



# A Survey on Visual Approaches for Analyzing Scientific Literature and Patents

Paolo Federico, Florian Heimerl,  
Steffen Koch, Silvia Miksch

# Scientific Literature and Patents

- Similar set of information
  - Textual information (+ formulas, tables, images, photo, video)
  - Metadata
  - Relational information (authorship, citations, classification)
  - Temporal information
- Similar analytical tasks



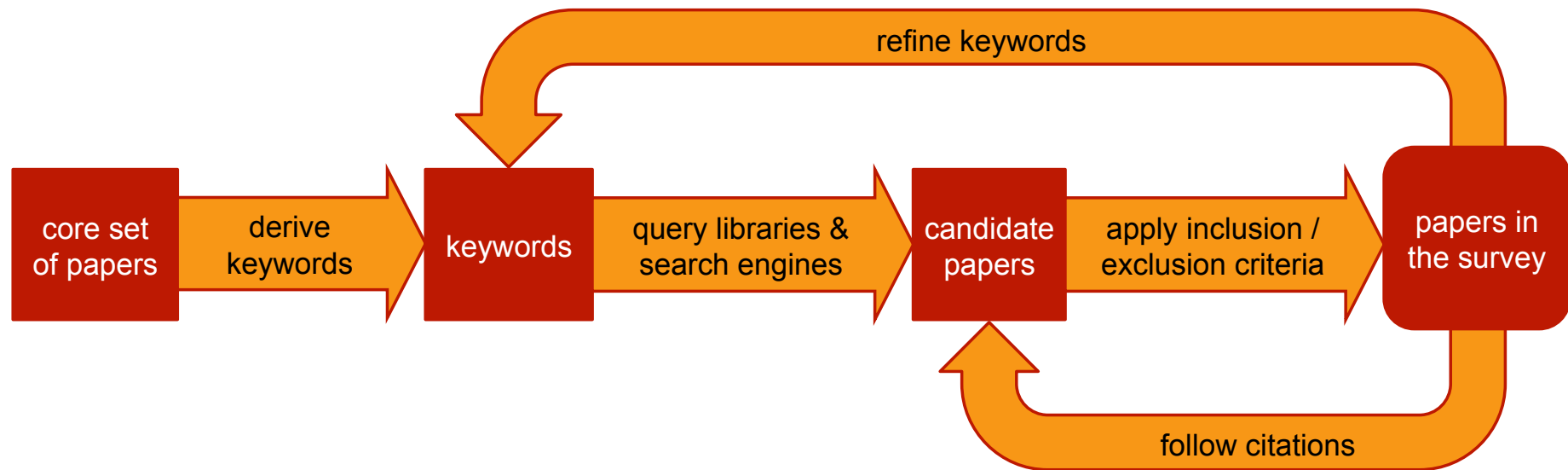
# Related Surveys

- Patents
  - Processing and retrieval (e.g., Bonino et al., 2010)
  - Commercial solutions (Yang et al., 2008)
- Scientific Literature
  - Evolution of knowledge (Chen, 2013)
  - Structure of research communities (Boerner, 2010)
- Text Analysis
  - Search and retrieval (Hearst, 2009)
  - Visualization (e.g., Kucher & Kerren, 2015)



source: scimaps.org

# Methodology



- Examples of exclusion criteria:
  - lack of interactivity, commercial products, patents

# Categorization

- Inspired by the task-by-datatype taxonomy (Shneiderman, 1996)
  - Domain-specific data types (bottom-up)
  - Task taxonomy by Andrienko & Andrienko (2006)



# Categorization

Data

Text

Citations

Authors

Metadata



# Categorization

Data	Tasks
Text	Entities
Citations	Relations
Authors	Patterns
Metadata	Temporal



# Categorization

<b>Data</b> Text Citations Authors Metadata	<b>Tasks</b> Entities Relations Patterns Temporal
Multiple	Connectional
<b>Level of integration</b> Aggregation, Labelling, Composition, Multiple Views, Tight integration	



# Survey Results: PaperViz.org

**PaperViz.org - A Survey on Visual Approaches for Analyzing Scientific Literature and Patents** (get pdf) SurVis

---

**Timeline**

publications per year

**Keywords** filter ... min 1

data: sci-lit<sub>109</sub> patents<sub>21</sub>

datatype: citations<sub>63</sub> meta<sub>51</sub> text<sub>50</sub> authors<sub>32</sub>

task: patterns<sub>74</sub> temporal<sub>41</sub> relations<sub>23</sub> entities<sub>21</sub>

multiple: none<sub>81</sub> views<sub>17</sub> aggregation<sub>15</sub> labelling<sub>7</sub> integration<sub>7</sub> composition<sub>4</sub>

other:

**Authors** filter ... min 3

Chen, Chaomei<sub>9</sub> Ertl, Thomas<sub>5</sub> Heer, Jeffrey<sub>4</sub> Henry Riche, Nathalie<sub>4</sub> Morris, Steven A.<sub>4</sub>  
 Boerner, Katy<sub>4</sub> White, Howard D.<sub>4</sub> Stasko, John<sub>3</sub> Fekete, Jean-Daniel<sub>3</sub> Park, Haesun<sub>3</sub> Chuang, Jason<sub>3</sub>  
 Manning, Christopher D.<sub>3</sub> Czerwinski, Mary<sub>3</sub> Giereth, Mark<sub>3</sub> Bosch, Harald<sub>3</sub> Shneiderman, Ben<sub>3</sub>  
 Choo, Jaegul<sub>3</sub> Havre, Susan<sub>3</sub> Hetzler, Elizabeth<sub>3</sub> Koch, Steffen<sub>3</sub> Lee, Bongshin<sub>3</sub> Lin, Xia<sub>3</sub>  
 Dou, Wenwen<sub>3</sub> Rosvali, Martin<sub>3</sub> Bergstrom, Carl T.<sub>3</sub>

**Series** filter ... min 1

TVCG<sub>18</sub> InfoVis<sub>11</sub> IV<sub>7</sub> VAST<sub>6</sub> JASIST<sub>5</sub> LNCS<sub>5</sub> PNAS<sub>5</sub> CGF<sub>5</sub> IVI<sub>4</sub> CHI<sub>4</sub> IPM<sub>3</sub> AVI<sub>3</sub>  
 CHI EA<sub>3</sub> Informetrics<sub>2</sub> APVis/PacificVis<sub>2</sub> DL<sub>2</sub> IVAPP<sub>2</sub> SIGIR<sub>2</sub> VIS Workshops<sub>2</sub> PLoS ONE<sub>2</sub> JASIS<sub>2</sub>  
 ICML<sub>1</sub> Scientometrics<sub>1</sub> CIKM<sub>1</sub> PICMET<sub>1</sub> JIS<sub>1</sub> GI<sub>1</sub> VISSYM<sub>1</sub> TIST<sub>1</sub> VIS Posters<sub>1</sub> UHC<sub>1</sub> ASOM<sub>1</sub> CGV<sub>1</sub>  
 JoV<sub>1</sub> IVICA<sub>1</sub> NN<sub>1</sub> AVI Workshops<sub>1</sub> WHAT<sub>1</sub> Computer<sub>1</sub> JNR<sub>1</sub> UDL<sub>1</sub> CASCON<sub>1</sub> JCIE<sub>1</sub> ISWC<sub>1</sub> ICDIM<sub>1</sub>  
 ISVC<sub>1</sub> ICSC<sub>1</sub> TKDE<sub>1</sub> Technovation<sub>1</sub> CG&A<sub>1</sub> JIS<sub>1</sub> ICDM<sub>1</sub> WIS<sub>1</sub> UHC<sub>1</sub> ICEI<sub>1</sub>

**Clusters**

**129 publications** sorted by selector agreement and publication key

---

1. Abello2014 TVCG (2014) | DOI | Google Scholar | Google  
**A Modular Degree-of-Interest Specification for the Visual Analysis of Large Dynamic Networks**  
 Abello, J. Hadlak, S. Schumann, H. Schulz, H. J.  
*Abstract: Large dynamic networks are targets of analysis in many fields. Tracking temporal changes at scale in these networks is challenging due in part to the fact that small changes can be missed or drowned out by the rest of the network. For static networks, current approaches allow the identification o...*  
 data:sci-lit datatype:citations task:temporal multiple:none  
 select similar BibTeX

---

2. Ahmed2004 InfoVis (2004) | DOI | Google Scholar | Google  
**InfoVis 2004 Contest: Wilmascope graph visualisation**  
 Ahmed, Adel Dwyer, Tim Murray, Colin Song, Le Wu, Ying Xin  
*Abstract: Our visualisation of the IEEE InfoVis citation network is based on 3D graph visualisation techniques. To make effective use of the third dimension we use a layered approach, constraining nodes to lie on parallel planes depending on parameters such as year of publication or link degree. Within the...*  
 data:sci-lit datatype:authors datatype:citations datatype:text task:temporal multiple:none  
 select similar BibTeX

---

3. Alper2011 TVCG (2011) | DOI | Google Scholar | Google  
**Design Study of LineSets, a Novel Set Visualization Technique**  
 Alper, B. Henry Riche, Nathalie Ramos, Gonzalo Czerwinski, Mary  
*Abstract: Computing and visualizing sets of elements and their relationships is one of the most common tasks one performs when analyzing and organizing large amounts of data. Common representations of sets such as convex or concave geometries can become cluttered and difficult to parse when these sets over...*  
 data:sci-lit datatype:authors datatype:meta task:patterns multiple:aggregation  
 select similar BibTeX

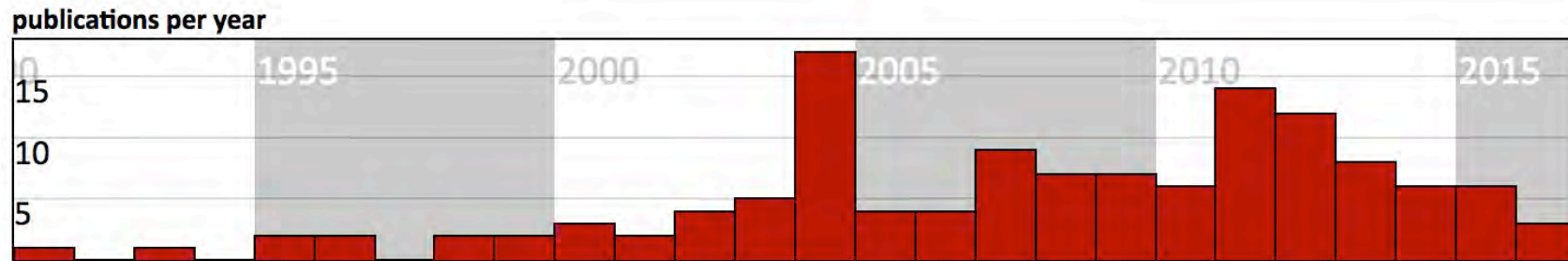
---

4. Alsallakh2013 TVCG (2013) | DOI | Google Scholar | Google  
**Radial Sets: Interactive Visual Analysis of Large Overlapping Sets**  
 Alsallakh, Bilal Aigner, Wolfgang Miksch, Silvia Hauser, Helwig  
*Abstract: In many applications, data tables contain multi-valued attributes that often store the memberships of the table entities to multiple sets such as which languages a person masters, which skills an applicant documents, or which features a product comes with. With a growing number of entities, the t...*  
 data:sci-lit datatype:meta task:relations multiple:none  
 select similar BibTeX

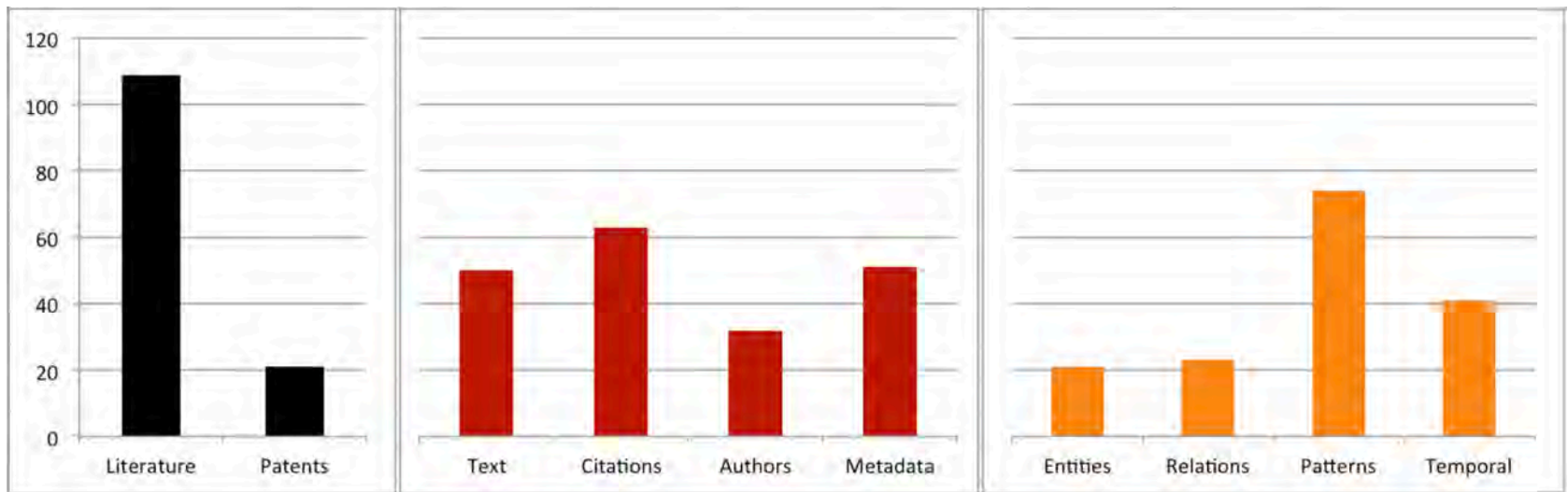


# Collection Overview

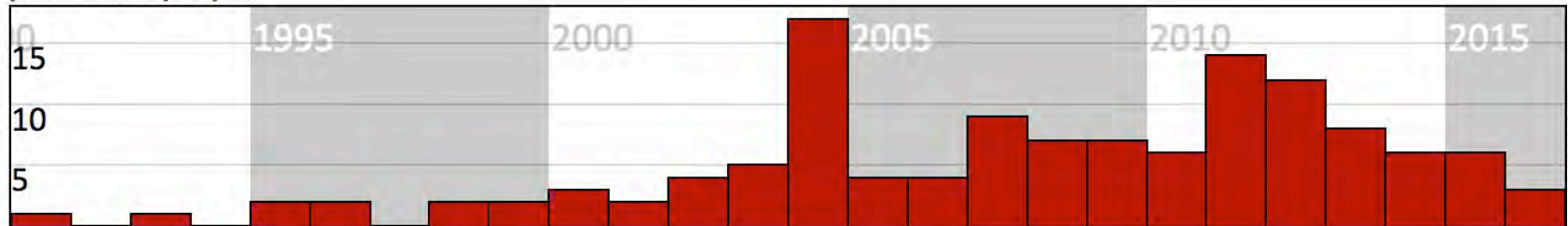
- Scientific literature 109
- Patents 21
- Overall 129 papers



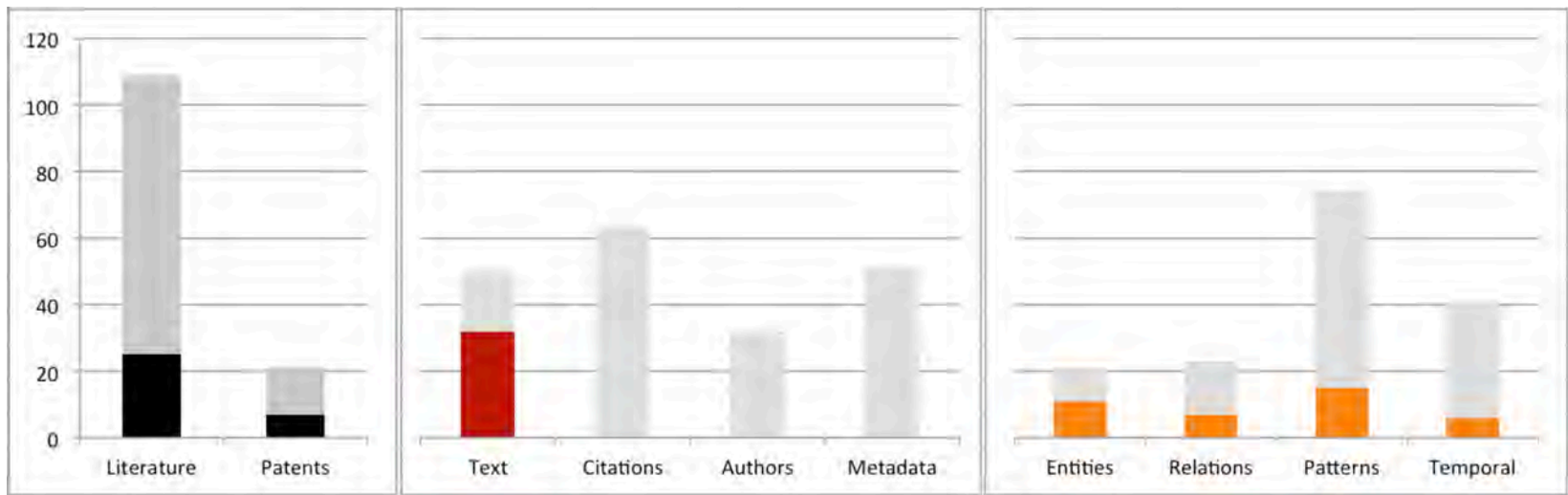
# Categories Overview



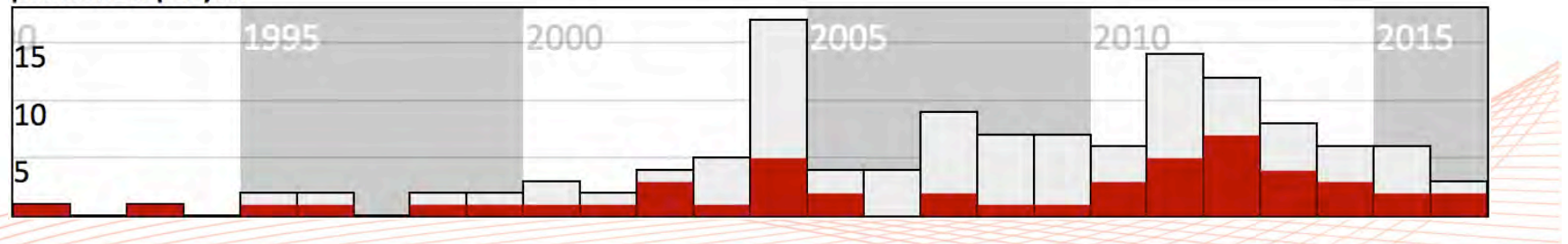
publications per year



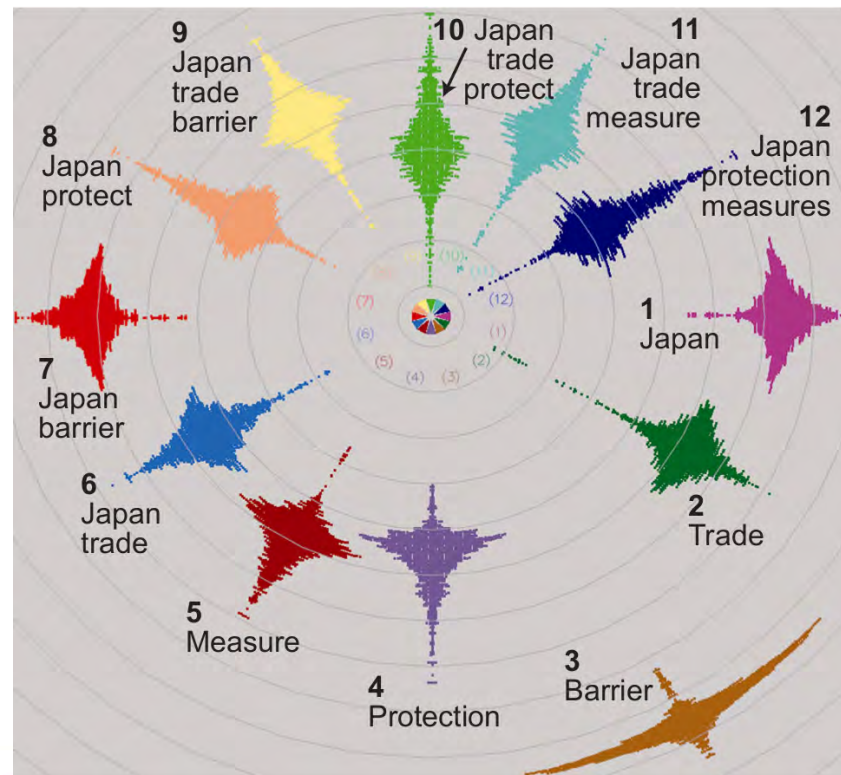
Text | Citations | Authors | Metadata | Multiple



publications per year

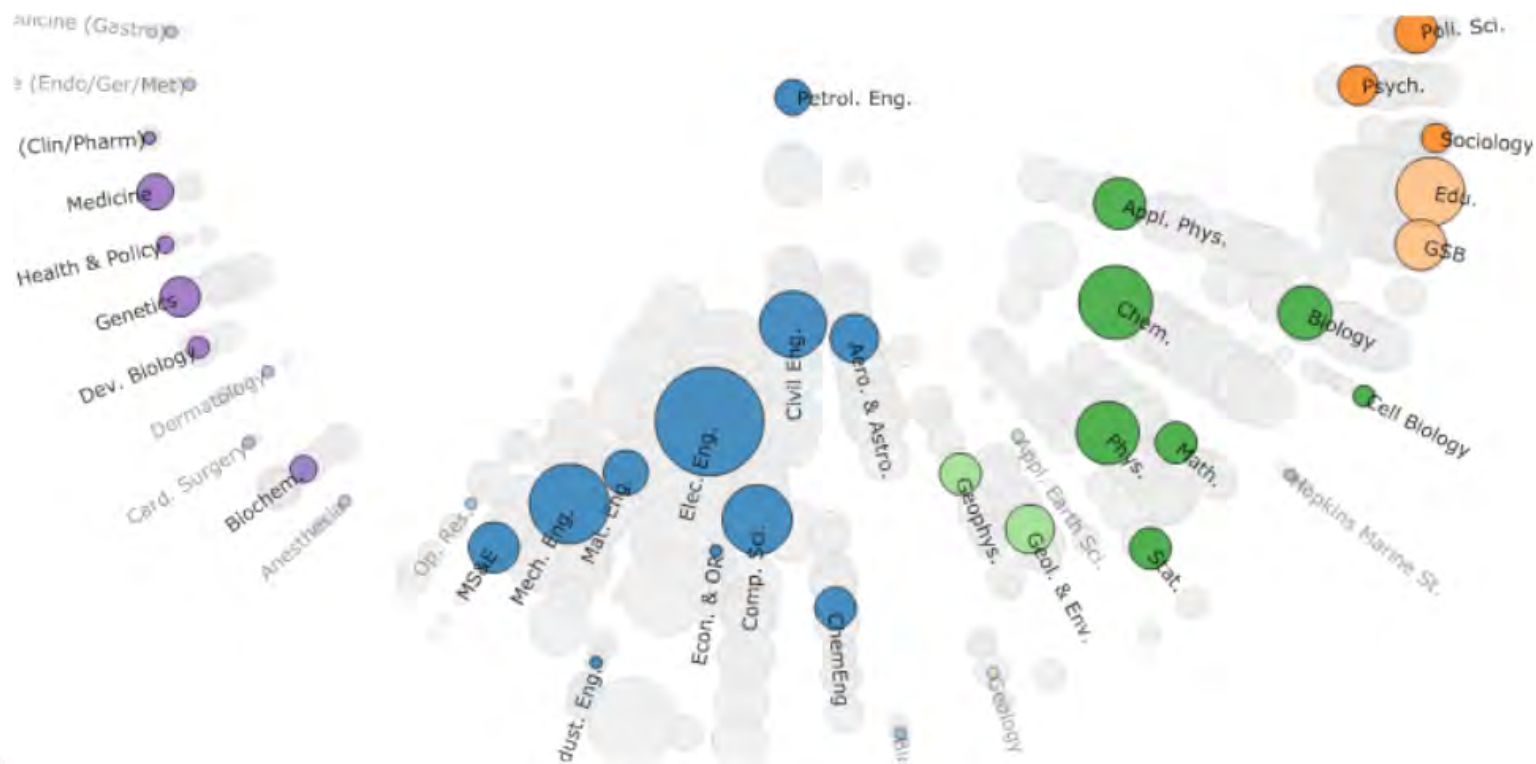


Text | Citations | Authors | Metadata | Multiple  
Entities | Relations | Patterns | Temporal | Connectional



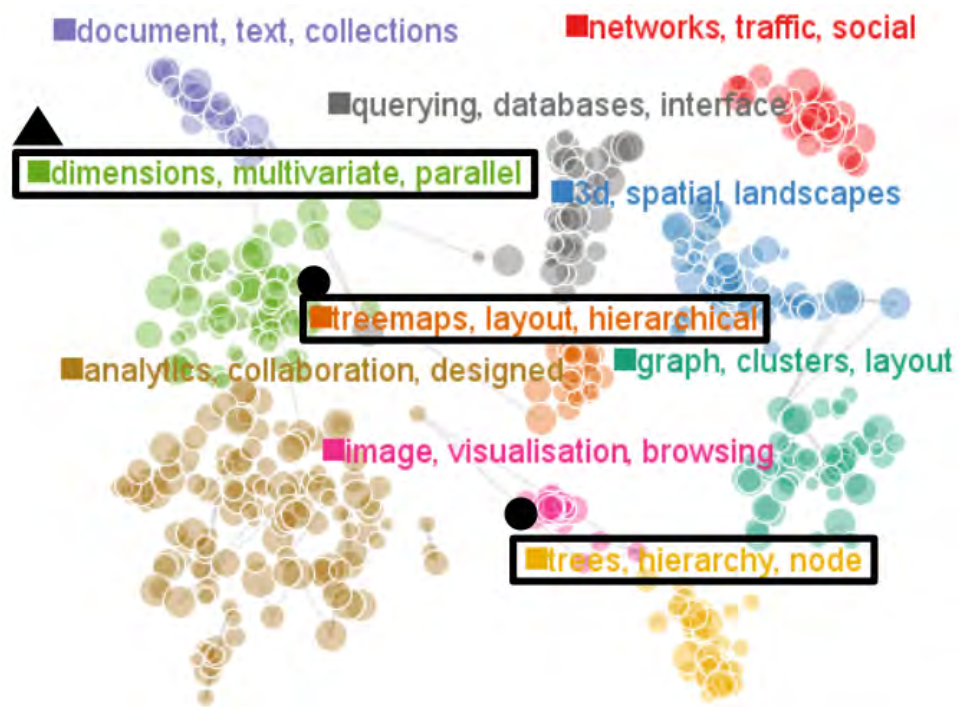
[Havre et al., 2002]

Text | Citations | Authors | Metadata | Multiple  
Entities | Relations | Patterns | Temporal | Connectional



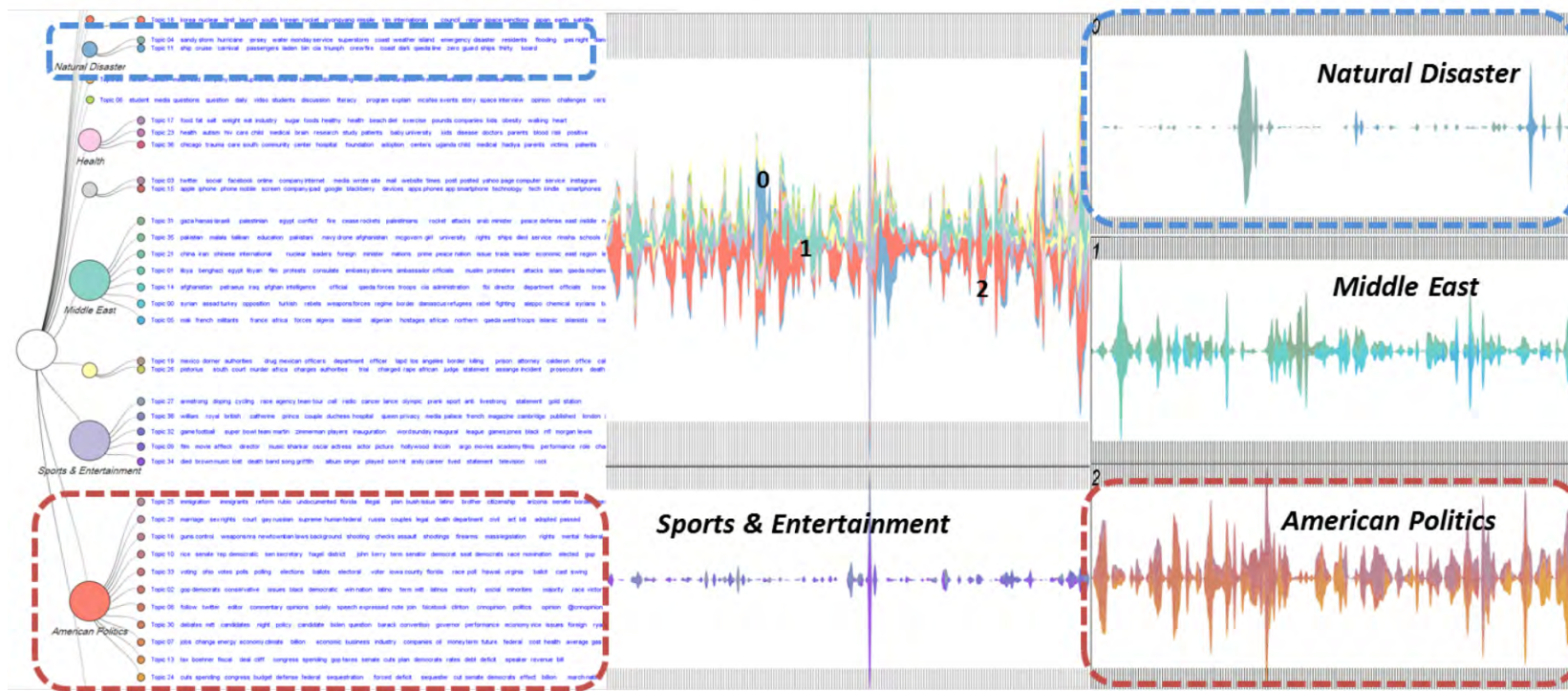
[Chuang et al., 2012]

Text | Citations | Authors | Metadata | Multiple  
Entities | Relations | **Patterns** | Temporal | Connectional



[Choo et al., 2013]

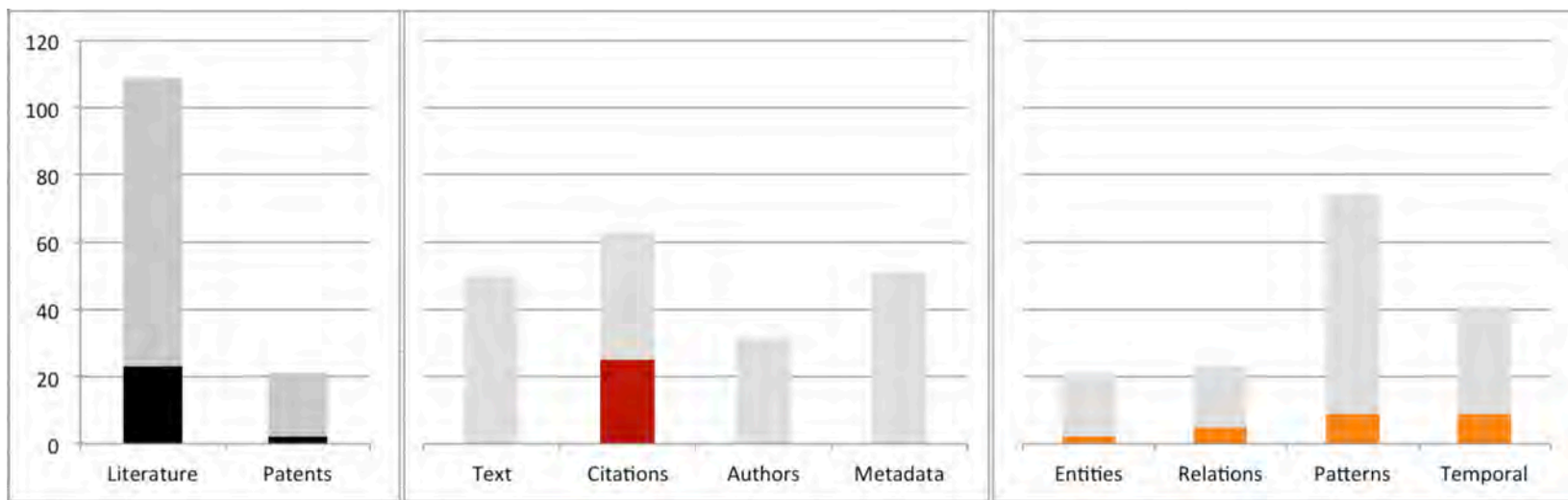
Text | Citations | Authors | Metadata | Multiple  
 Entities | Relations | Patterns | Temporal | Connectional



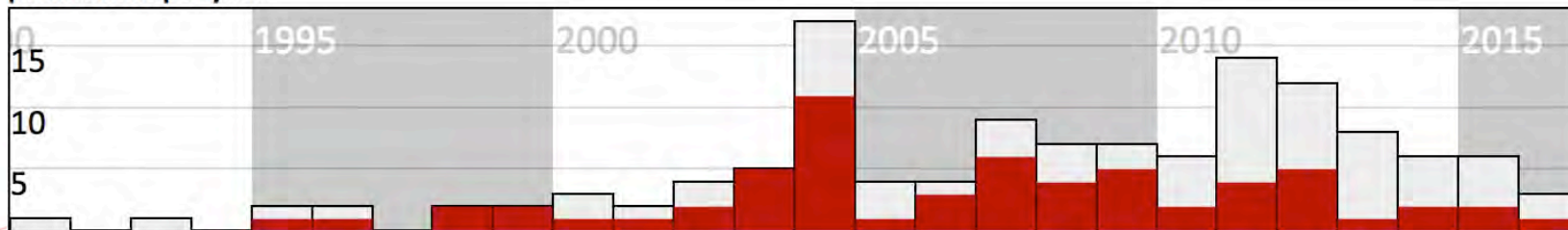
[Dou et al., 2013]



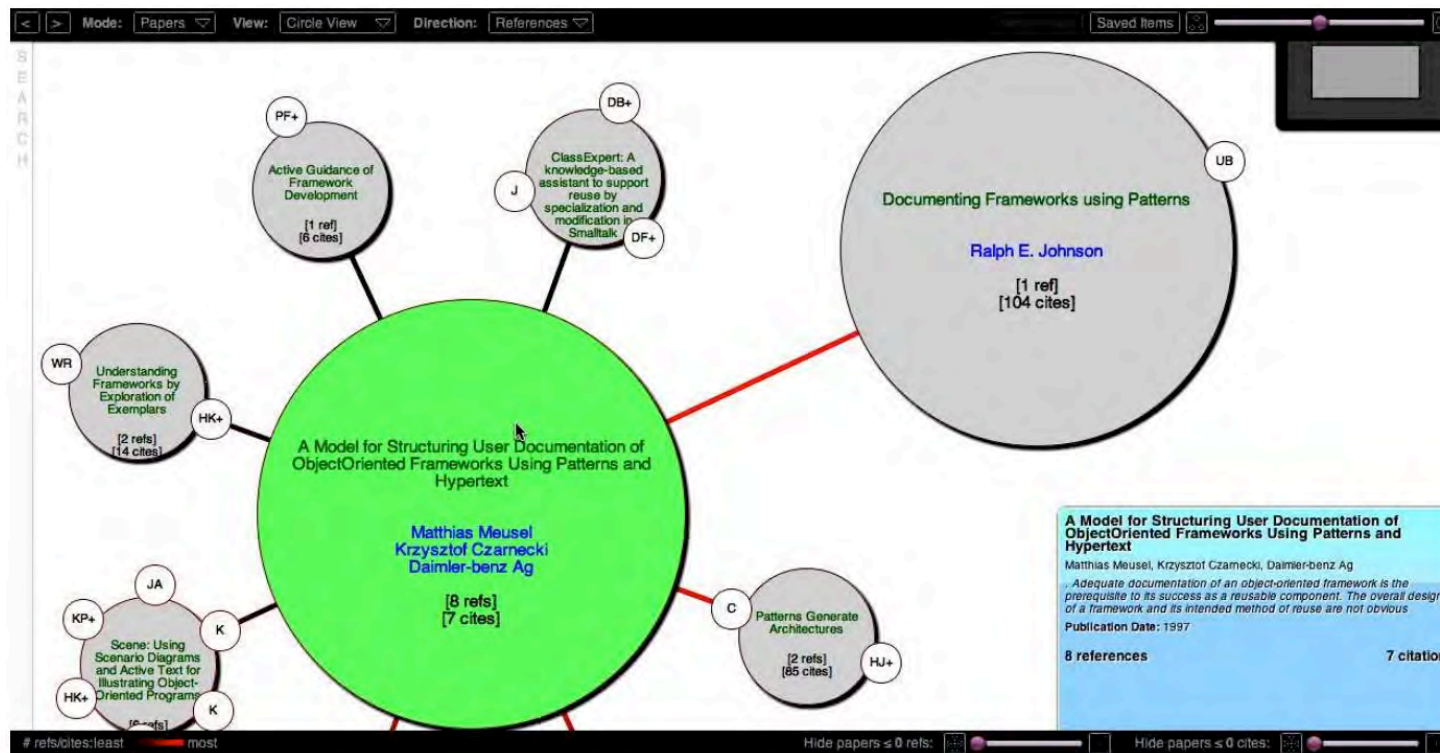
Text | Citations | Authors | Metadata | Multiple



publications per year

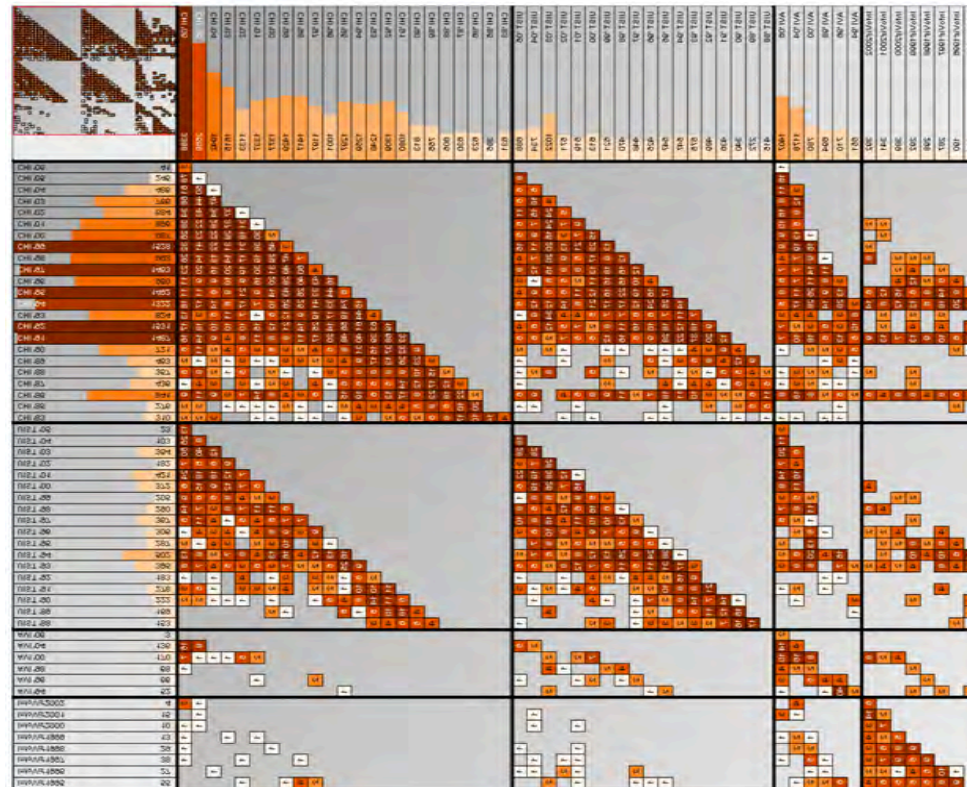


Text | Citations | Authors | Metadata | Multiple  
 Entities | Relations | Patterns | Temporal | Connectional



[Bergstrom, 2009]

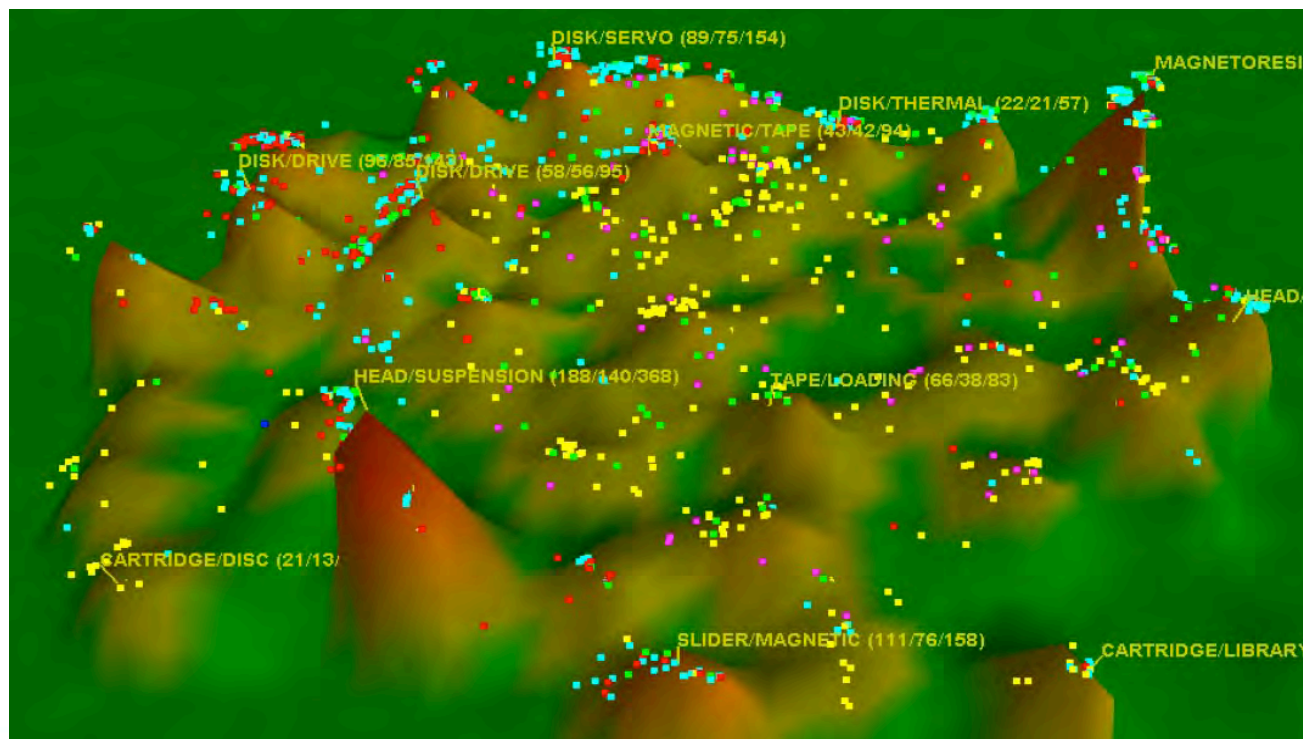
Text | Citations | Authors | Metadata | Multiple  
 Entities | Relations | Patterns | Temporal | Connectional



[Henry, 2007]

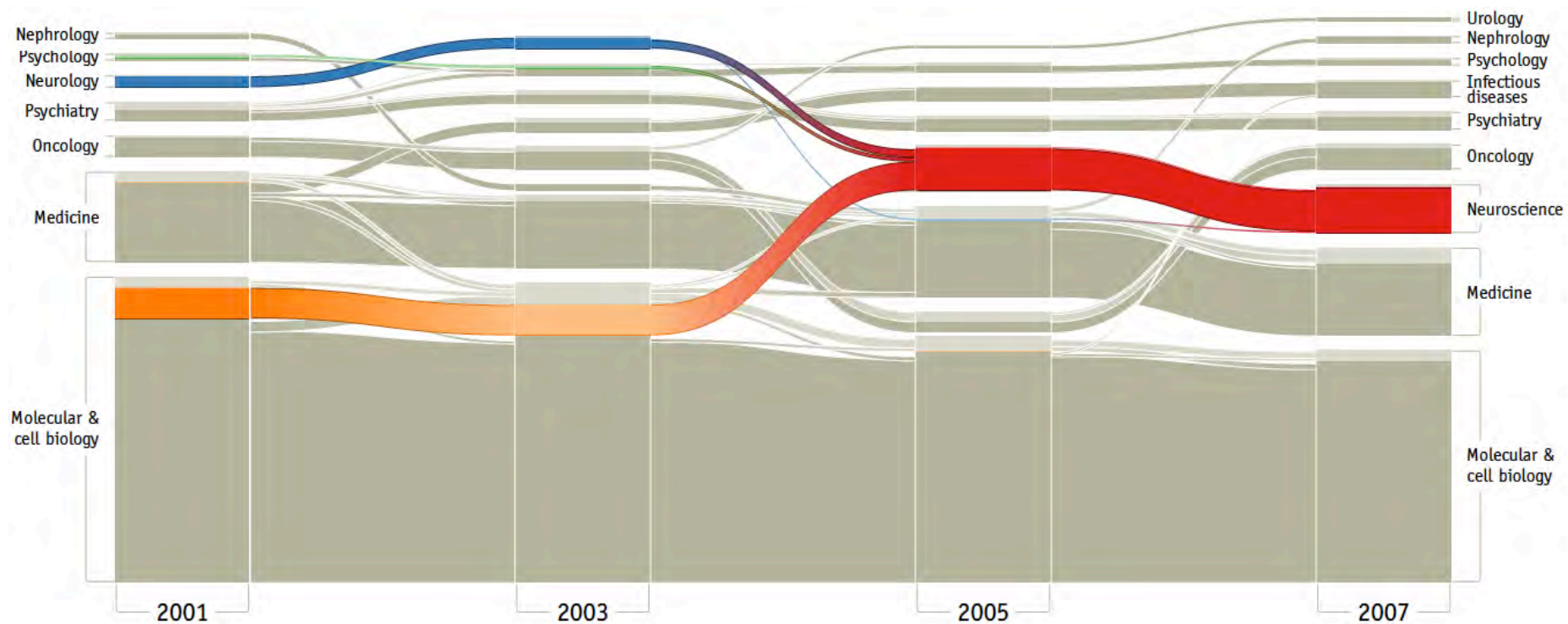


Text | Citations | Authors | Metadata | Multiple  
Entities | Relations | Patterns | Temporal | Connectional

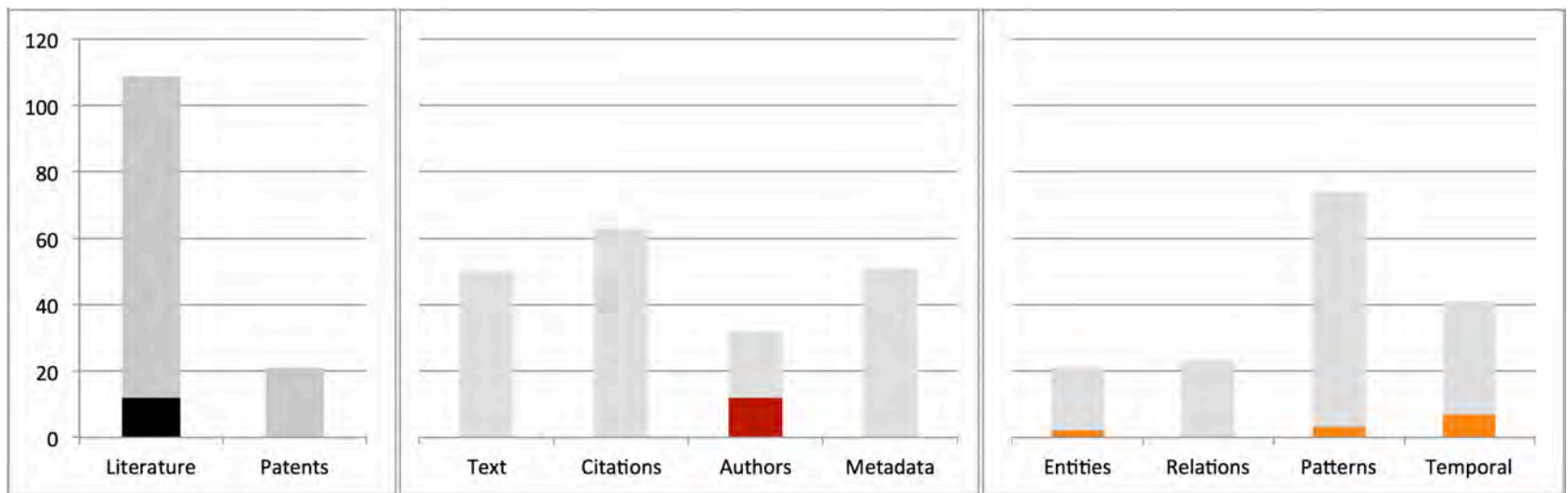


[Boyack, 2000]

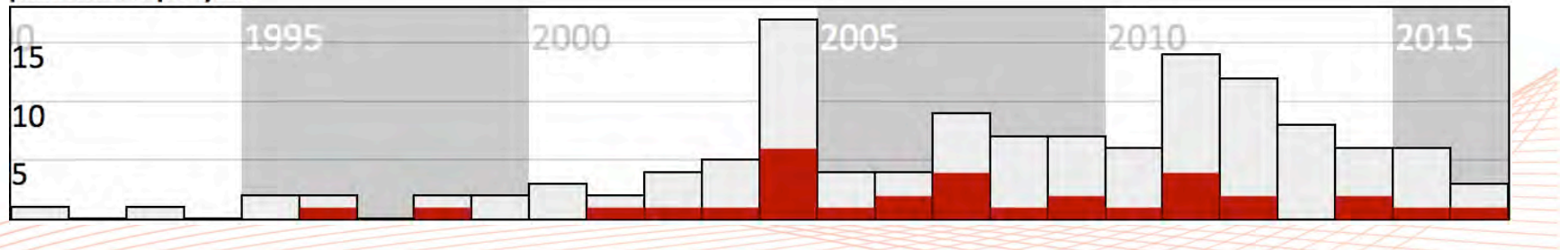
Text | Citations | Authors | Metadata | Multiple  
Entities | Relations | Patterns | Temporal | Connectional



Text | Citations | **Authors** | Metadata | Multiple



publications per year



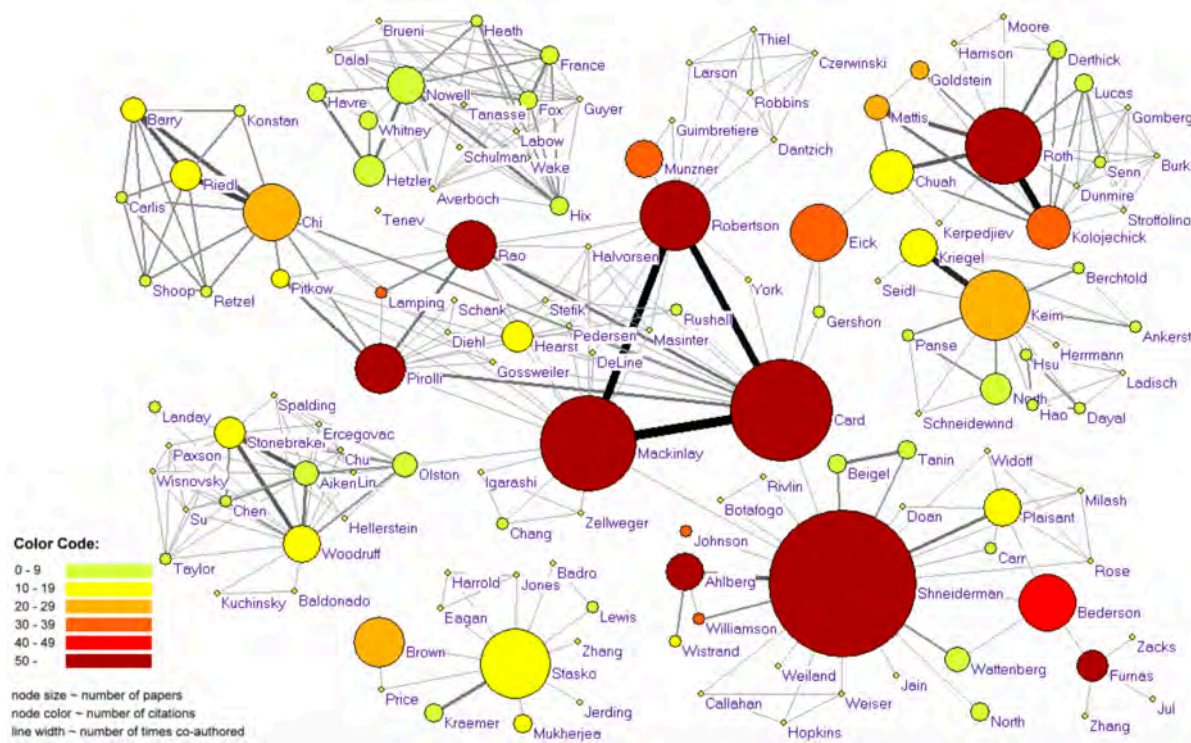
Text | Citations | **Authors** | Metadata | Multiple  
 Entities | Relations | Patterns | Temporal | Connectional



[Kang, 2007]



Text | Citations | **Authors** | Metadata | Multiple  
 Entities | **Relations** | Patterns | Temporal | Connectional

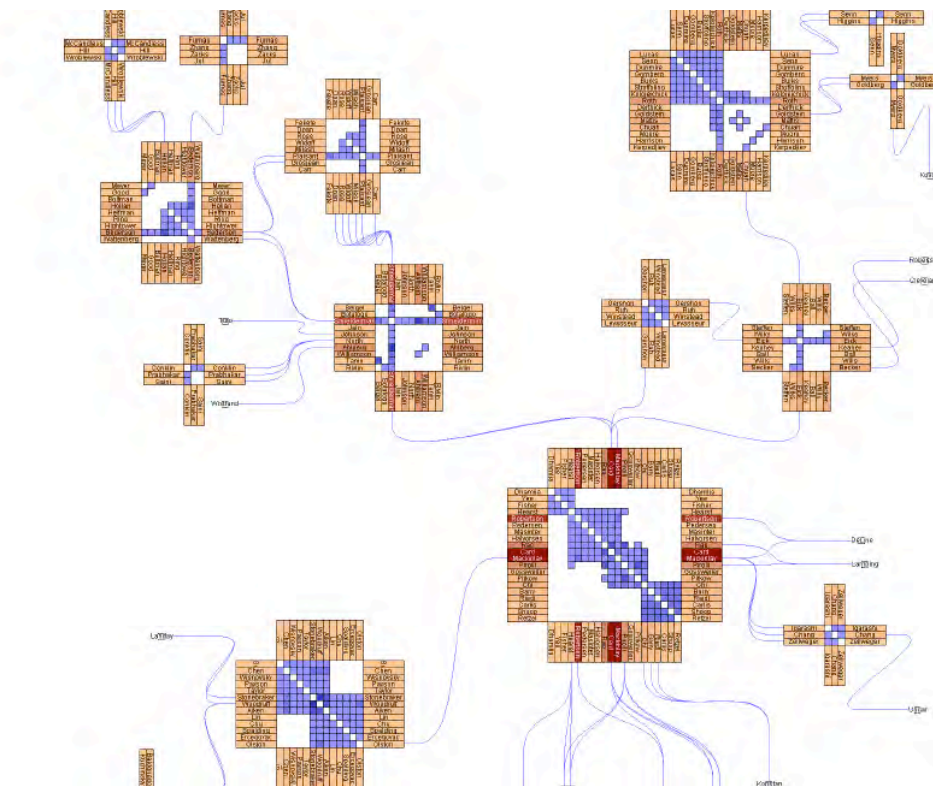


[Ke, 2004]



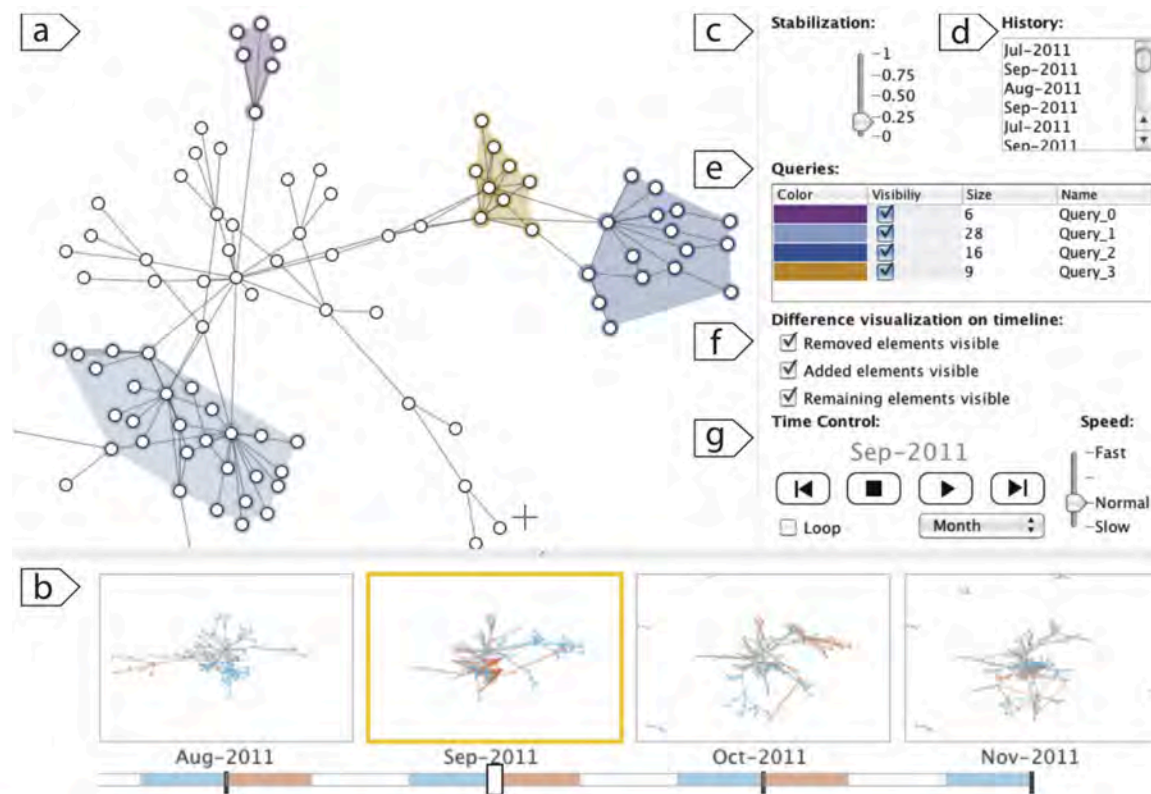


Text | Citations | **Authors** | Metadata | Multiple  
Entities | Relations | **Patterns** | Temporal | Connectional



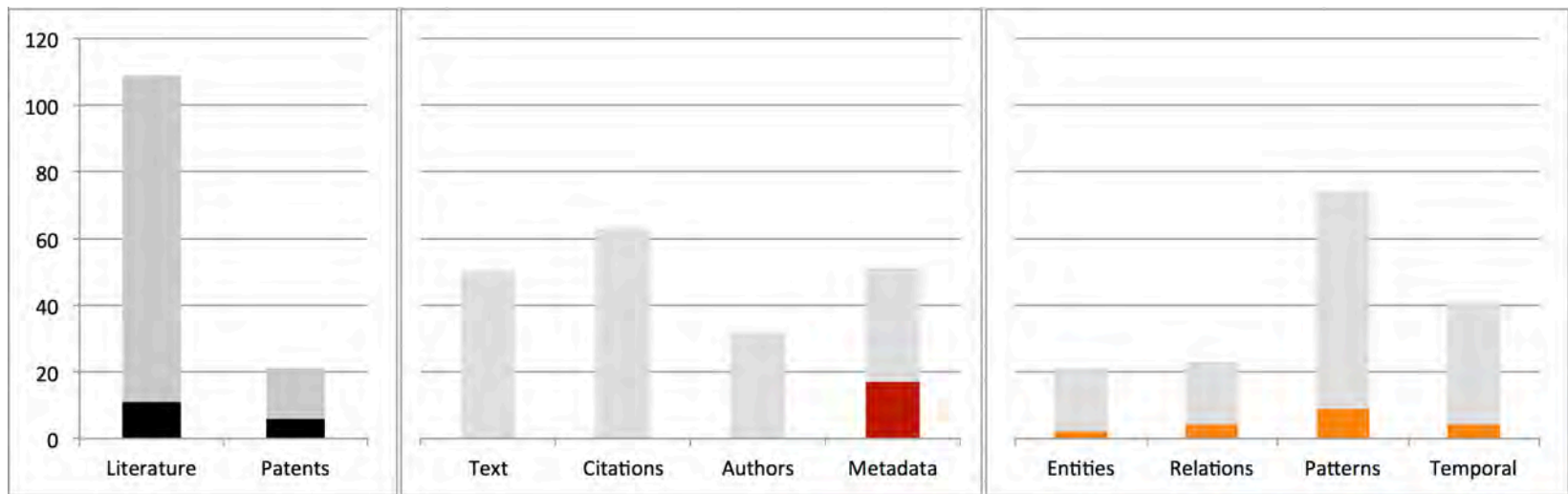
[Henry, 2007]

Text | Citations | **Authors** | Metadata | Multiple  
Entities | Relations | Patterns | **Temporal** | Connectional

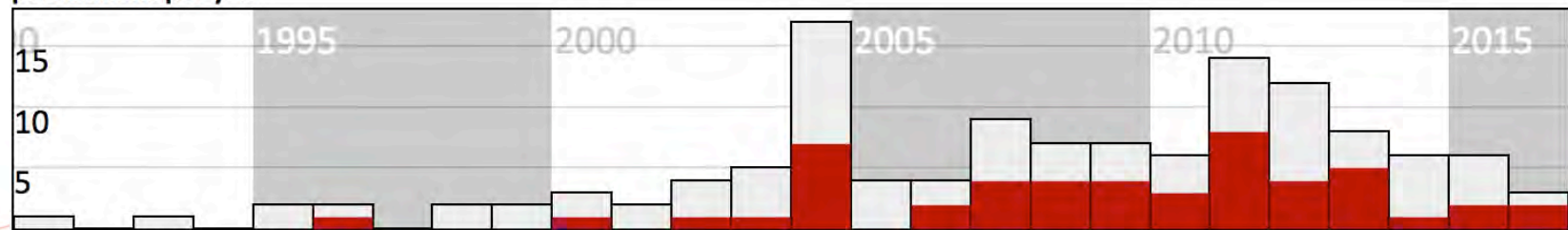


[Bach, 2014]

Text | Citations | Authors | **Metadata** | Multiple



publications per year



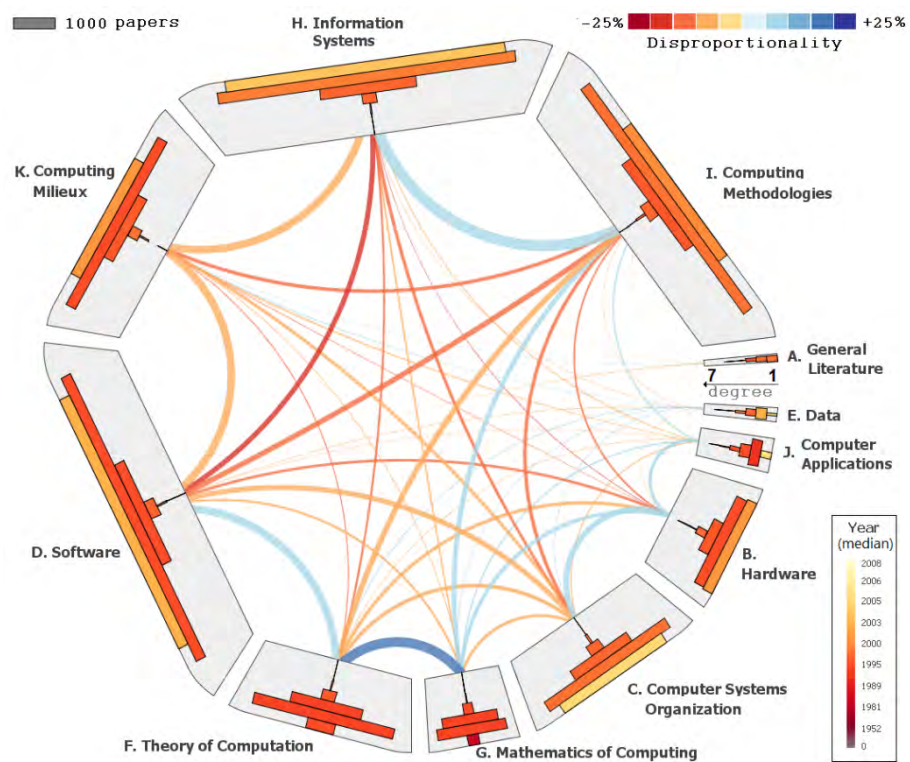
Text | Citations | Authors | **Metadata** | Multiple  
**Entities** | Relations | Patterns | Temporal | Connectional

The screenshot shows the SurVis interface with a search bar at the top. On the left, there is a 'Literature overview' section featuring a horizontal bar chart of citations per publication and a list of keywords with their counts. The main area displays a 'List of publications' with two entries, each including a title, authors, abstract, and metadata tags.

This screenshot shows the 'Selectors' panel in SurVis. It includes a 'publications per year' bar chart with a timeline from 1995 to 2014. Below the chart is a 'Keywords' section with a grid of checkboxes for various categories like 'technique', 'application', 'evaluation', 'time', 'paradigm', 'evaluation', and 'application', each with a corresponding count.

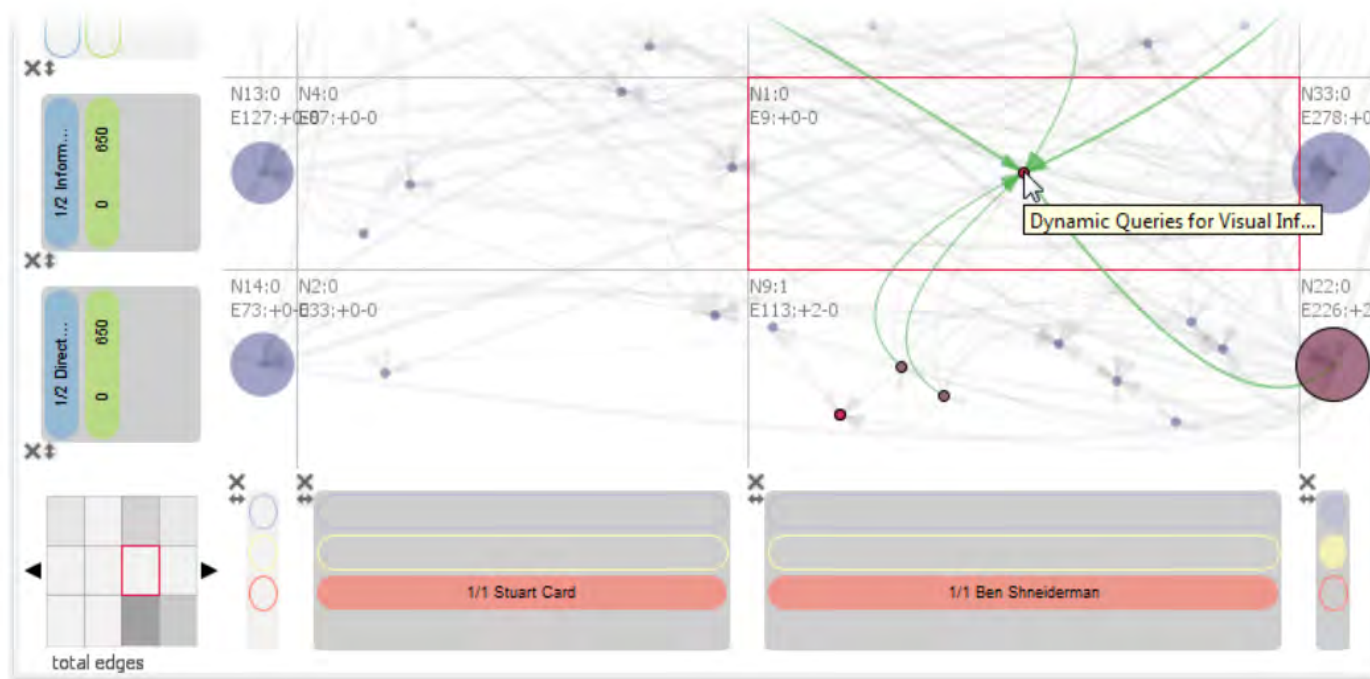
[Beck, 2015]

Text | Citations | Authors | **Metadata** | Multiple  
 Entities | **Relations** | Patterns | Temporal | Connectional

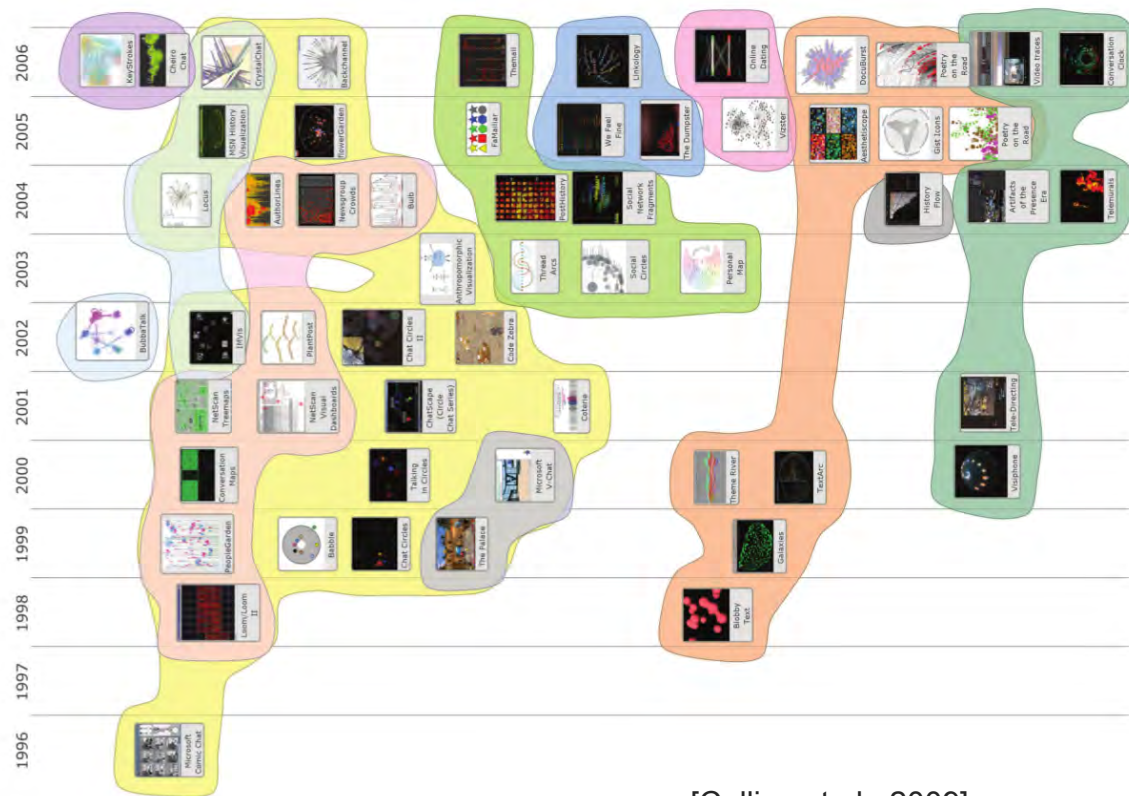


[Alsallakh et al., 2013]

Text | Citations | Authors | **Metadata** | Multiple  
Entities | Relations | **Patterns** | Temporal | Connectional

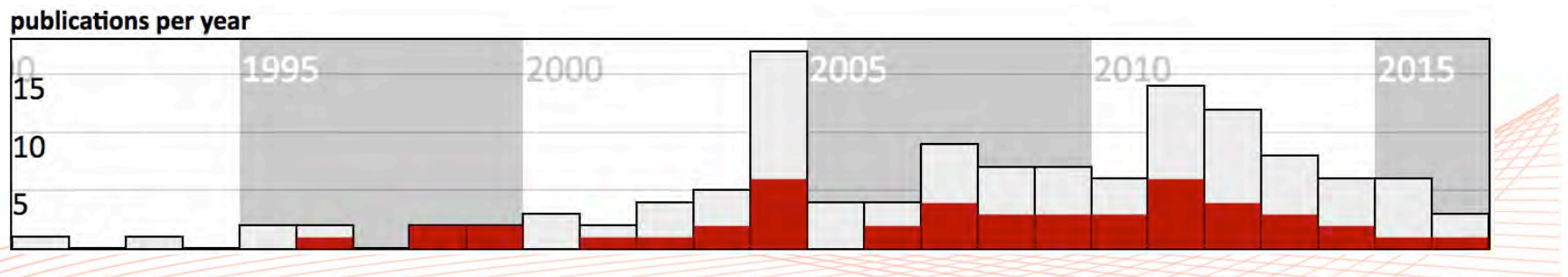
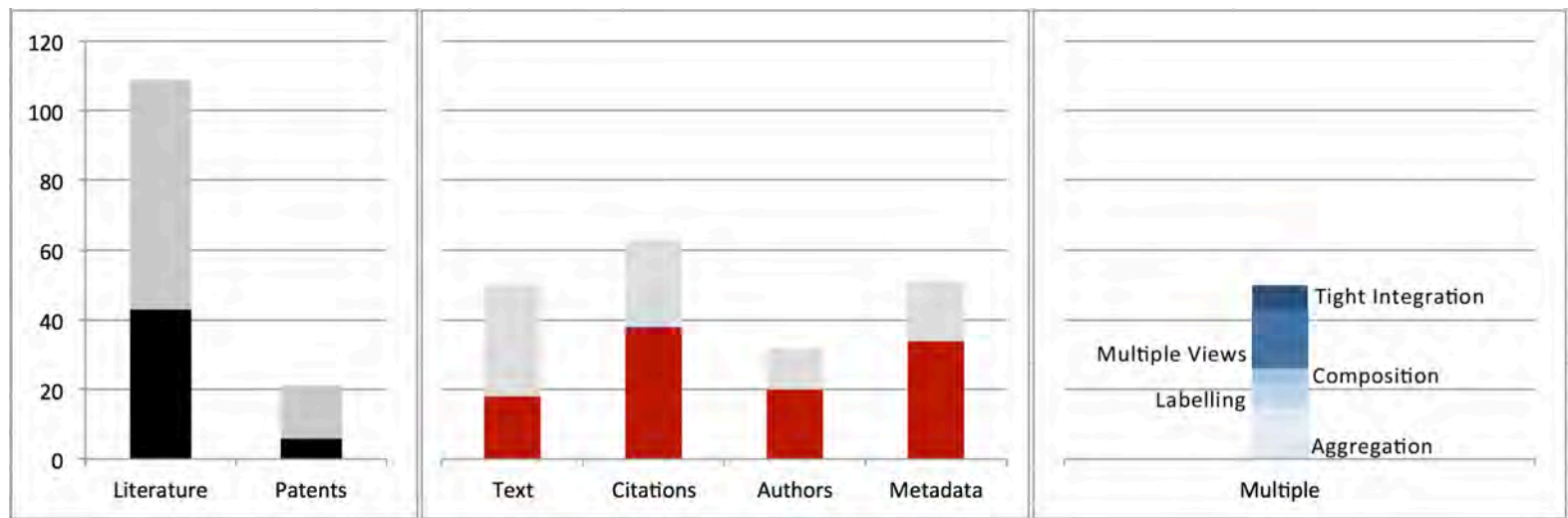


Text | Citations | Authors | **Metadata** | Multiple  
 Entities | Relations | Patterns | **Temporal** | Connectional



[Collins et al., 2009]

# Multiple Connectional

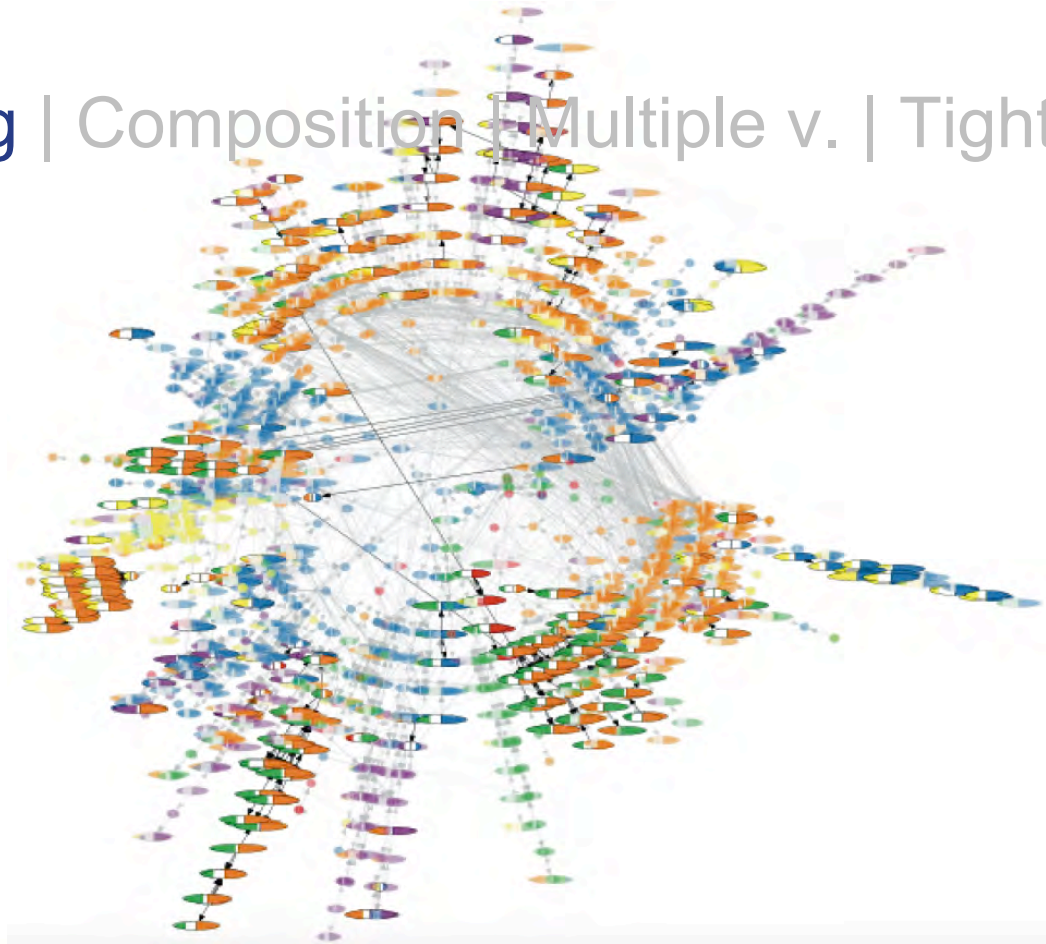






# Multiple Connectional

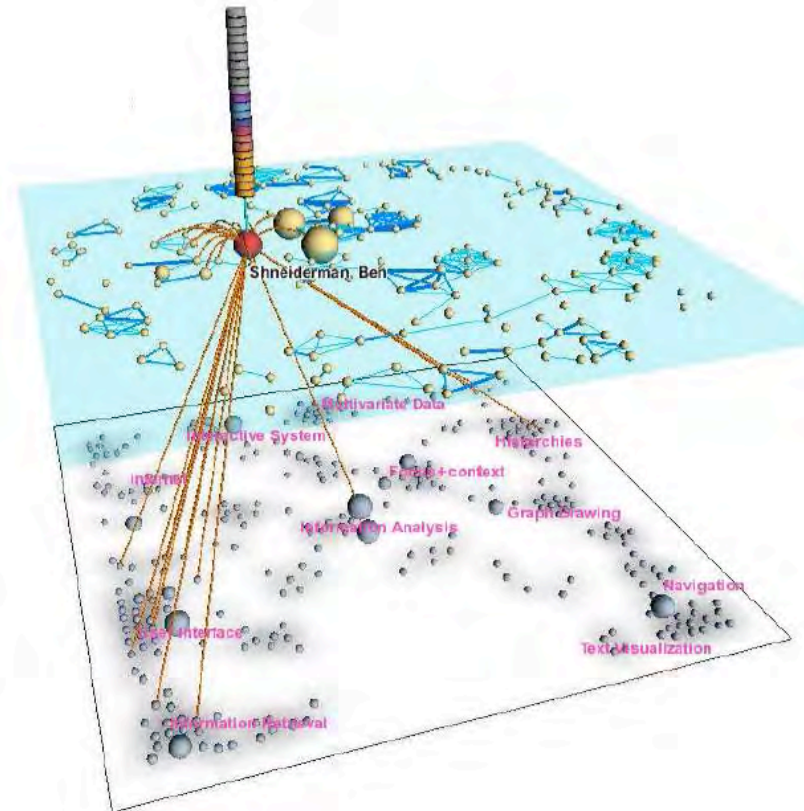
Aggregation | Labelling | Composition [Multiple v. | Tight int.



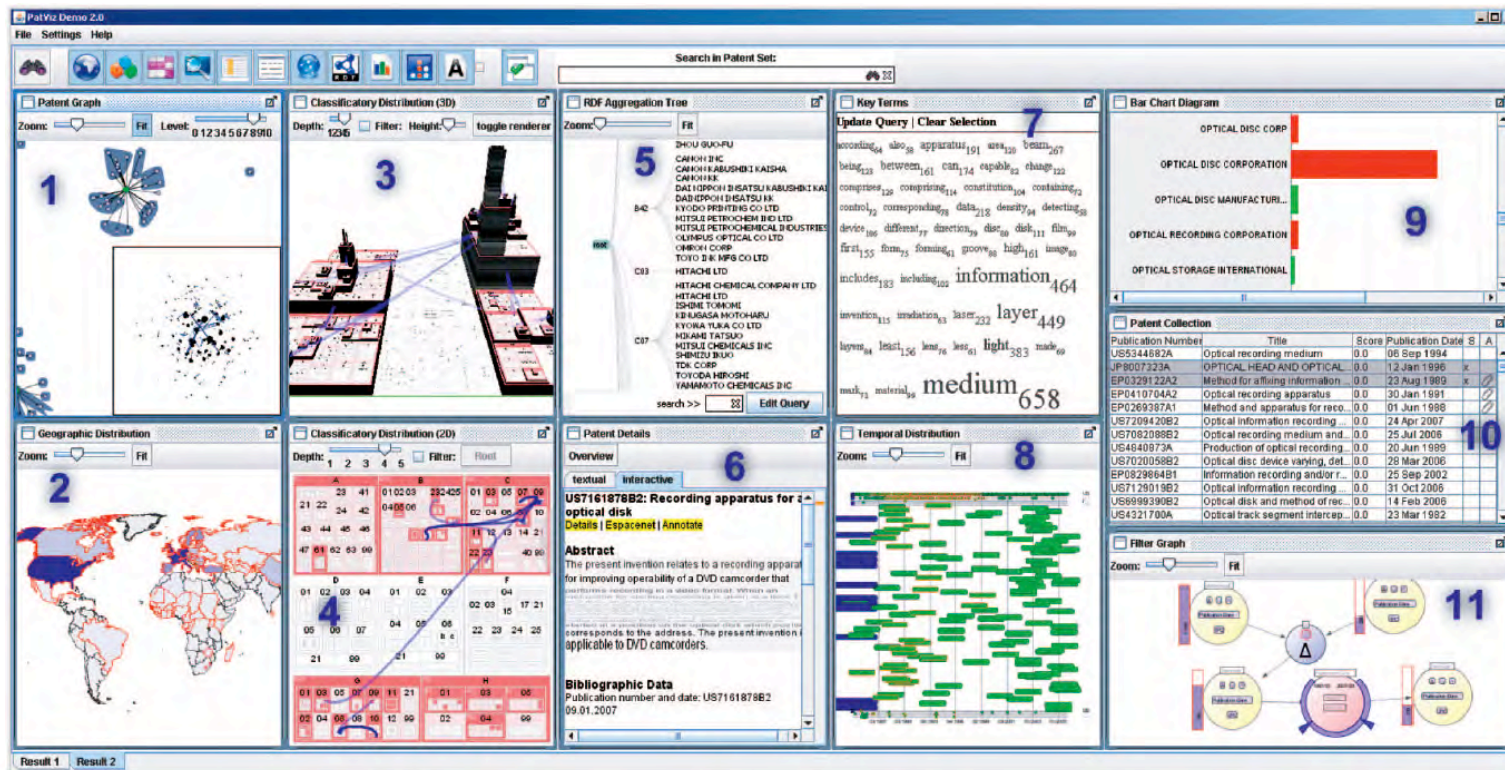
[Sharara, 2011]

# Multiple Connectional

Aggregation | Labelling | Composition | Multiple v. | Tight int.



# Multiple Connectional Aggregation | Labelling | Composition | Multiple v. | Tight int.



# Multiple Connectional

Aggregation | Labelling | Composition | Multiple v. | Tight int.



# Challenges

- Text: +context, -text
- Authors: name disambiguation
- Citations: citation context, sentiment analysis
- Metadata: data-specific techniques (classification, images)
  
- Multiple: scalability, interaction, full VA, predictive analytics
  
- General: evaluation, multidisciplinary approaches

# Challenges

- Text: -text , +context
- Authors: name disambiguation
- Citations: citation context, sentiment analysis
- Metadata: data-specific techniques (classification, images)
  
- Multiple: scalability, interaction, full VA, predictive analytics
  
- General: evaluation, multidisciplinary approaches

