in Faceted Datasets

- Overview first
- Pivot and slice
- Relate and extract
- Details-on-demand
- History available

[1] B. Shneiderman. The eyes have it: a task by data type taxonomy for information visualizations. *Proceedings of IEEE Symposium on Visual Languages,* **1996**, 336-343

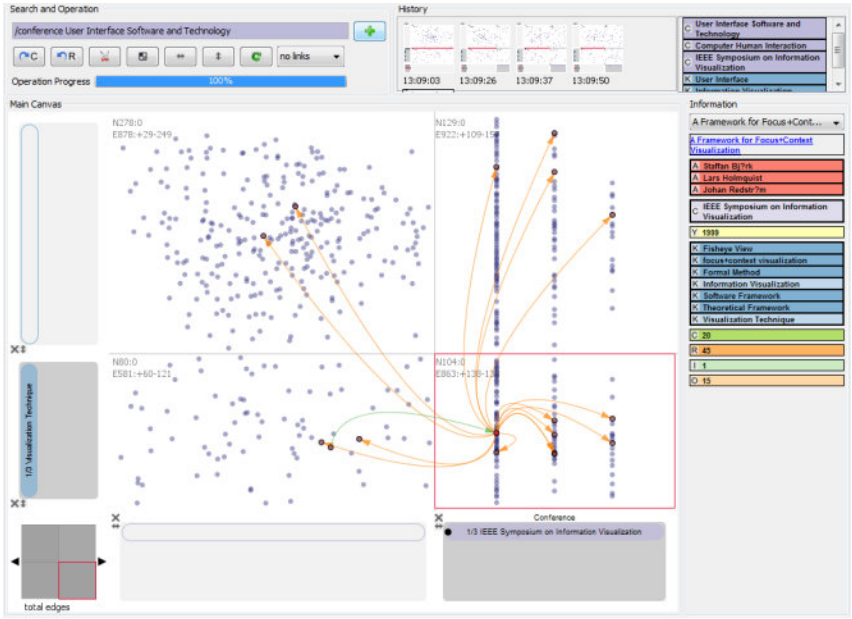
Information Seeking in Faceted Datasets

- *Overview first*
- Pivot and slice
- Relate and extract
- Details-on-demand
- History available



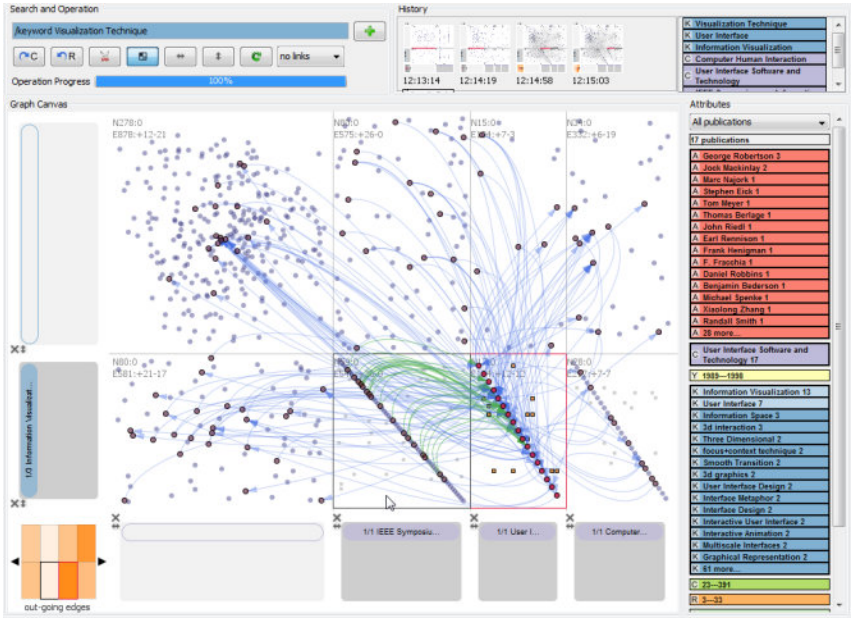
Information Seeking in Faceted Datasets

- Overview first
- *Pivot and slice*
- Relate and extract
- Details-on-demand
- History available



Information Seeking in Faceted Datasets

- Overview first
- Pivot and slice
- *Relate and extract*
- Details-on-demand
- History available

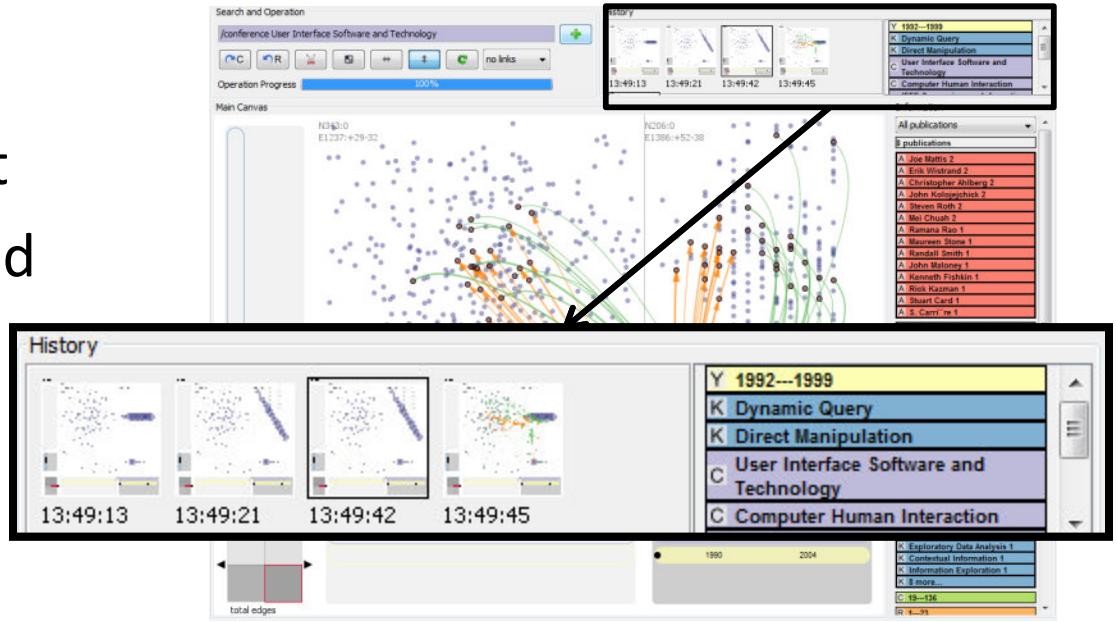


Information Seeking in Faceted Datasets

- Overview first
- Pivot and slice
- Relate and extract
- *Details-on-demand*
- History available

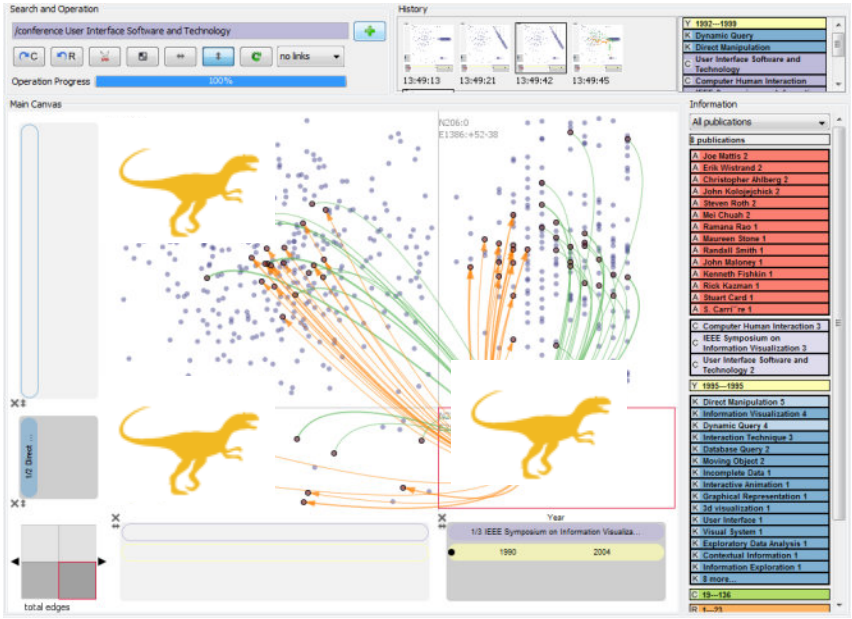
Information Seeking in Faceted Datasets

- Overview first
- Pivot and slice
- Relate and extract
- Details-on-demand
- *History available*



Information Seeking in Faceted Datasets

- Overview first
- Pivot and slice
- Relate and extract
- Details-on-demand
- History available
- *Dinosaurs*



Qualitative Evaluation

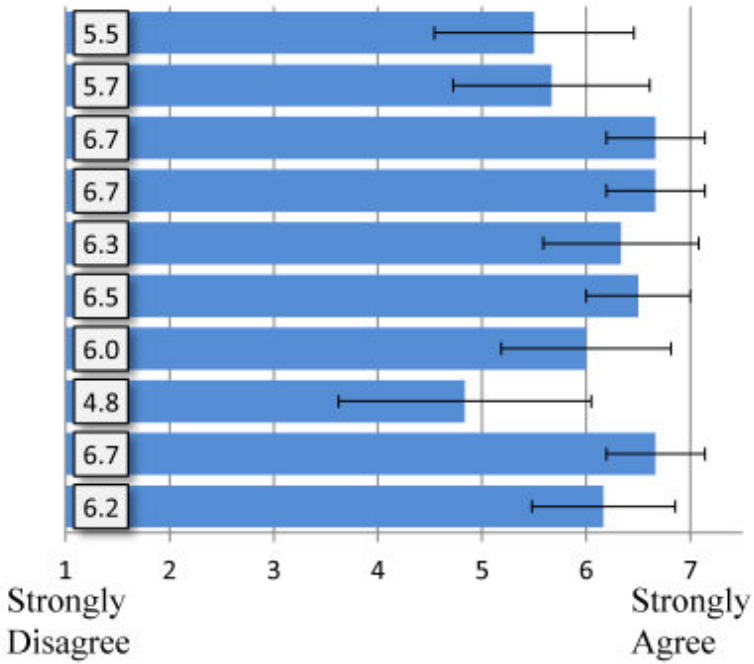
- Participants
  - 6 graduate students (2 females and 4 males)
  - Aged 24-33
  - Had at least 2 years of research experience
  - Had conducted literature searches as part of research

Qualitative Evaluation

- Tasks
  - 20 low-level analytical tasks, classified into 10 groups according to Amar et al. [2]
  - 3 high-level exploratory tasks

[2] R. Amar, J. Eagan, and J. Stasko. Low-level components of analytic activity in information visualization *IEEE Symposium on Information Visualization,* **2005**, 111 - 117

Questionnaire Results



Q1. Easy to learn.
Q2. Easy to use.
Q3. Helpful to organize and browse data.
Q4. Helpful to locate and query specific data.
Q5. Helpful to reveal and obtain data information.
Q6. Helpful to identify and interpret data relationships.
Q7. I feel the animations are useful.
Q8. I feel the graphical histories are useful.
Q9. I feel the integration of online searching is useful.
Q10. I would like to use PivotSlice to explore faceted datasets.

Observation and Feedback

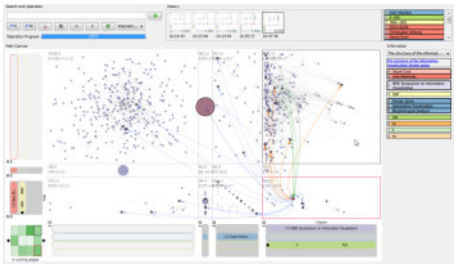
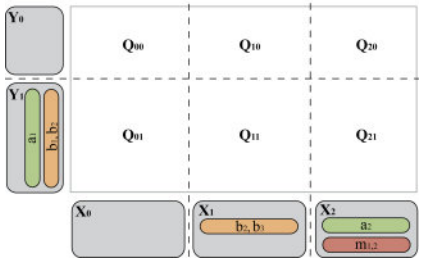
- Interaction observation
  - Low-level tasks
    - All completed relatively quickly (~1.5min each)
    - Various approaches emerged for the same task
  - High-level tasks
    - Used different functions to find insights and explore more deeply in the data

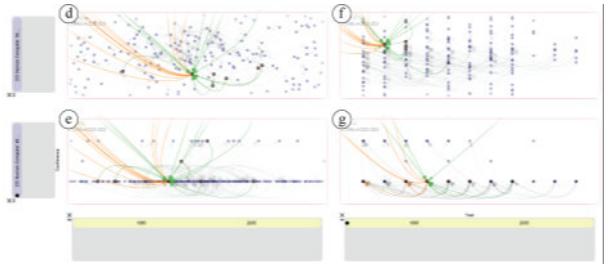
Observation and Feedback

- Sample feedback
  - + "[PivotSlice] makes it easy for organizing complicated filters and searches"
  - + "it is extremely helpful in uncovering trends and distributions between data attributes [...] relating different parts of data is easy by using the connections and different alignments"
  - – "it is sometimes confusing that you can perform the same task in many different ways"
  - – "consider using similar visual cues from the field of logical circuit design [for representing AND/OR operations]"

Future Work

- Apply PivotSlice to other types of faceted data
  - Movie and music databases
  - Large scale datasets
- Conduct further evaluation
  - Laboratory and practical settings (deployment)
  - Quantitative and comparative study
- Improve the visualization and interaction techniques
  - Better node layout (currently force-directed)
  - Semantic zooming
  - Edge bundling



Interactive Exploration of Implicit and Explicit Relations in Faceted Datasets

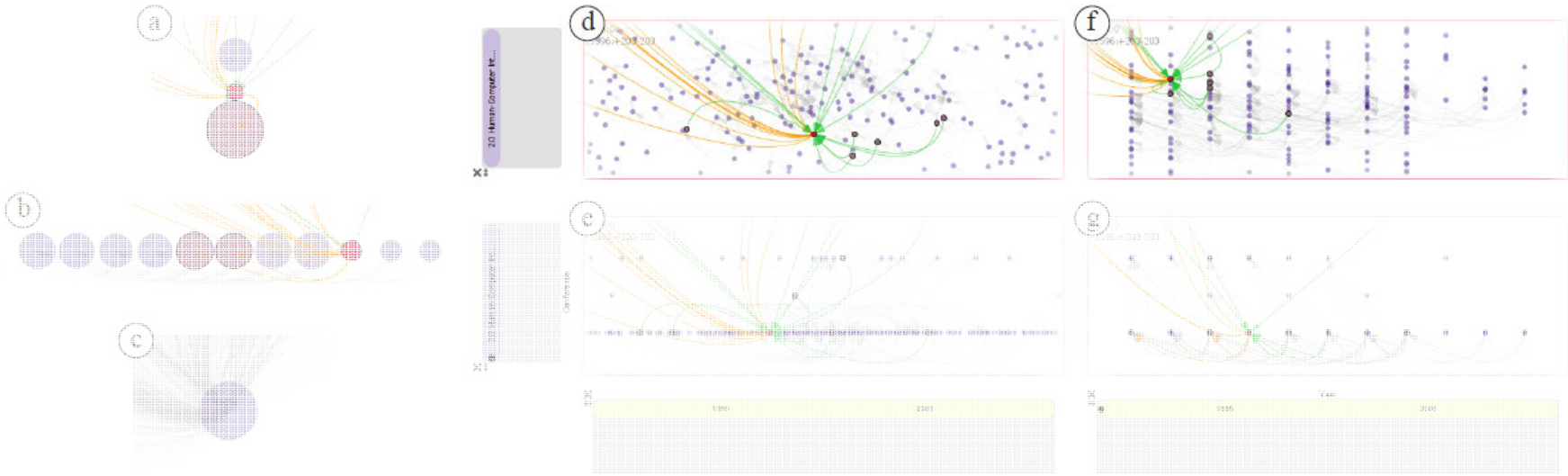
Jian Zhao, Christopher Collins, Fanny Chevalier, and Ravin Balakrishnan

Try it yourself: www.cs.toronto.edu/~jianzhao/ or **tiny.cc/pivotslice**

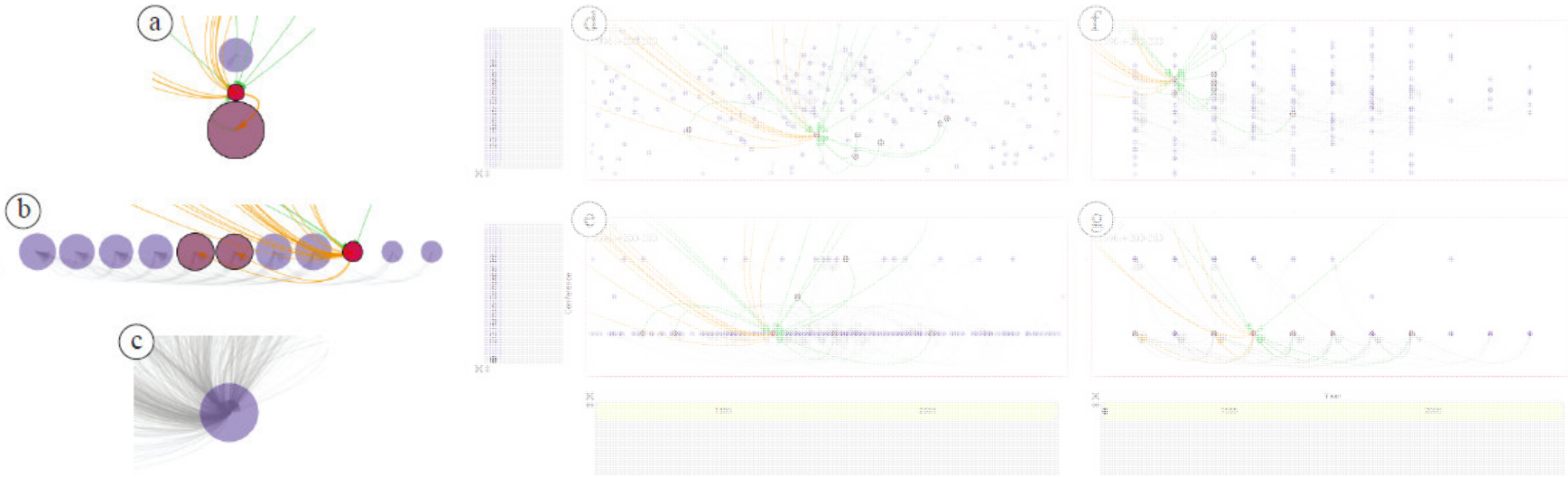
Tasks

Retrieve Value	T1. Who are the authors of Cone trees: Animated 3D visualization of hierarchical information? T2. What are the papers cited by paper H3: Laying out large directed graph in 3D hyperbolic space?
Filter	T3. What are the papers coauthored by John Stasko and Robert Amar? T4. What are the papers with keyword Information Visualization in 2000–2002? How do papers in these years cite each other?
Determine Range	T5. What are the publication years of author Catherine Plaisant? T6. What is the range of citation counts for papers written by Chris North?
Find Extremum	T7. What are the most popular keywords of papers in conferences excluding Computer Human Interaction? T8. Which author publishes the most papers in conference IEEE Symposium on Information Visualization?
Compute Derived Value	T9. Find all the citations of the papers with keyword Information Visualization in conference Computer Human Interaction. T10. Who are the collaborators of author Stuart Card? Are there any patterns in his coauthors?
Sort	T11. Order papers with keywords Information Space or Dynamic Query, by year. Tell us one insight about the citation patterns among papers in different years. T12. Rank papers written by author George Robertson with and without Jock Mackinlay by citation count. What is the most cited paper and how do other papers cite it?
Characterize Distribution	T13. What is the distribution of Stuart Card's papers by conference? How do papers in those conferences refer to each other? T14. Compare the distributions of papers by year between conference Computer Human Interaction and IEEE Symposium on Information Visualization. What can you identify?
Find Anomalies	T15. Are there any exceptions for the conference IEEE Symposium on Information Visualization in terms of paper year trends? T16. Observe papers with keywords Information Seeking or Dynamic Query in this dataset. Is there any incomplete data in terms of citations and references? What are the papers?
Cluster	T17. Identify groups of papers in conference IEEE Symposium on Information Visualization, in terms of similar year and citation count attribute values. T18. Find clusters of papers written by Stuart Card, George Robertson, or Jock Mackinlay. Can you tell more about the collaboration patterns between them?
Correlate	T19. Compare papers containing both keywords Information Visualization and User Interface, and the ones containing neither of them. Who are the authors that publish the most in the two categories? What are the popular keywords in addition to the two above? T20. Compare papers published in years 1990–2000 and 2001–2010. Of keywords Three Dimensional, Information Space and World Wide Web, which one is more popular in the year ranges? How many papers exactly? Can you tell more about the citations across those paper categories?
High Level Tasks	T21. Tell three facts about the author Ben Shneiderman. T22. Give three comments on the conference User Interface Software and Technology. T23. Name three interesting findings for the Direct Manipulation research keyword.

Layout - Force Directed



Layout - Aggregated



Layout - Facet-Aligned