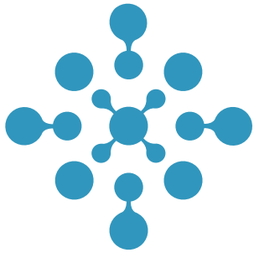


VIS2019

Galex: Exploring the Evolution and Intersection of Disciplines

Zeyu Li, Changhong Zhang, Shichao Jia and Jiawan Zhang
Tianjin University

Background

- The increasing availability of scientific literature data offers a great deal of opportunities to explore the evolution and intersection of a knowledge domain.

Previous work

- They has shown that **visualization** is effective in understanding the literature data.
- They explored …
  - Multiple perspectives
    - by data type: network based, text based
  - Multiple observation levels
    - keywords/topics
    - areas/disciplines
    - the whole science

Motivation

Multiple perspectives

document space
semantic information
+
entity networks
structural information

---

evolution + intersection

+

Multiple observation levels

discipline level

↓ zoom in

area level

↓ filtering

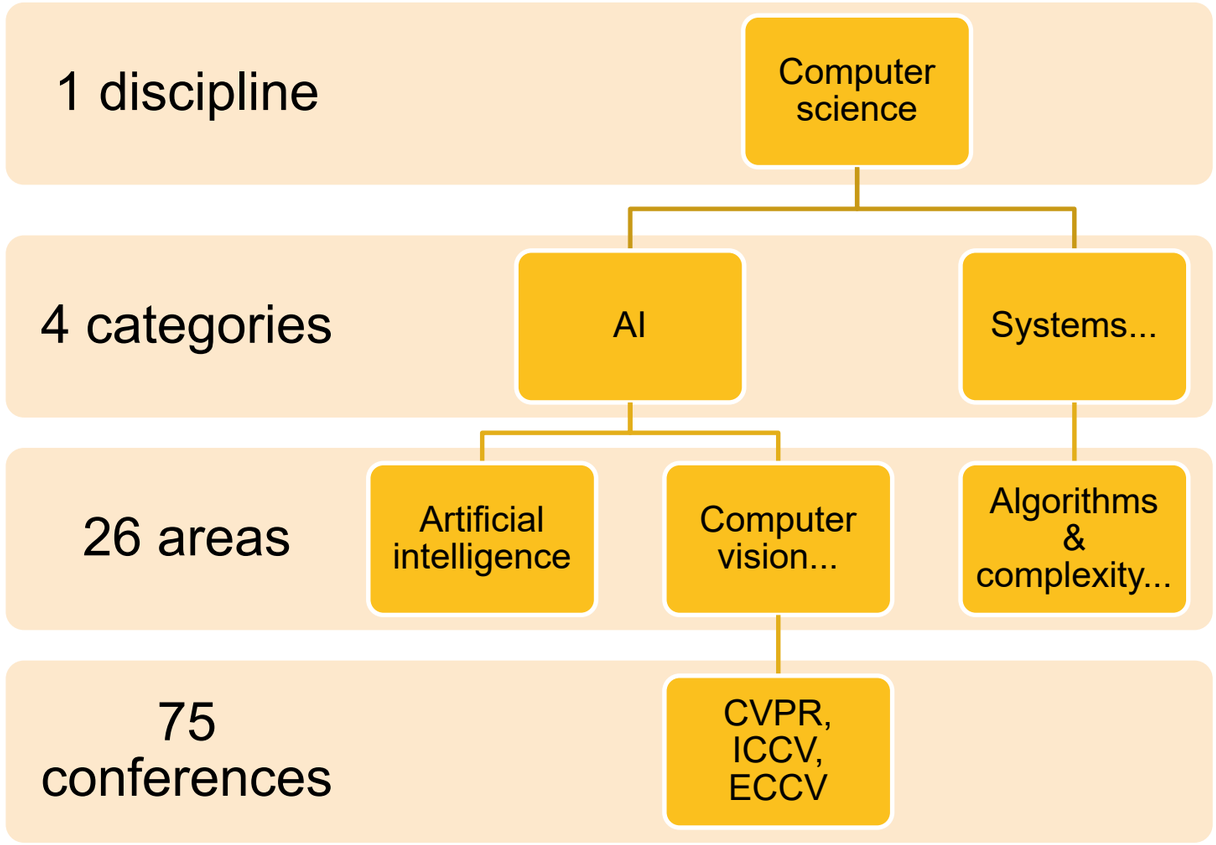
Institution level

Tasks

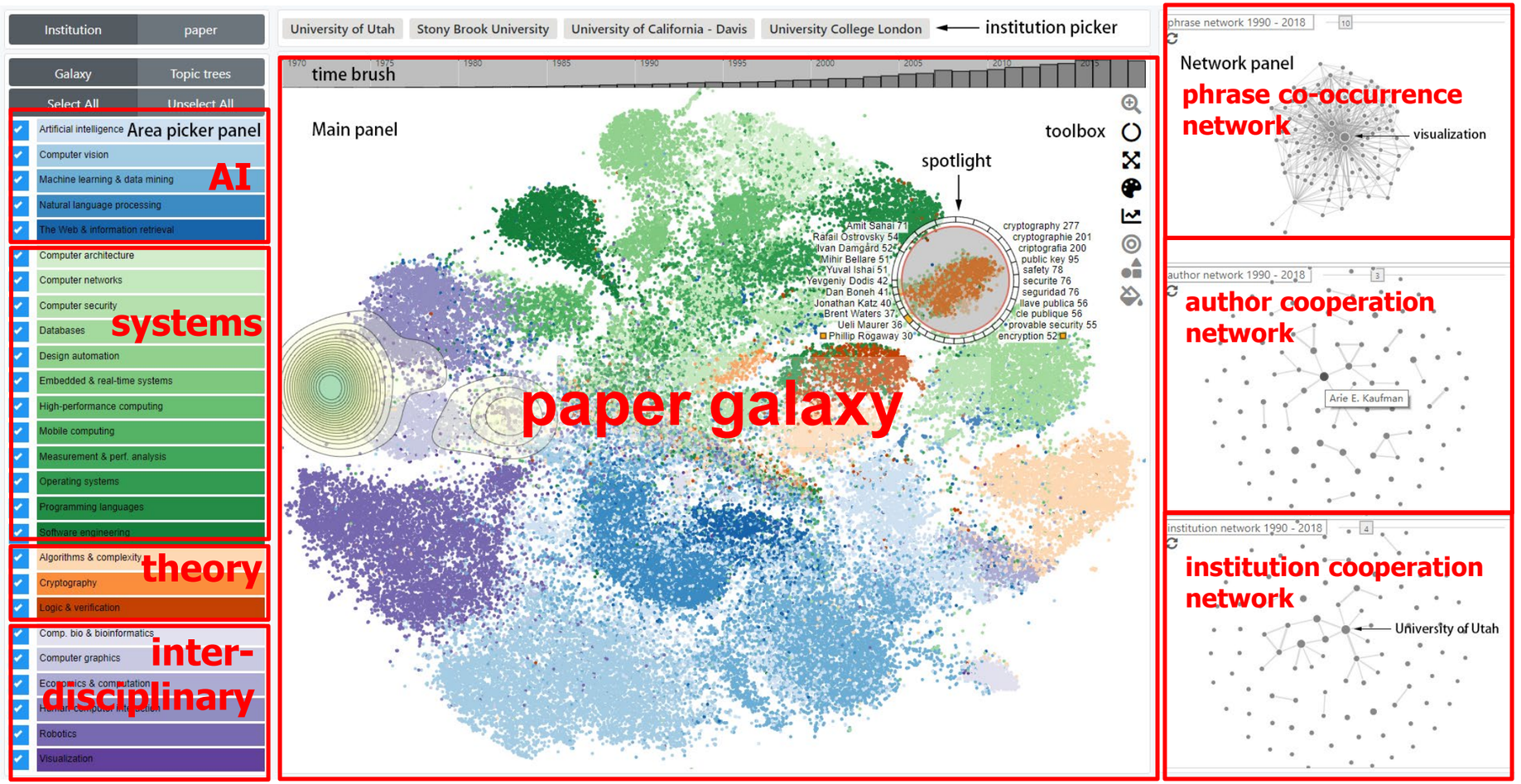
- Understand the content of a field
  - sub-fields, hotspots and entities…
- Perceive the structure of a field
  - entity networks, academic communities…
- Comprehend the evolution of a field
  - origin, rise and decline…
- Detect the intersection between sub-fields
  - papers and common topics that act as bridge…
- Compare the performance among institutions
  - topic distributions

Data description

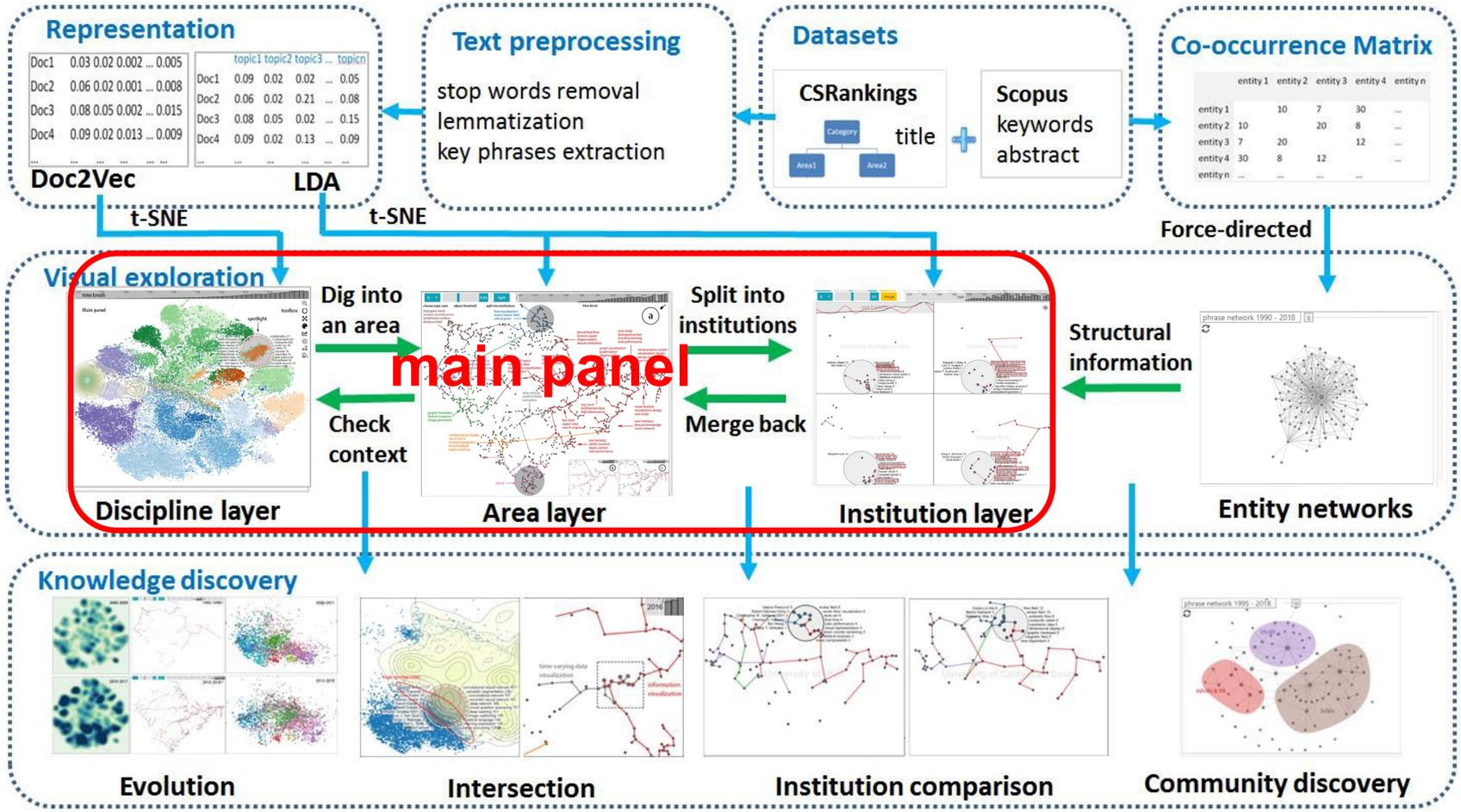
- computer science as an example
- over 86, 000 papers, 1970-2018
  - **CSRankings**
    - title, authors
    - institutions
    - published year
  - **Scopus(text data)**
    - title
    - abstract
    - keywords



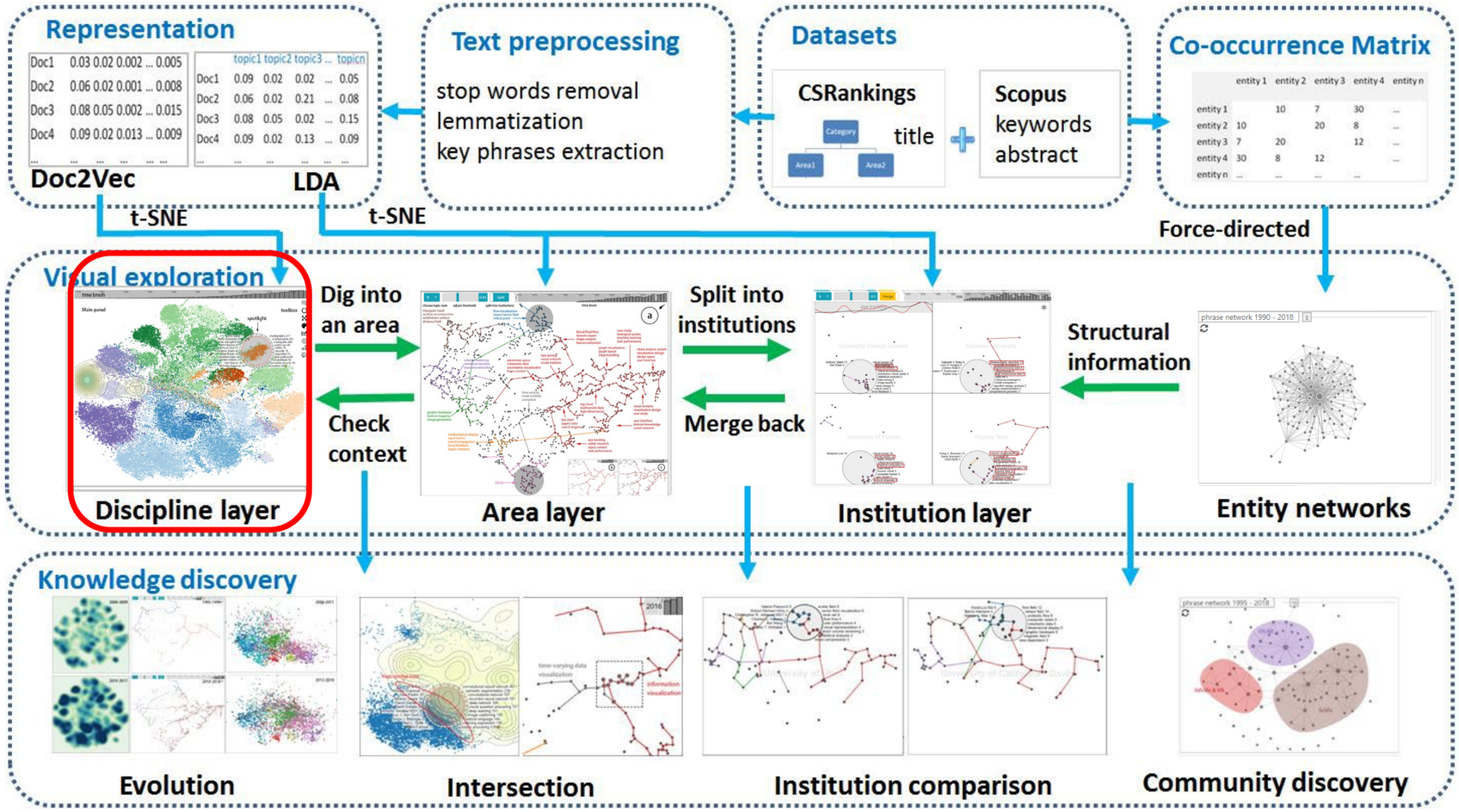
Galex(Galaxy evolution explorer)



Pipeline of Galex

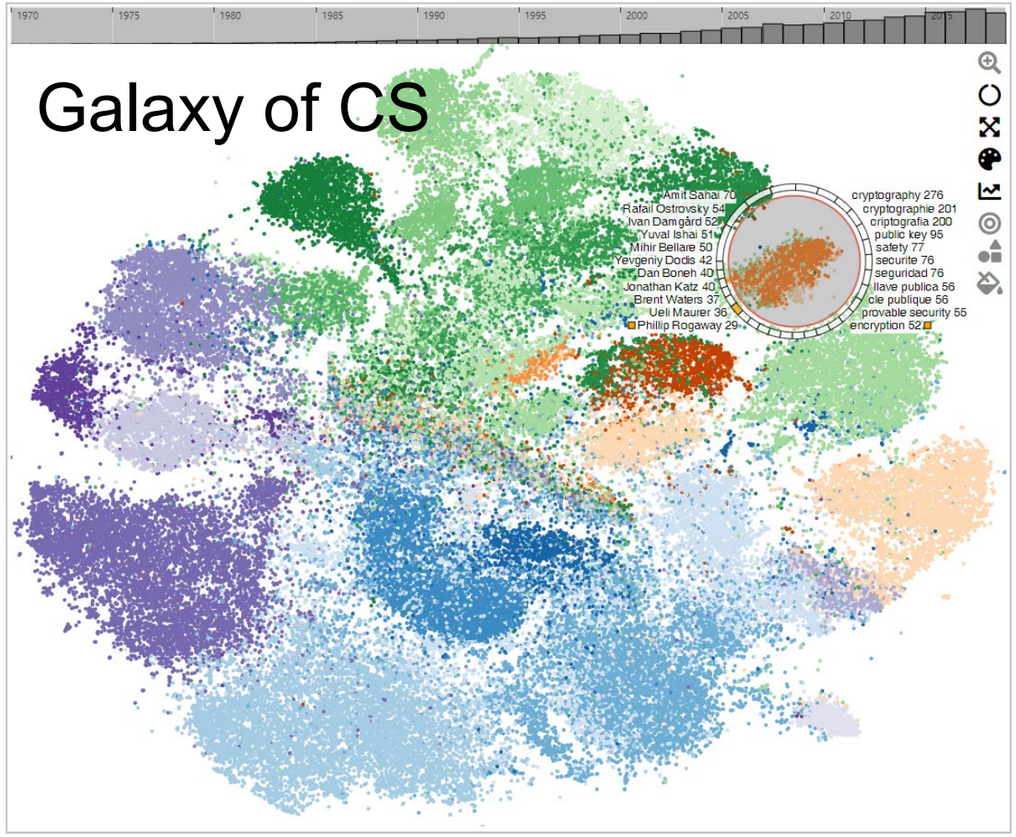


Pipeline of Galex



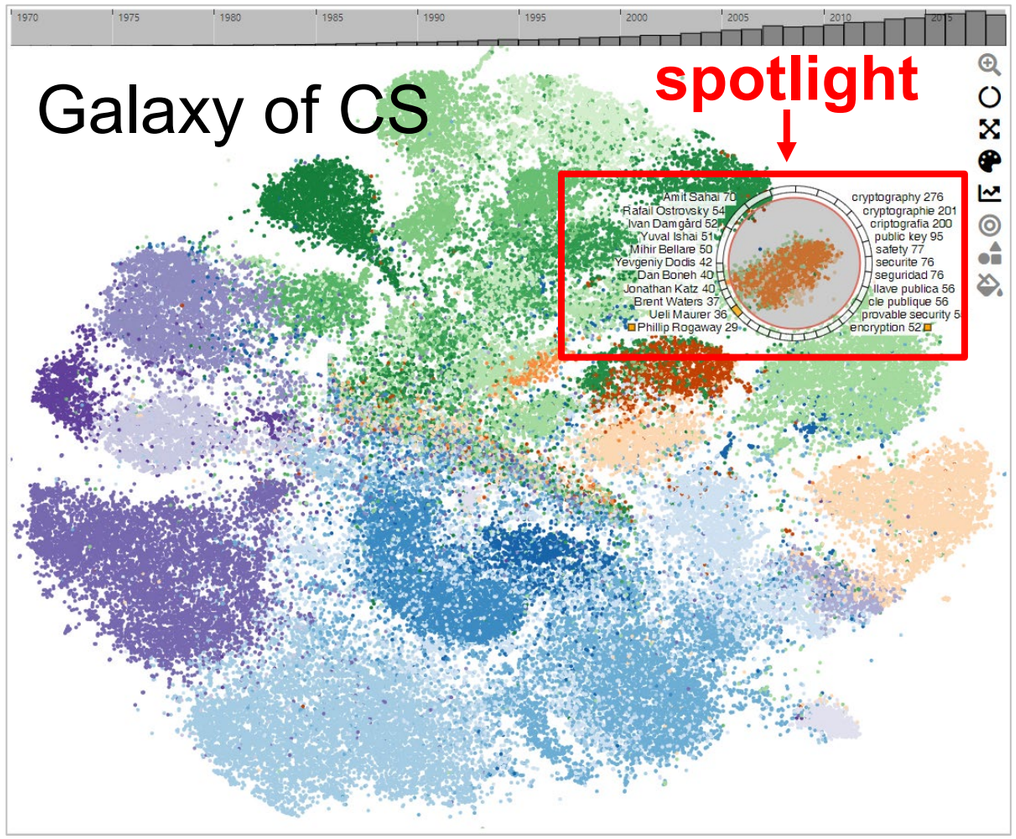
Discipline layer

- **Task**: present an overview of computer science
- **Solution**: **galaxy of computer science**
  - text data —doc2vec→ paper vectors —t-SNE→ coordinates
  - dots represent papers
  - colored clusters represent areas



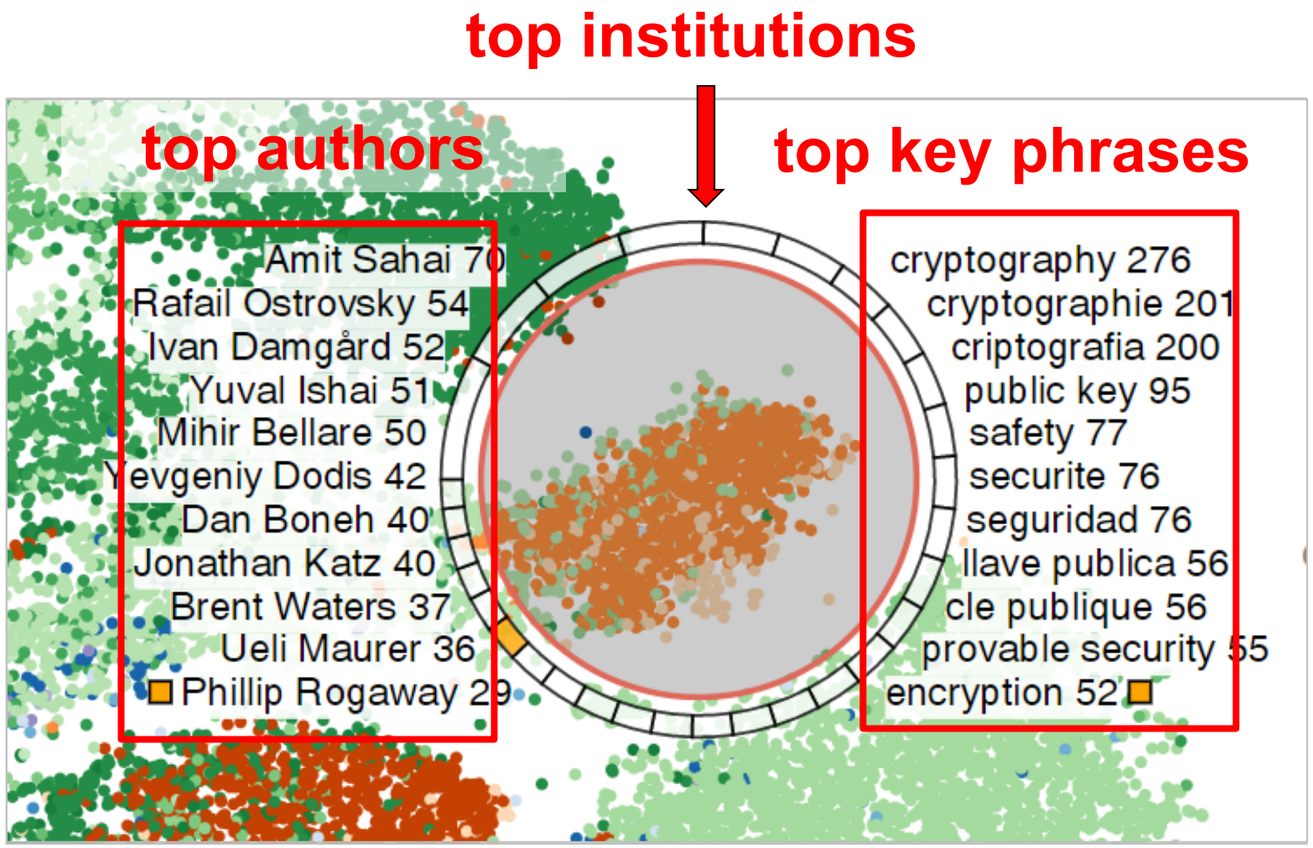
Discipline layer

- **Task**: interpret the galaxy
- **Solution**: **spotlight**
  - can be moved freely
  - entities are shown around it



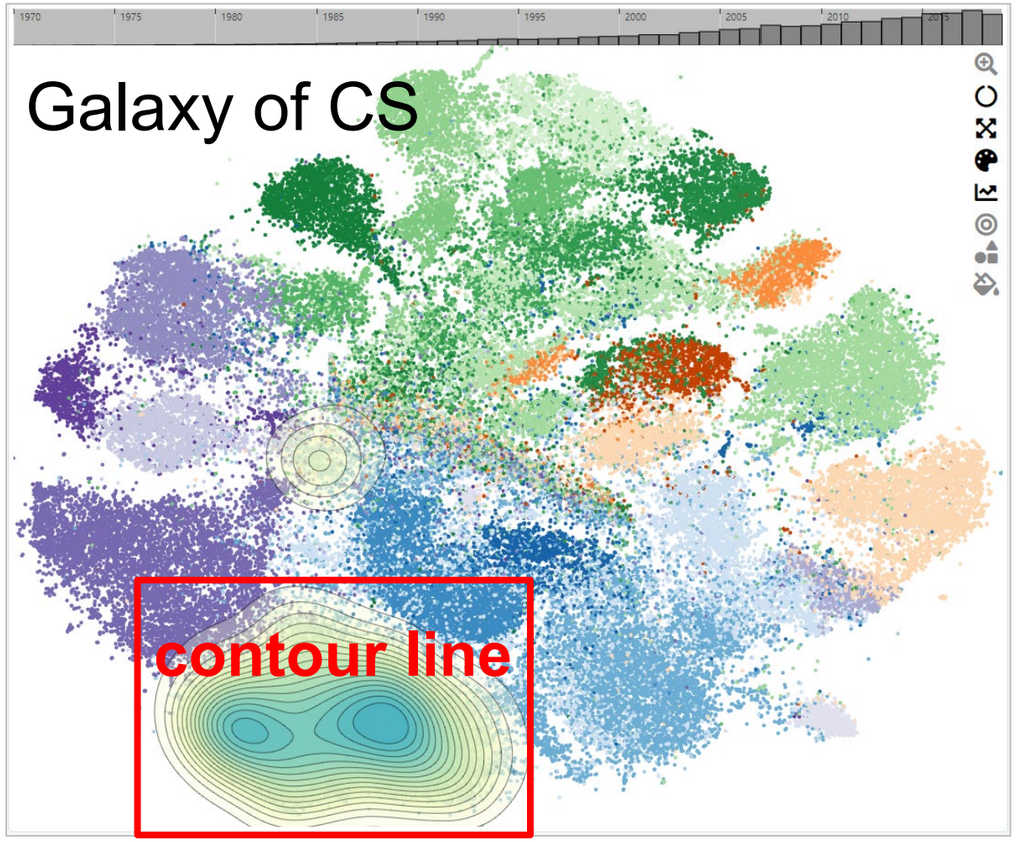
Discipline layer

- **Task**: interpret the galaxy
- **Solution**: **spotlight**
  - can be moved freely
  - entities are shown around it



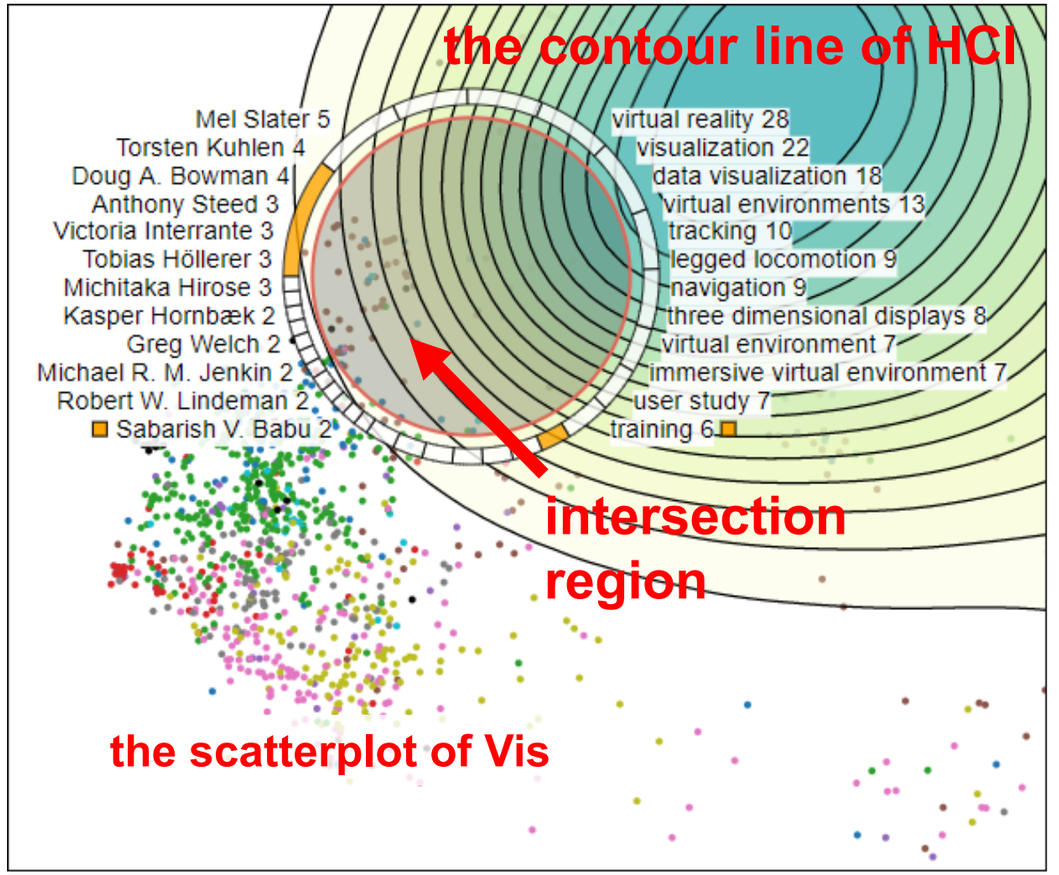
Discipline layer

- **Task**: quickly locate an area and perceive its paper distribution
- **Solution**: **contour line**
  - Color encodes the density of papers.
  - It has extra benefits, it helps to
    - identify the cores
    - detect the intersections between two areas



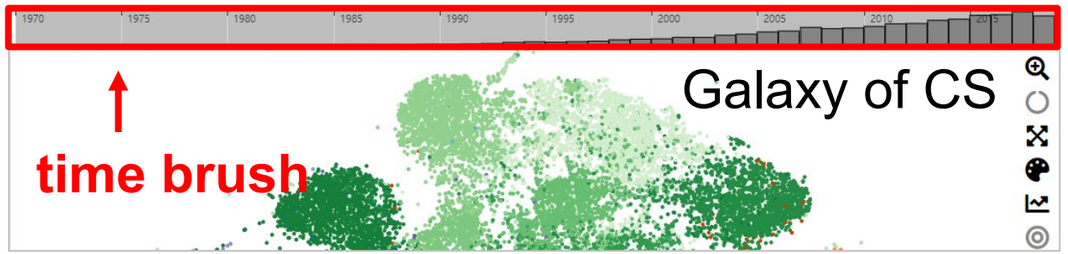
Discipline layer

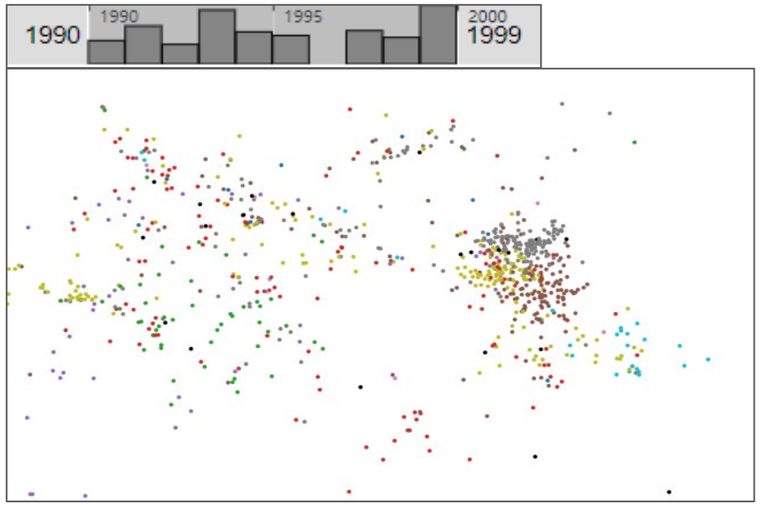
- **Task**: detect the intersection between two areas
- **Solution**: **contour line + scatterplot**
  - Overlay the contour line of an area on the scatterplot of another area.



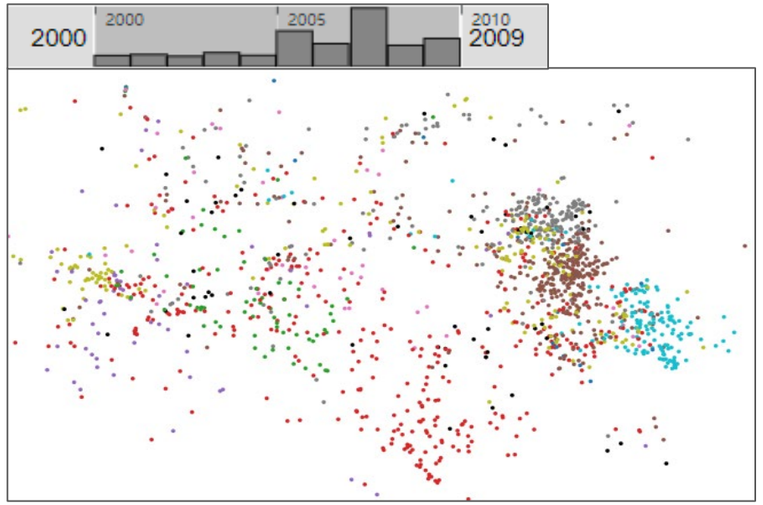
Discipline layer

- **Task**: check the evolution of selected areas
- **Solution1**: **time brush**
  - It can be dragged and re-sized freely.



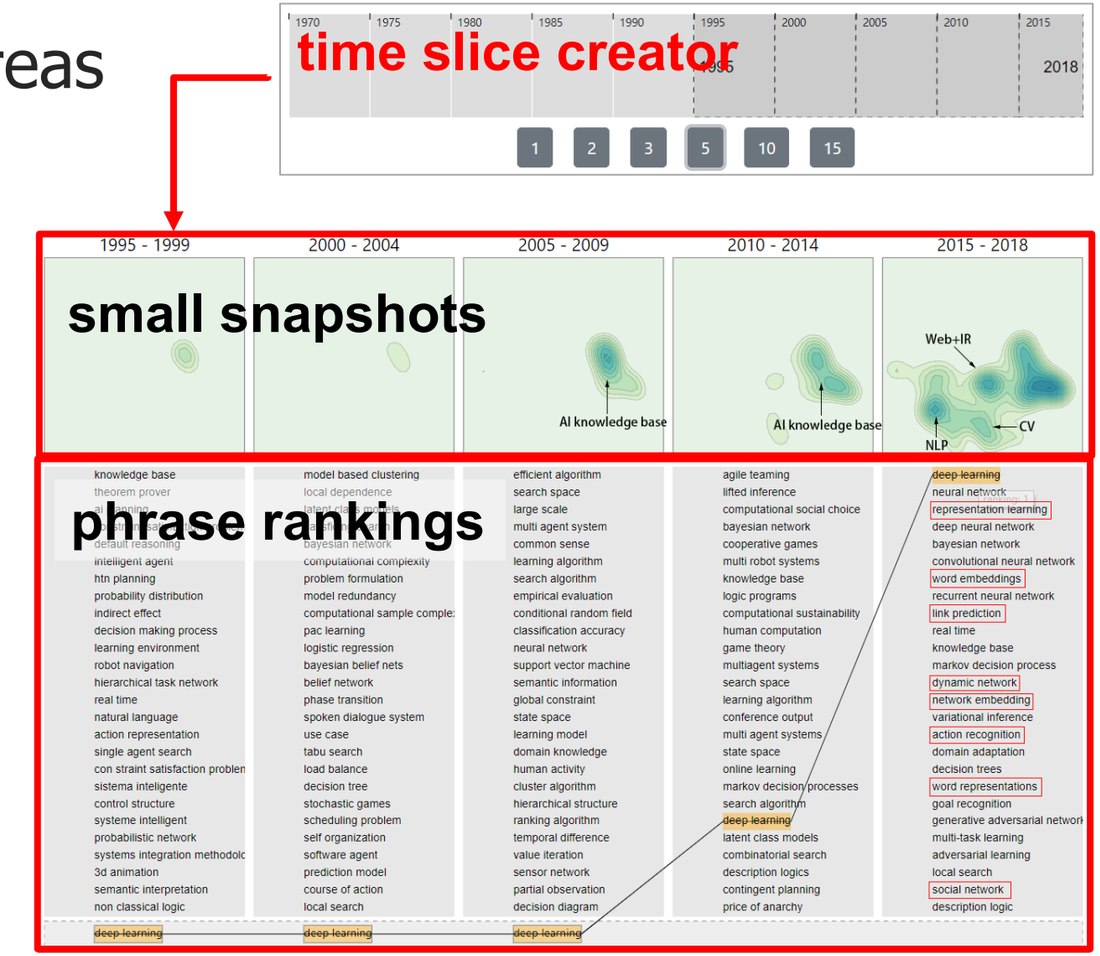


from 1990s to 2000s

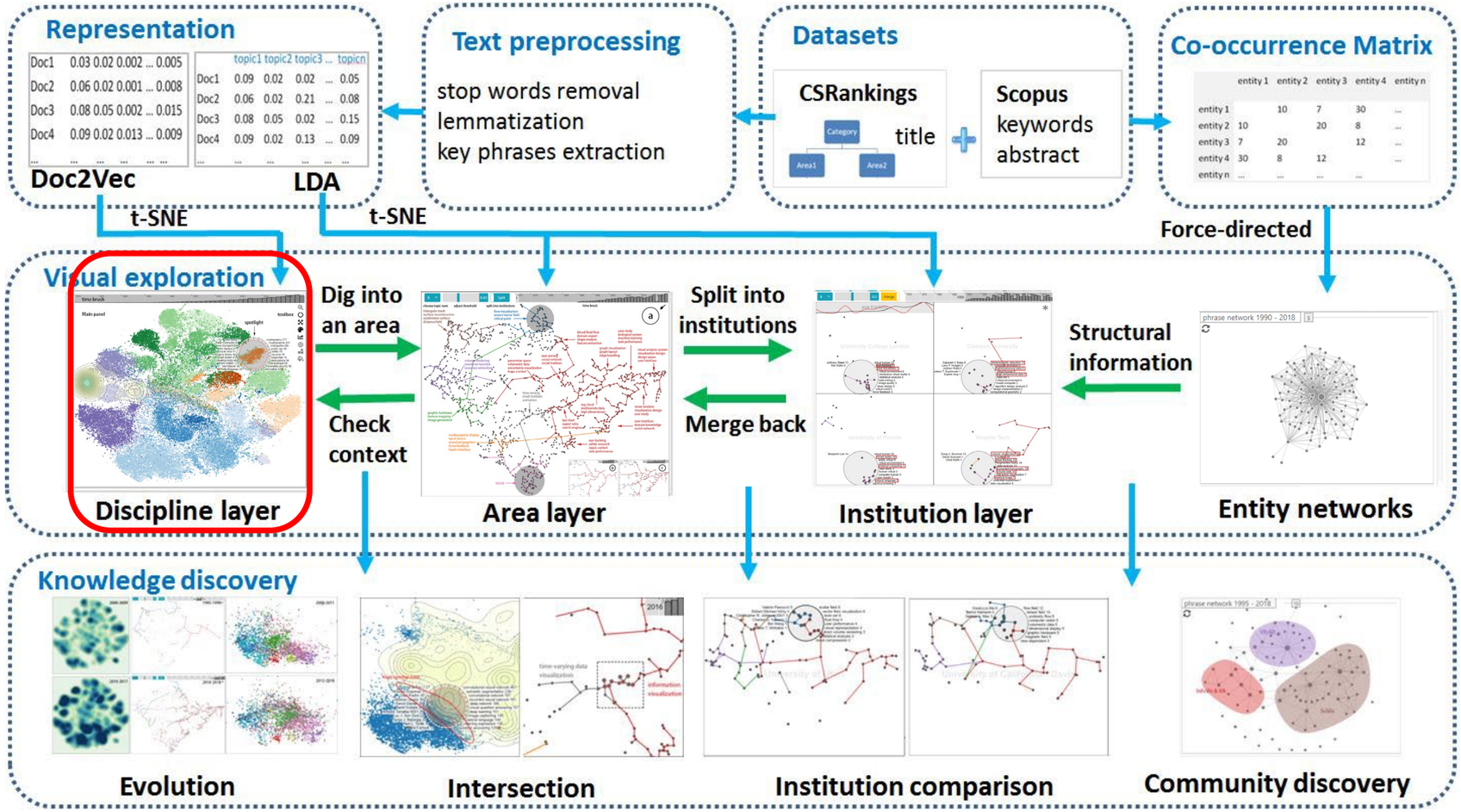


Discipline layer

- **Task**: check the evolution of selected areas
- **Solution2**: **time slice creator**
  - It creates equal-length time slices
  - small snapshot: high-level description
  - phrase rankings: detail interpretation

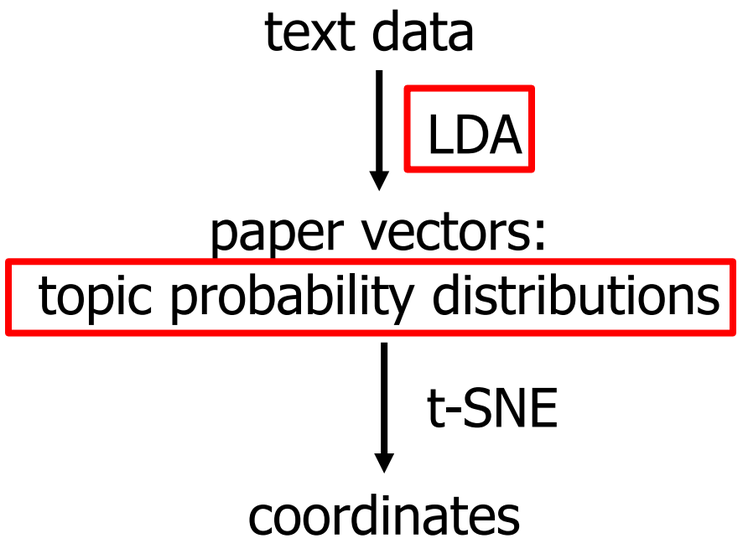


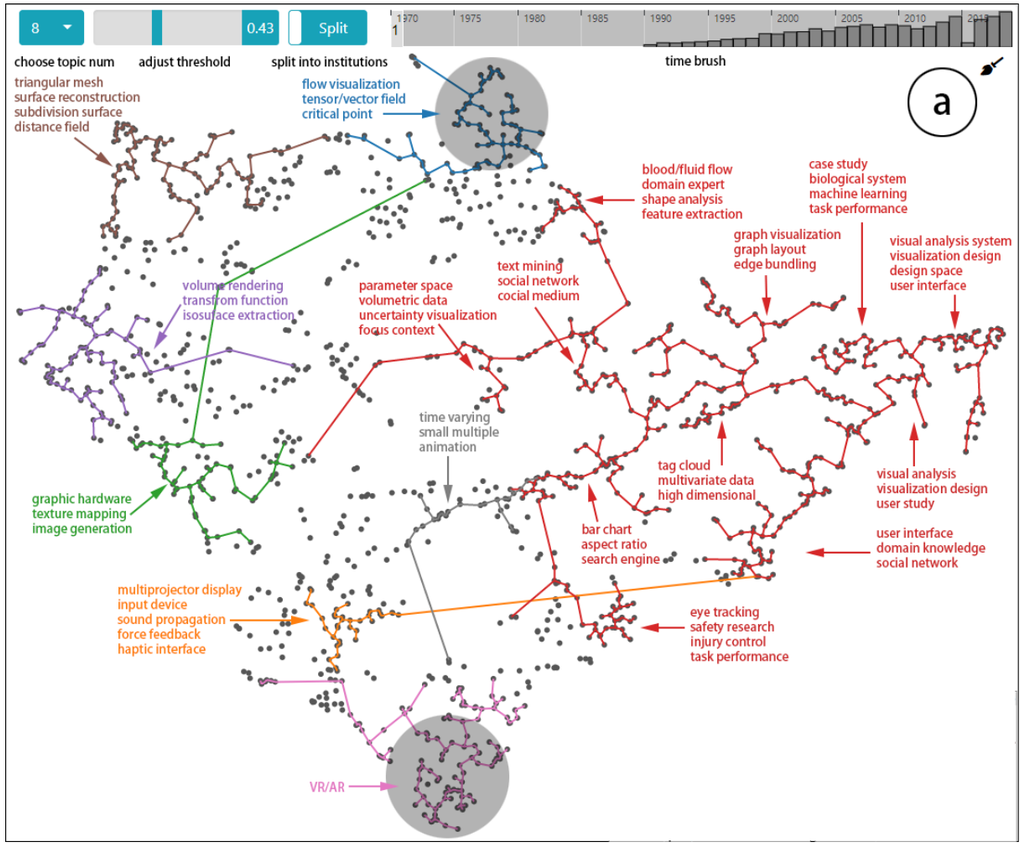
Pipeline of Galex



Area layer

- **Task**: get an overview of an area
- **Solution**: **LDA model +t-SNE**





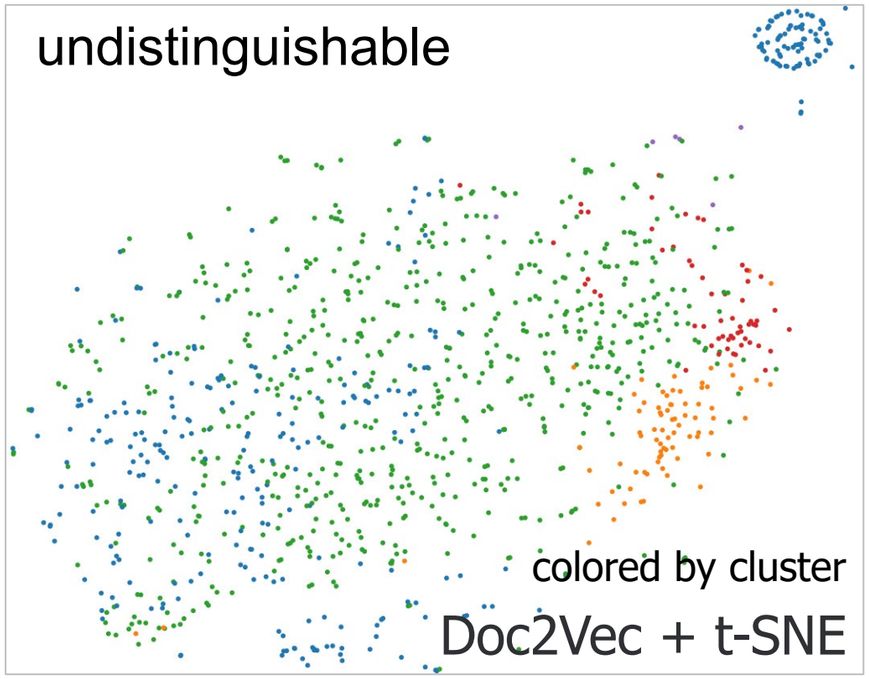
Design decision:

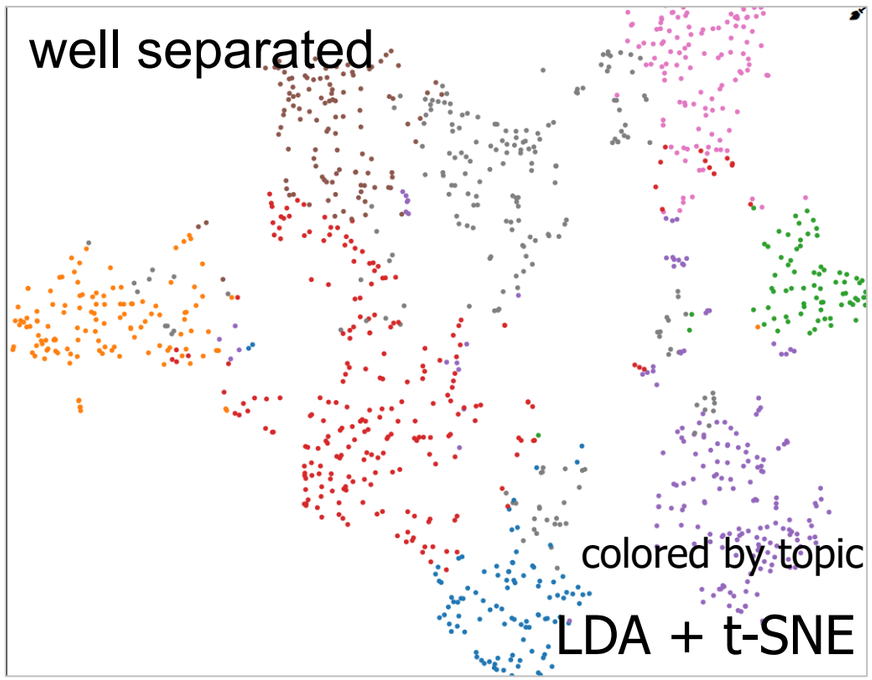
Why do we abandon the vectors trained by doc2vec in the area layer?

Area layer

Design decision:

Vectors trained by doc2vec emphasize the diversity between areas, but they are not good at distinguishing the topics in a specific area.

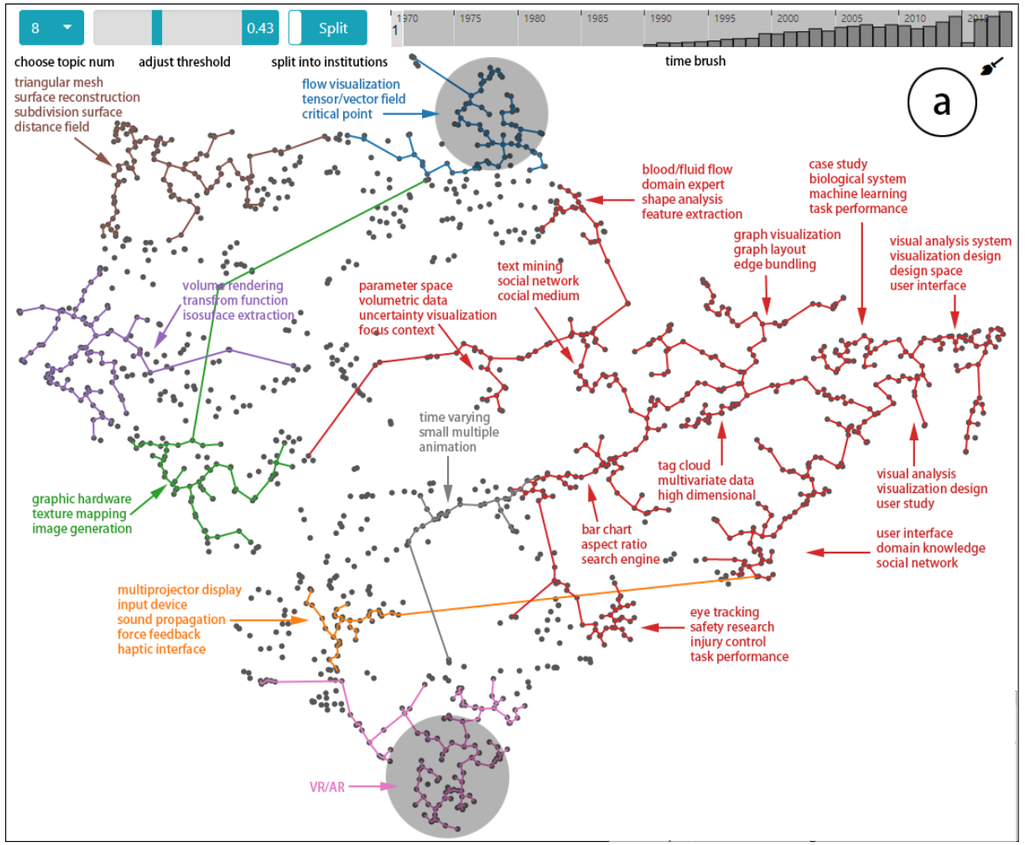




Area layer

- **Task**: get an overview of a specific area
- **Solution**: **LDA model + topic trees**

  - Papers belong to the same topic are connected by a minimum spanning tree (MST, topic tree) of that topic, using the 2D coordinates of papers.



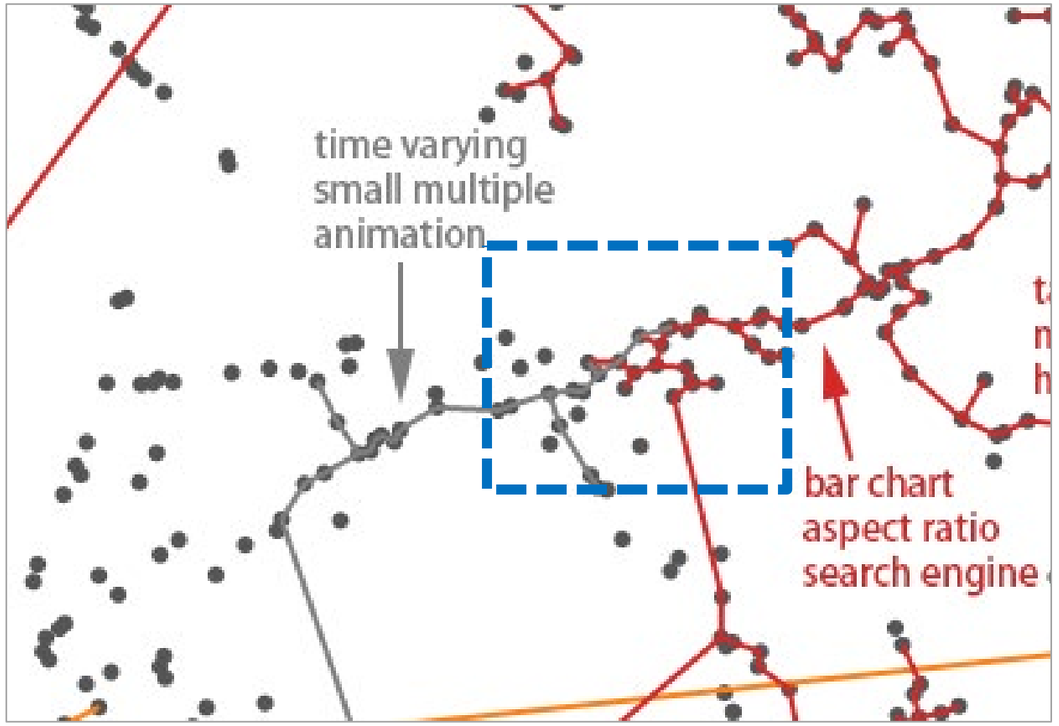
Design decision:

Why do we use the MSTs to represent topics rather than simply coloring papers by its leading topic?

Area layer

- **Task**: check the intersection between topics
- **Solution**: **topic trees**

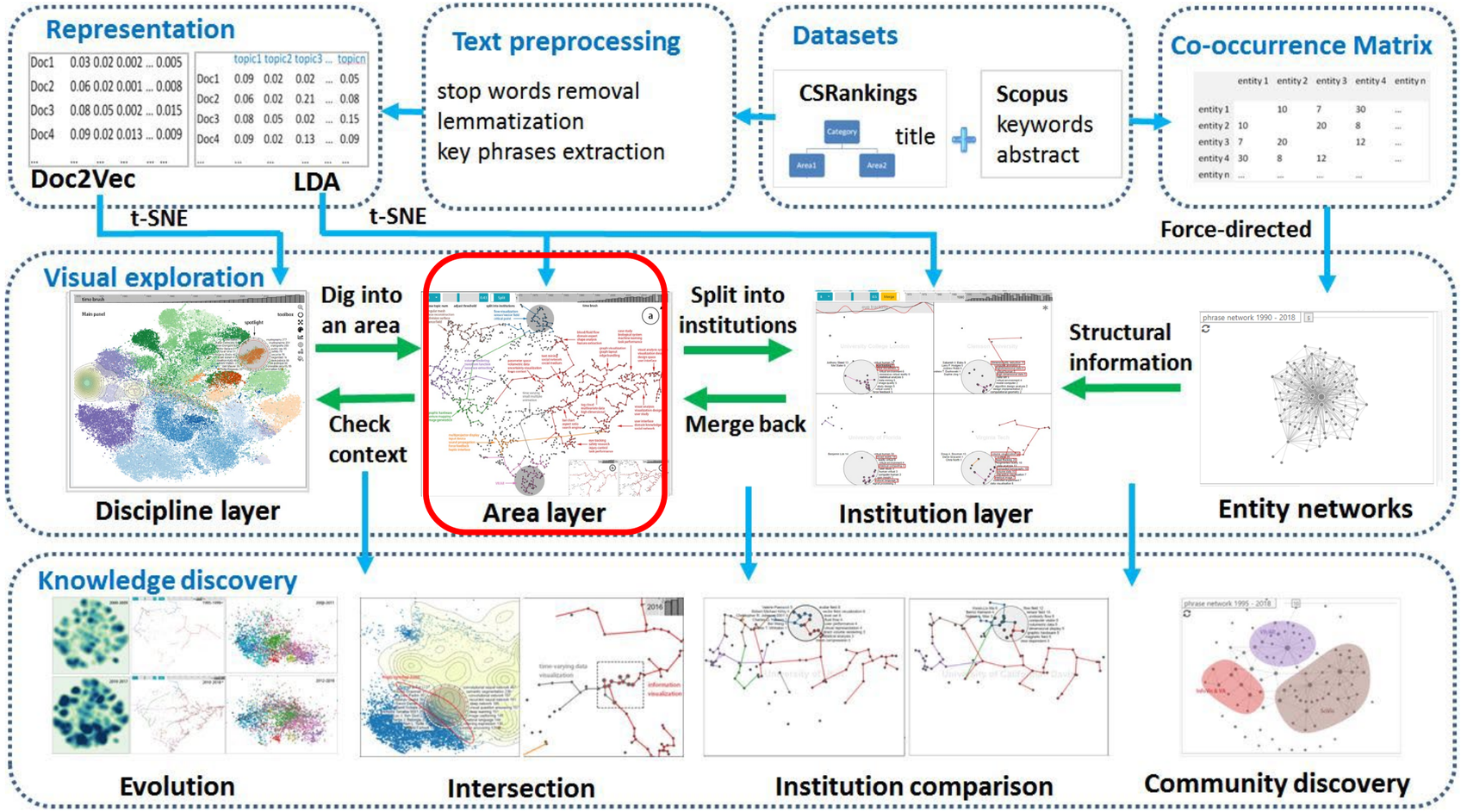
Papers that are connected by different colored lines are those that involve multiple topics.



Area layer

- **Task**: learn the evolution of topics
- **Solution**: **topic trees + time brush**
  - The evolution of topics can be represented by the growth of the topics trees.

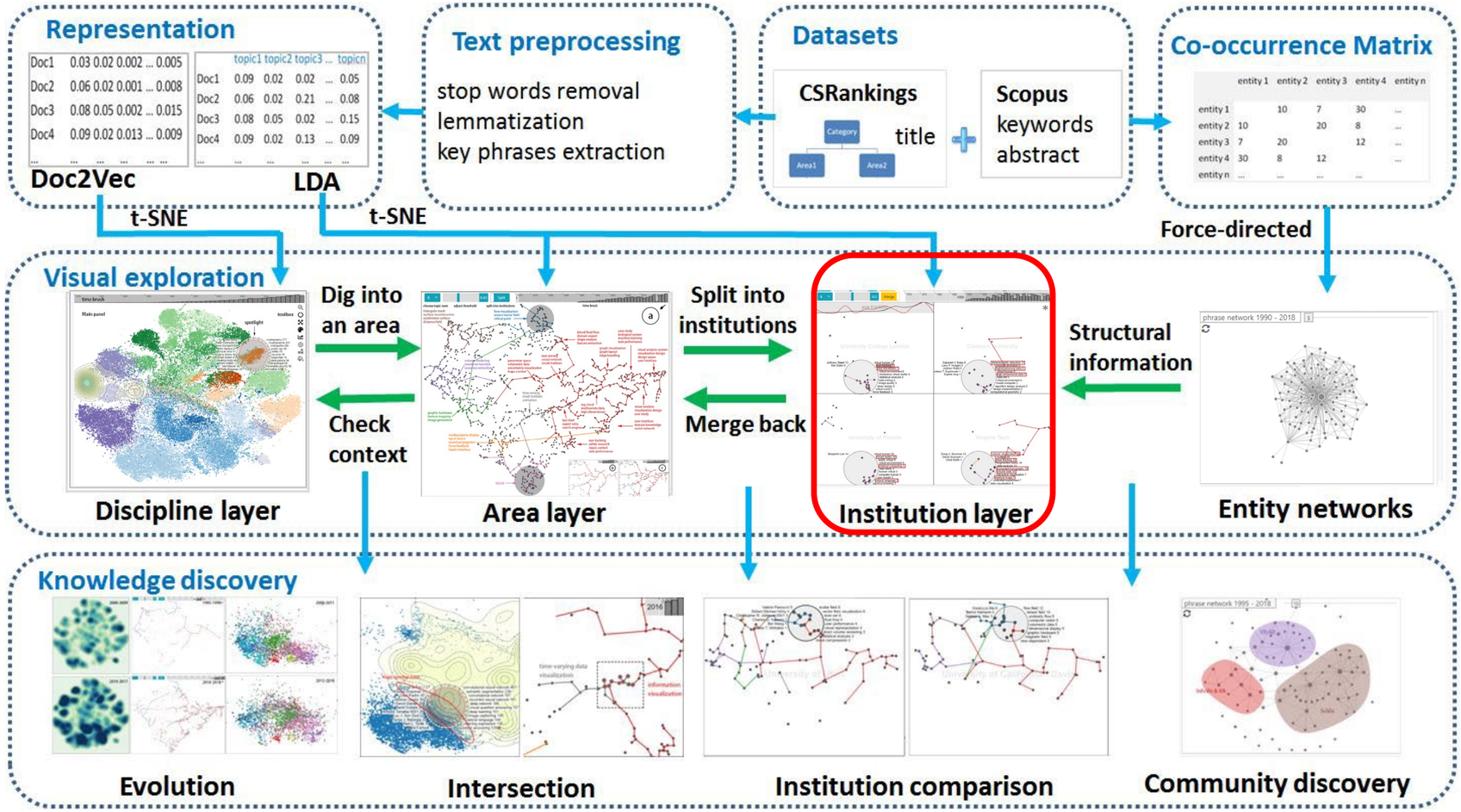
Pipeline of Galex



Institution layer

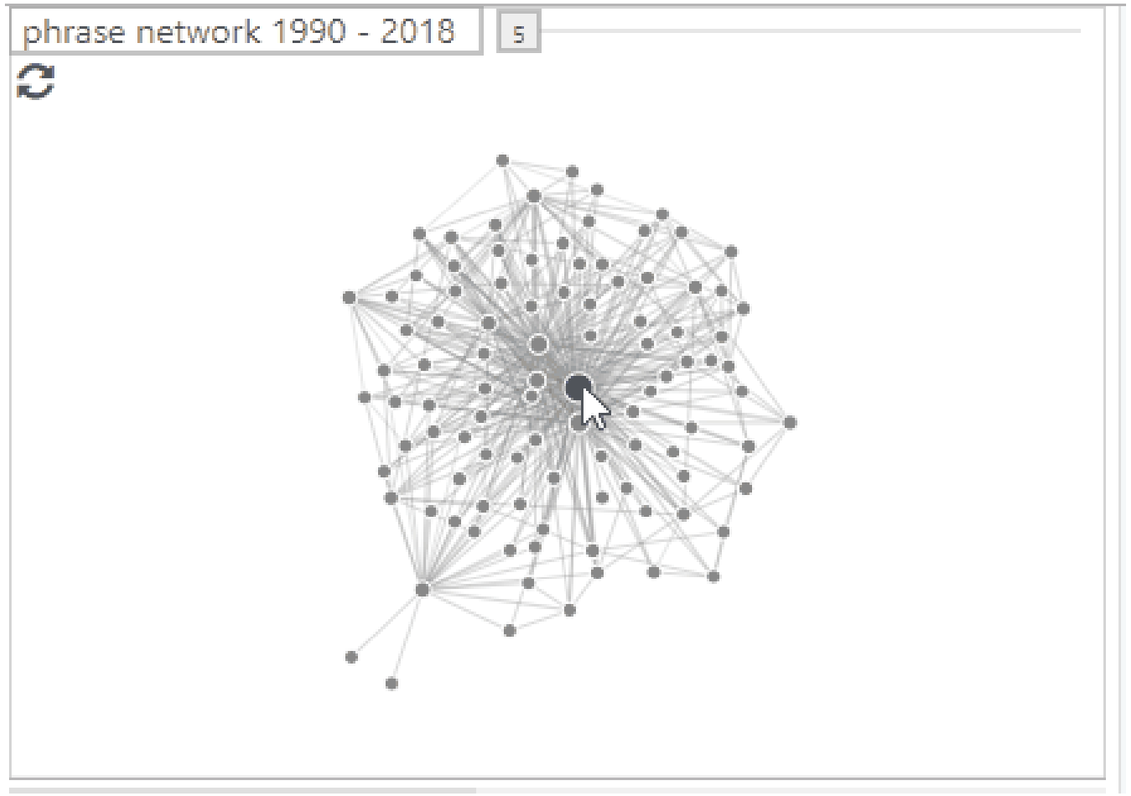
- **Task**: compare the research contents between institutions
- **Solution**: **synchronous spotlights**

Pipeline of Galex

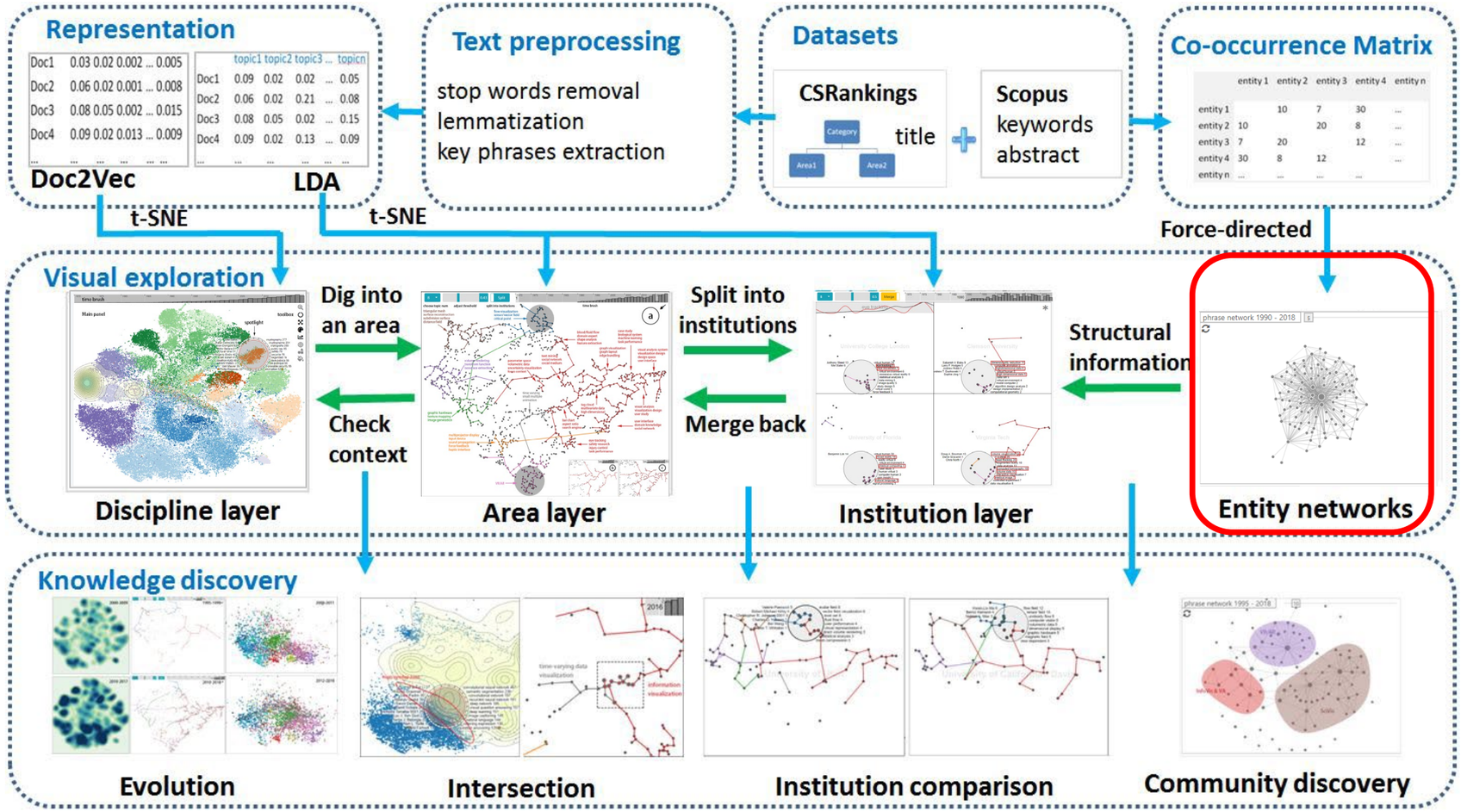


Entity networks

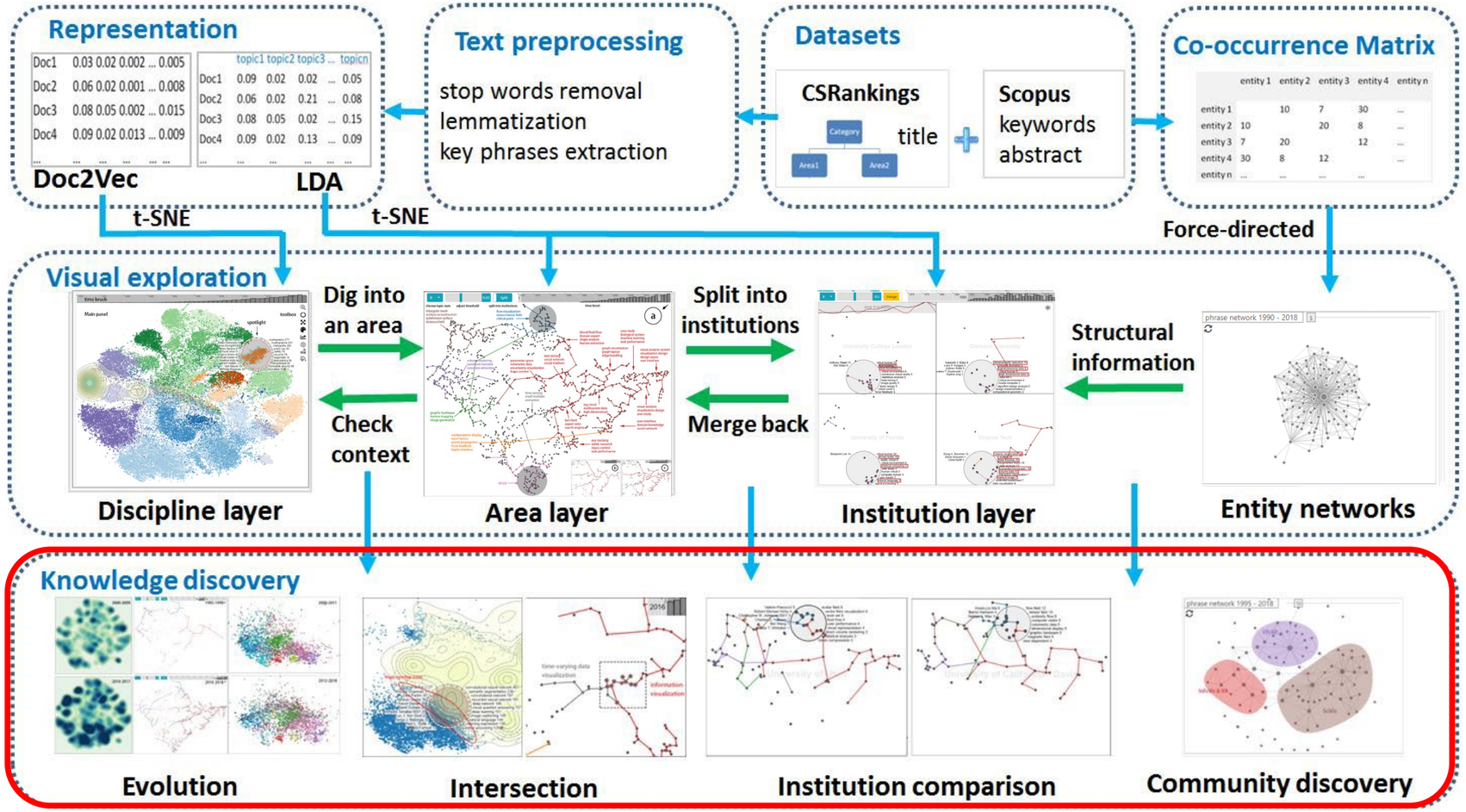
- **Task**: check the relationship between entities and detect communities
- **Solution**: **networks + interactions**
  - Interactions:
    - filter links by weight
    - delete meaningless phrases



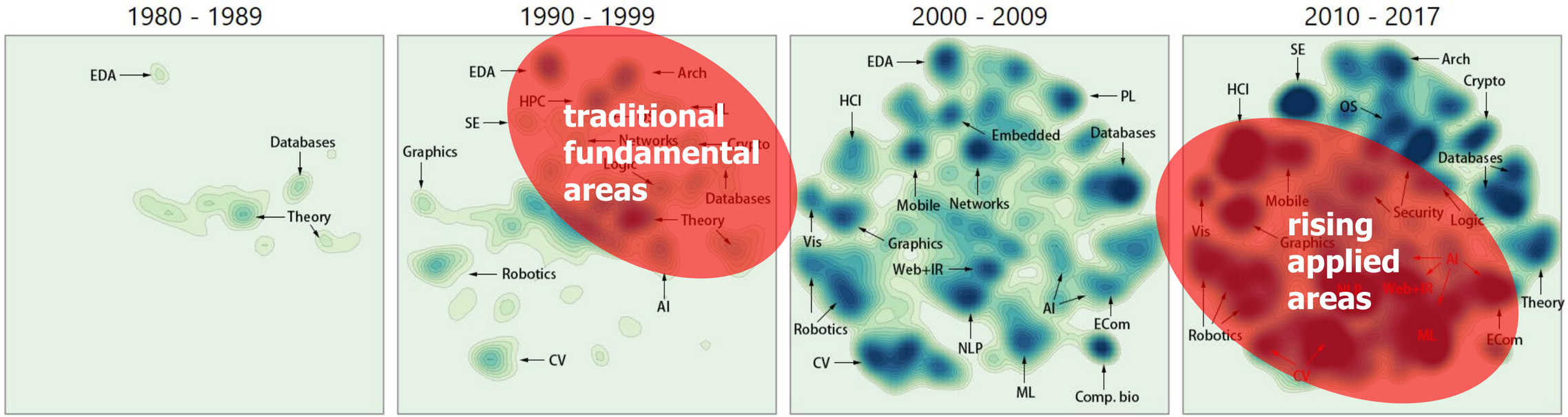
Pipeline of Galex



Pipeline of Galex



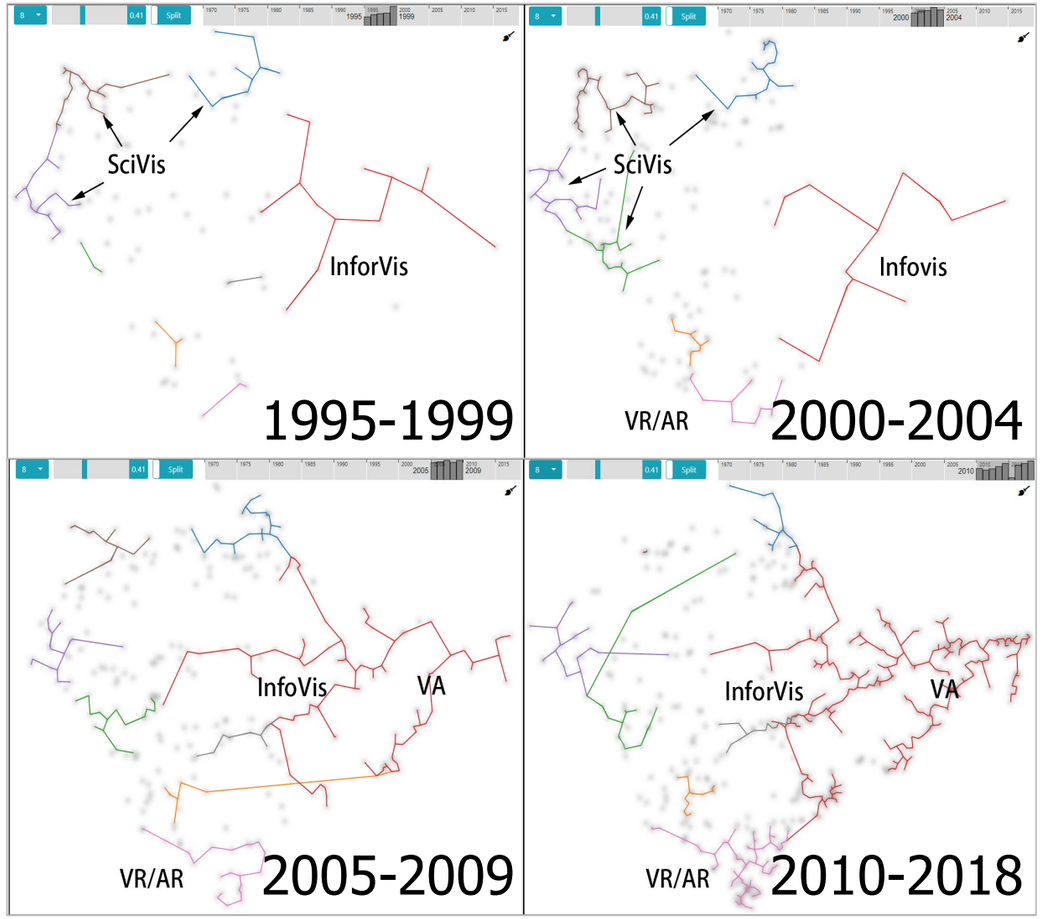
Evolution of entire computer science



The core force driving the development of computer science changed from **the traditional fundamental areas** to the **rising applied areas**.

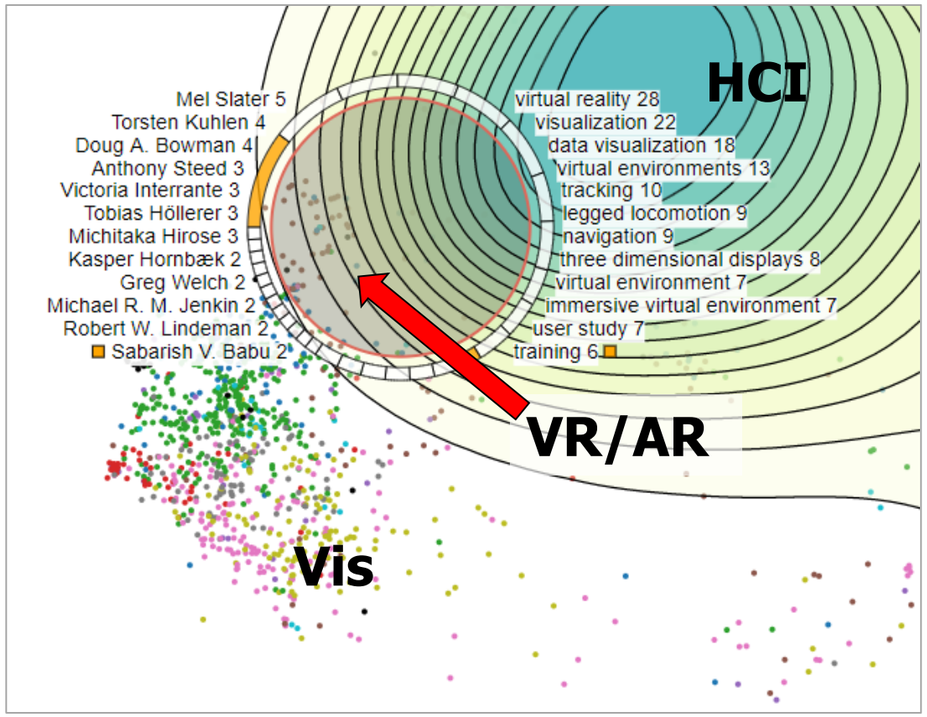
Evolution of visualization area

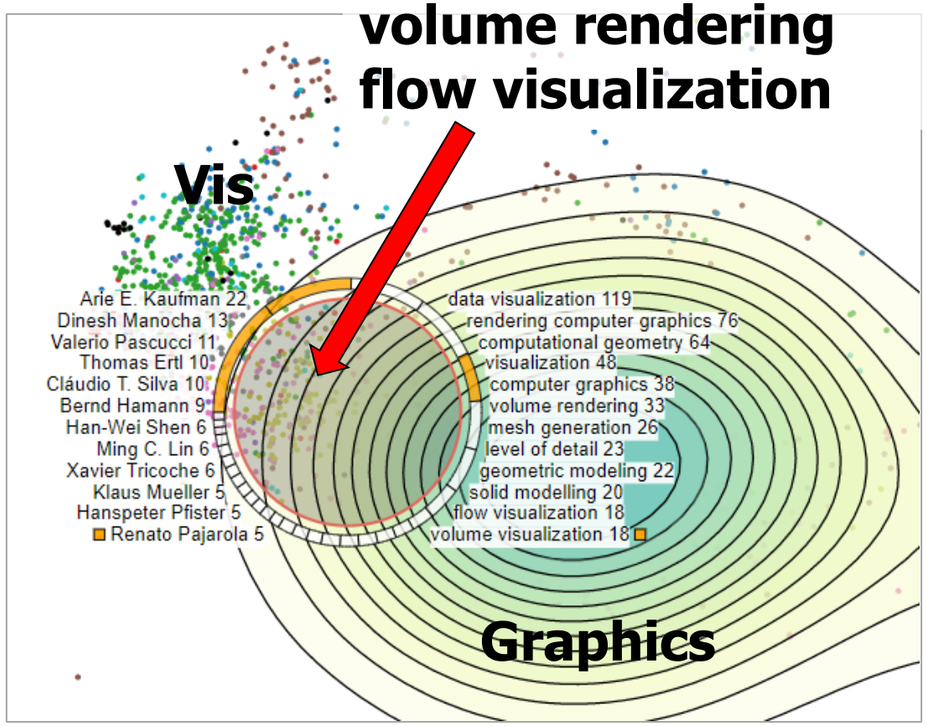
Before 2004, SciVis lead the research topics of Vis area, and for the next 15 years, VA & InfoVis and VR & AR grew rapidly.



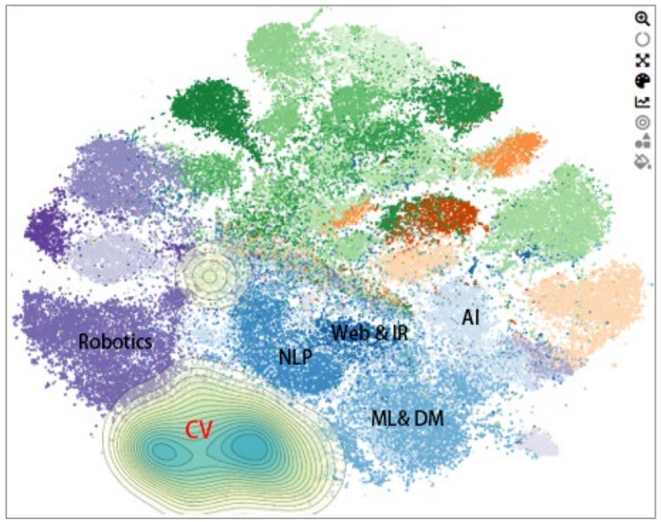
Intersection between Vis and HCI/Graphics

- Overlay the contour line of HCI/Computer Graphics area on Visualization area (scatterplot).

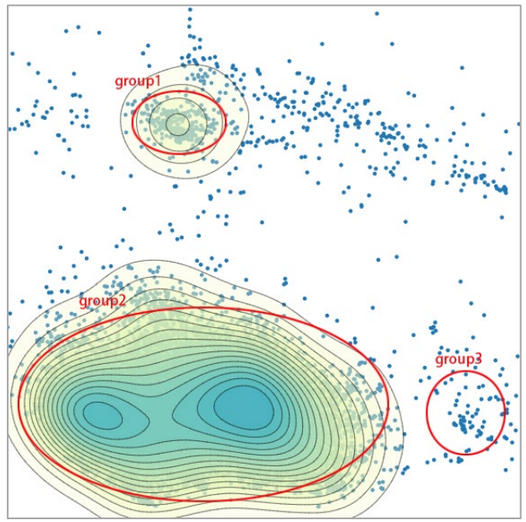




Understand Computer Vision



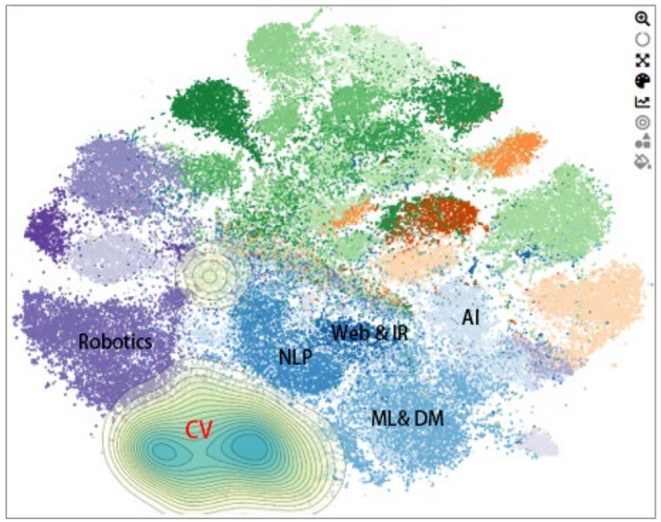
check the location of CV

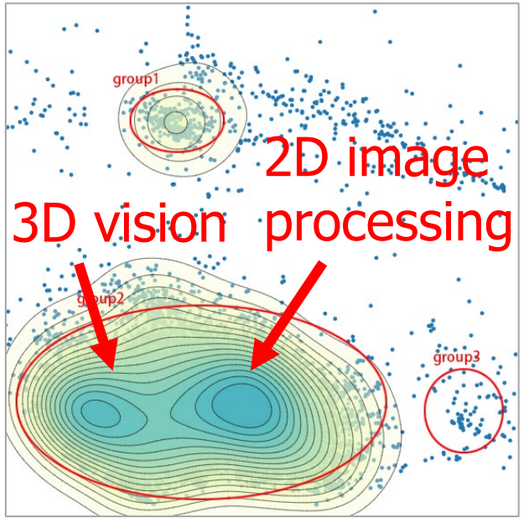


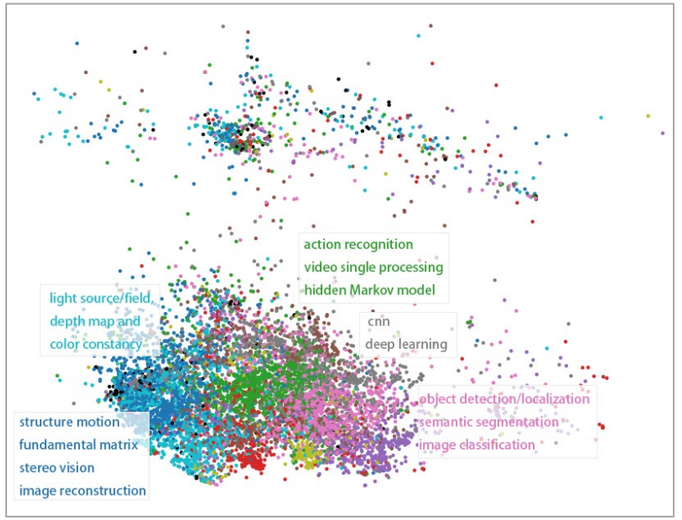
find two cores

hide other areas,
overlay contour line of CV

Understand Computer Vision







check the location of CV

find two cores

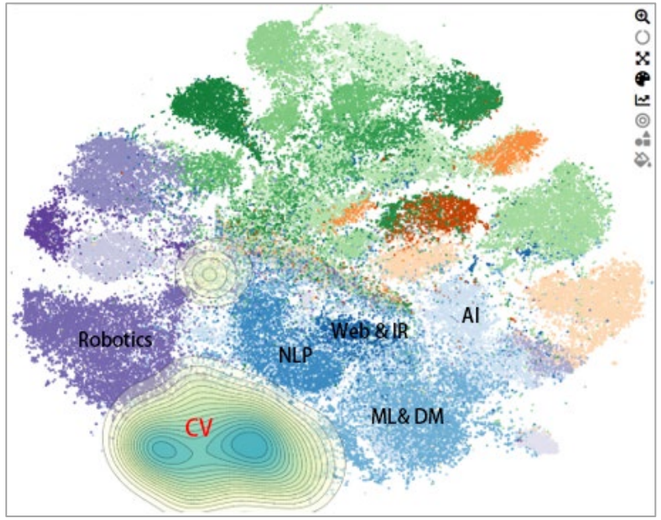
check the topics

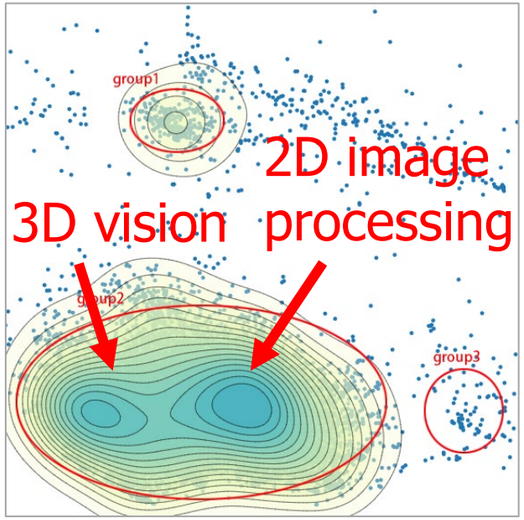
interpret

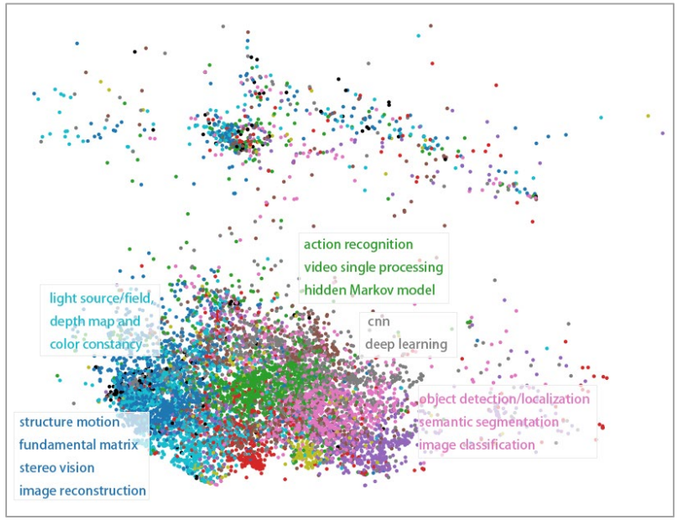
hide other areas,
overlay contour line of CV

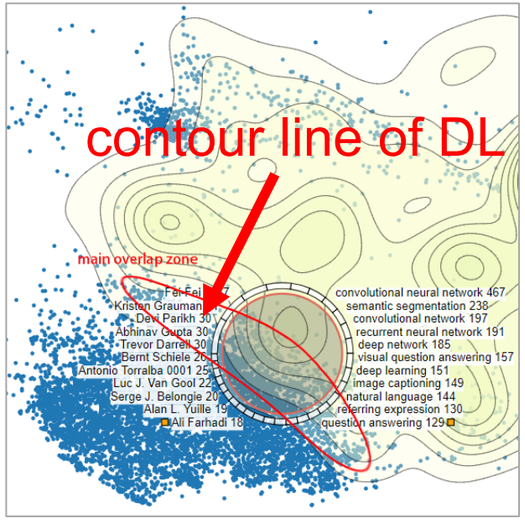
enable topic assistant,
use spotlight

Understand Computer Vision









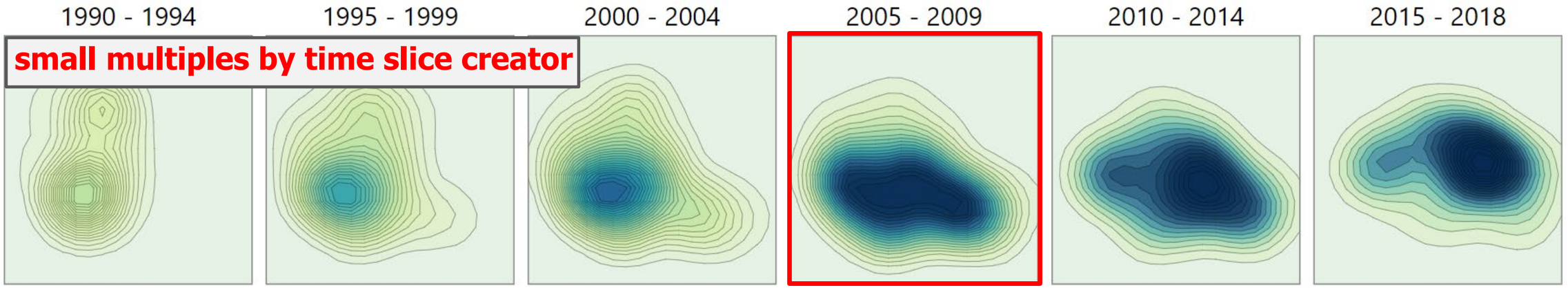
check the location of CV → find two cores ← (interpret) check the topics ← (verify) check the intersections

hide other areas, overlay contour line of CV

enable topic assistant, use spotlight

overlay contour line of AI/ML area

Evolution of Computer Vision

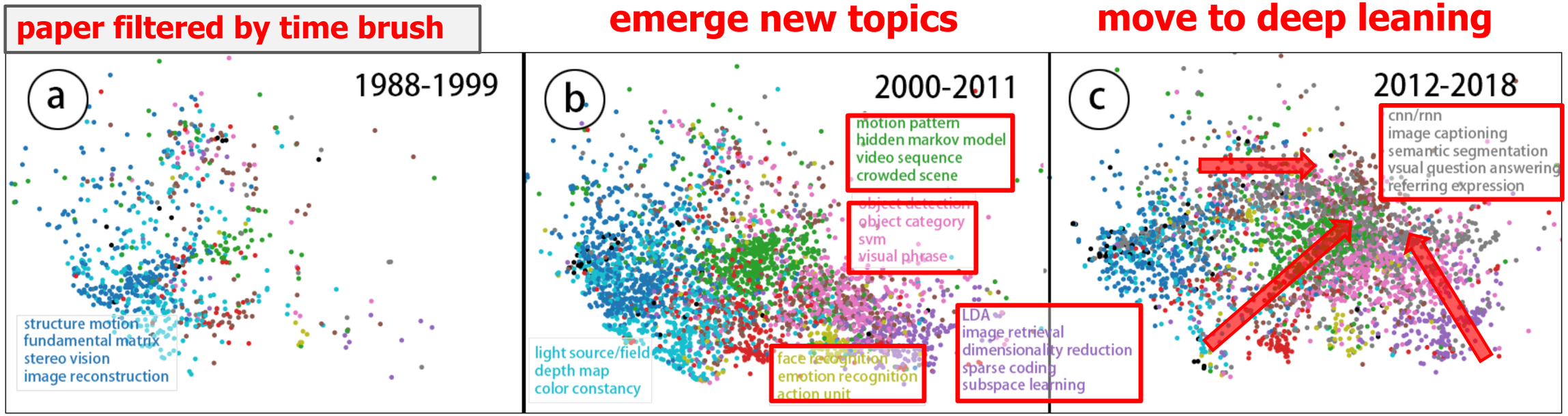


The hotspot shows a noticeable **right shift.**

Is this simply caused by deep learning techniques?

No! The rise of deep learning techniques in academia began around 2012. However, the right shift is already significant in the period of 2005-2009.

Evolution of Computer Vision



The **first shift** is **topic-driven**, from ⓐ to ⓑ with the emerging of several new topics.

The **second shift** is **technique-driven**, from ⓑ to ⓒ with the spread of deep learning.

Conclusions

- Galex: a hierarchical and integrated visual analysis system
- Reasonable visual design and rich interactive components
  - spotlight, synchronous spotlights
  - contour line
  - topic trees
  - time brush/time slice creator, small snapshots, phrase rankings
- Knowledge discoveries of a discipline
  - check its evolution and intersection
  - compare research contents among institutions
  - detect communities of multiple entities.

Thanks

Galex

Exploring the Evolution and Intersection of Disciplines

Zeyu Li, Changhong Zhang, Shichao Jia and Jiawan Zhang
Visual Computing Lab, Tianjin University
lzytianda@tju.edu.cn