The Structure of Science Linkage
- A Case Study of Solar Cell

The University of Tokyo, School of Engineering
Innovation Research Policy Center
Assistant Prof.
Naoki Shibata

2009/6/22

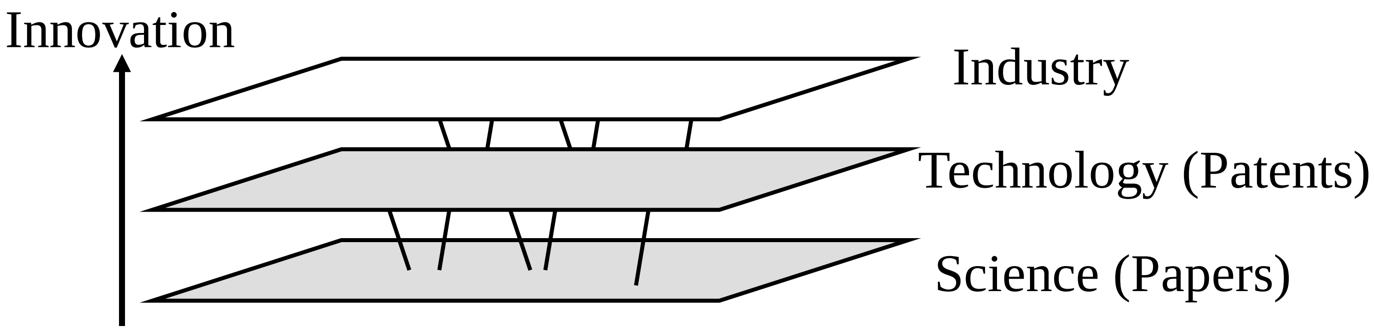
Index

1. Introduction
2. Method
3. Result
4. Discussion
5. Conclusion

1. Introduction

What's "Science Linkage"?

Model of Technology Innovation

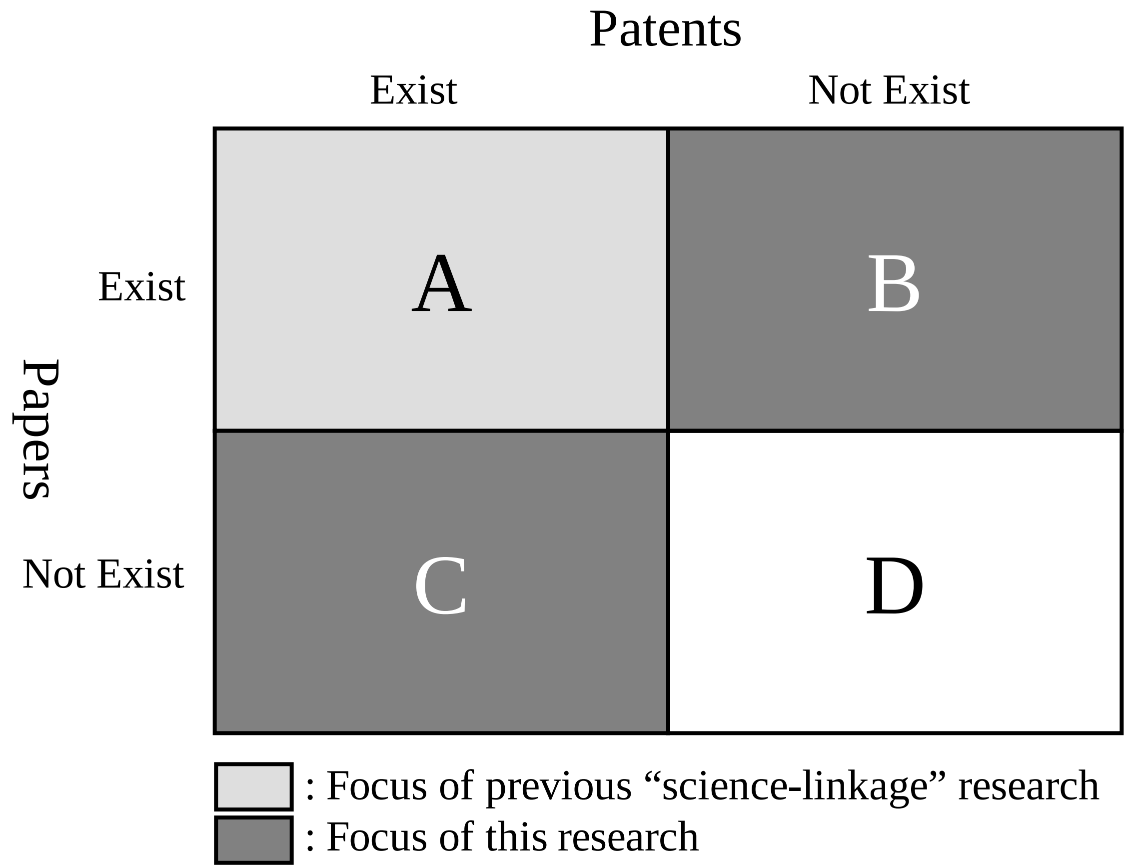


Definition of Science Linkage : citations from patent to paper

- Significant fact that "Science pushes Technology"
  - This is the main reason why governments support academic research.

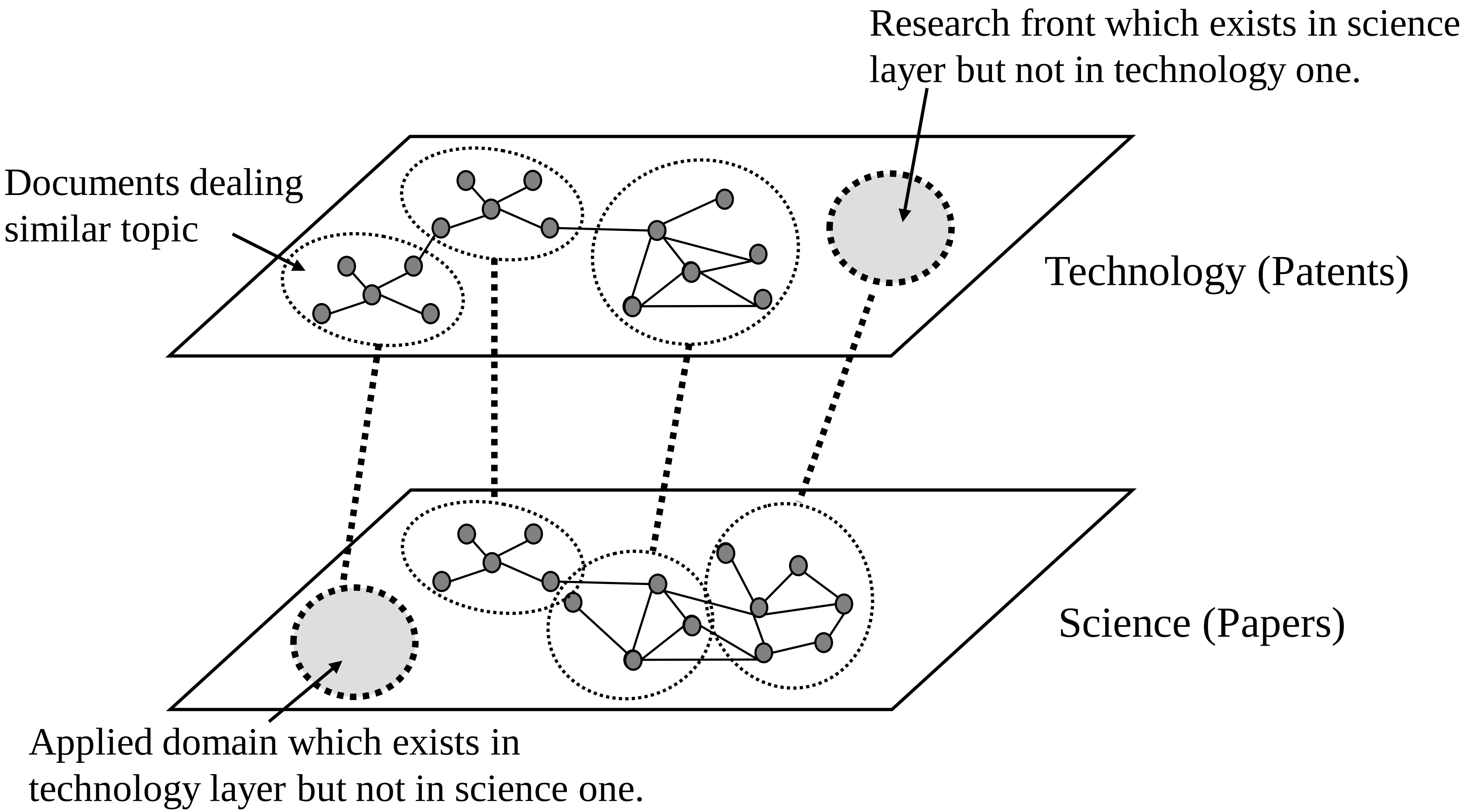
Relationship between Science and technology

- Relationship between Science and technology
  - A: Patents cite and refer related papers. (Previous “science-linkage” research.)
  - B: “Chances for industry” OR “Too basic research”
  - C: “Too applied” OR “lack of science”



Goal: a methodology to detect the gap

Strategy = Gap between Science and Technology

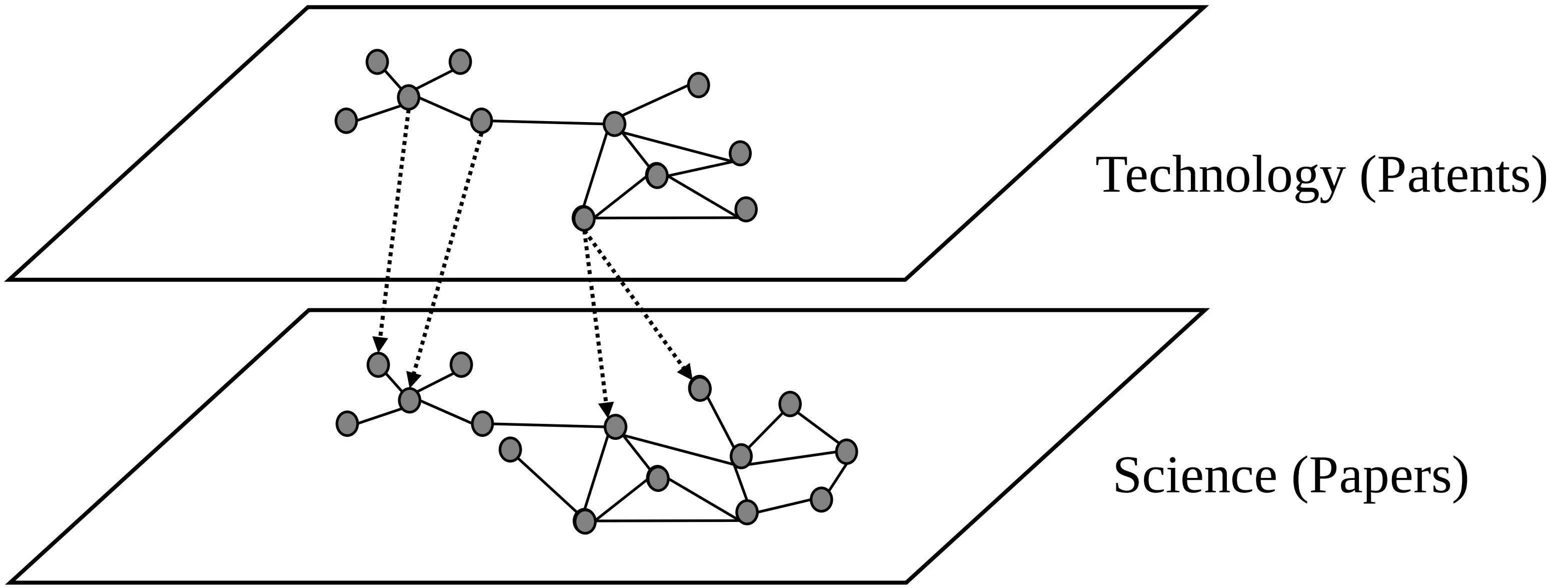


2. Method

How to detect gaps between technology and science? (1)

Previous research:

- analyzing visible citations (science linkages) from patents to papers.



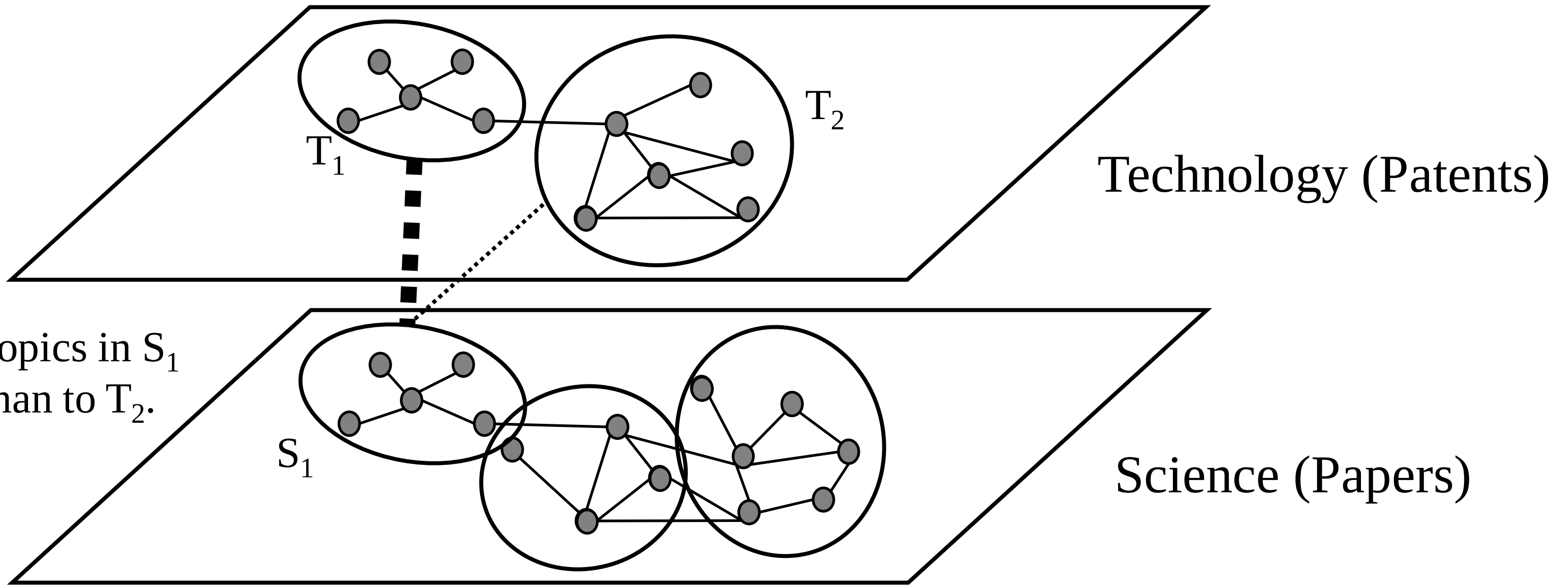
Pros	Easy to analyze.
Cons	Citation from patents to papers are sometimes inadequate. - Patents don’t always cite adequate papers. (Examiners could not understand all scientific achievements.) - Difficult to compare the meanings between patents and papers. Difficult to detect white space, which exists in one but not in the other.

How to detect gaps between technology and science? (2)

Our approach:

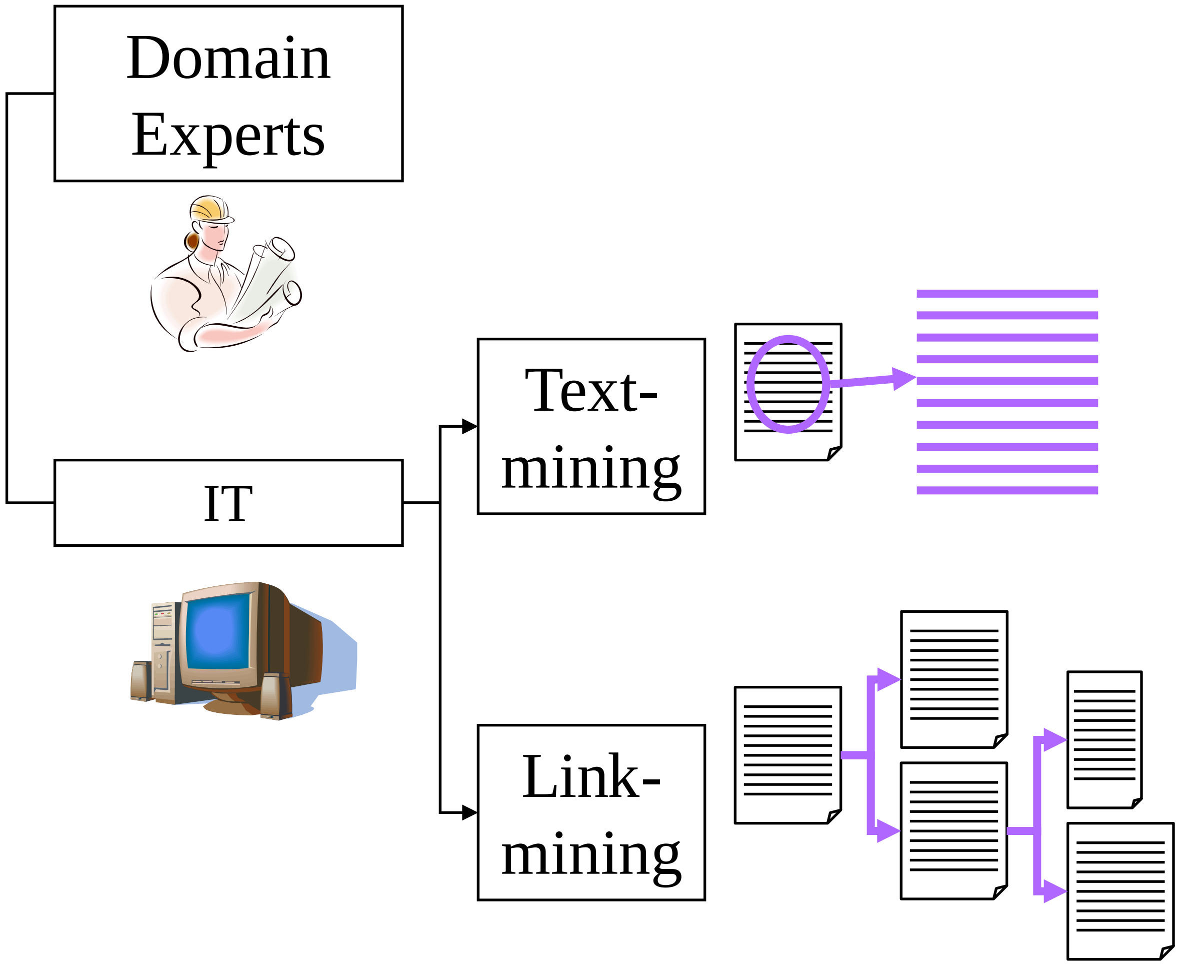
- Papers and patents are divided into clusters as closely-related documents are in same cluster.
- Comparing the topics between sets of patents and ones of papers.

(For instance,) Topics in S_1 is similar to T_1 than to T_2.

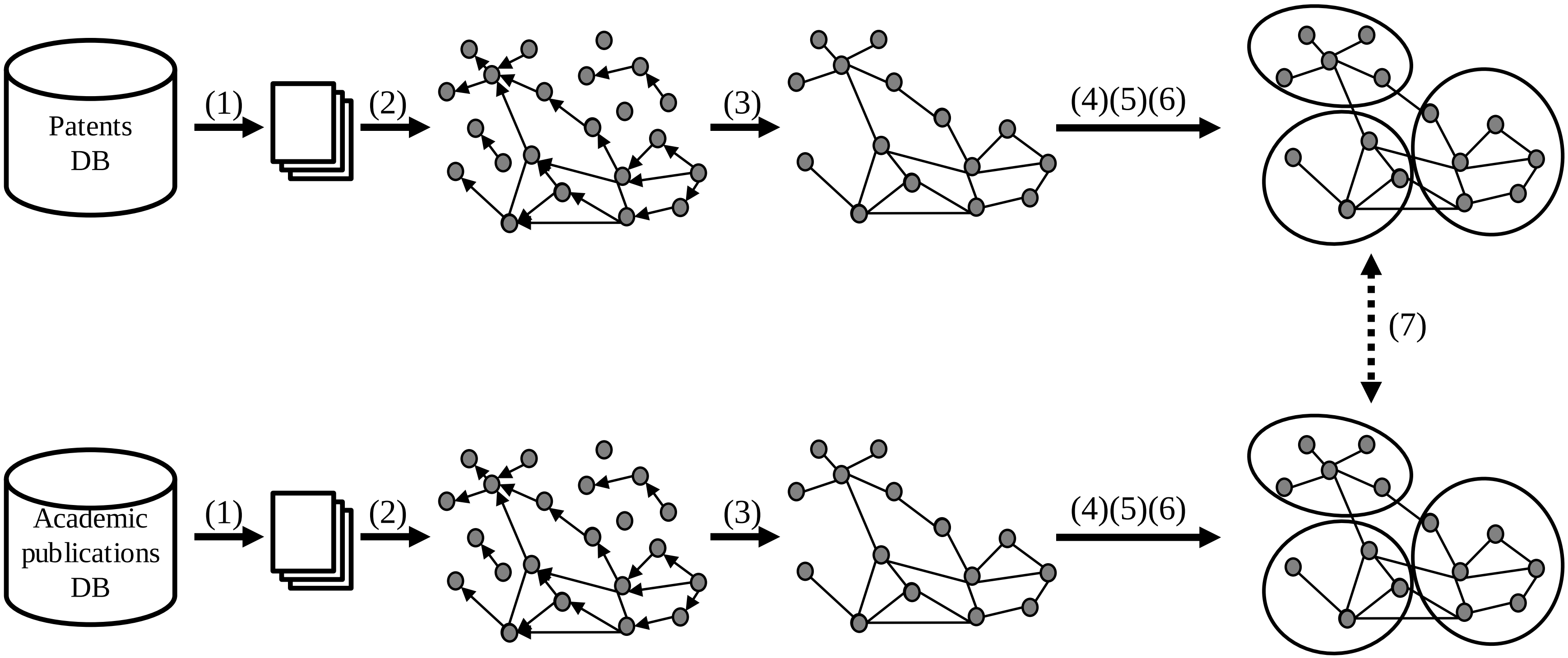


Pros	Possible to detect white space, which exists in one but not in the other. Papers and patents could be dealt as clusters of similar documents. Topics between a set of papers and one of patents are comparable semantically.
Cons	Papers/patents without citations could not be included.

How to summarize documents (in both science and technology layer)?



Method

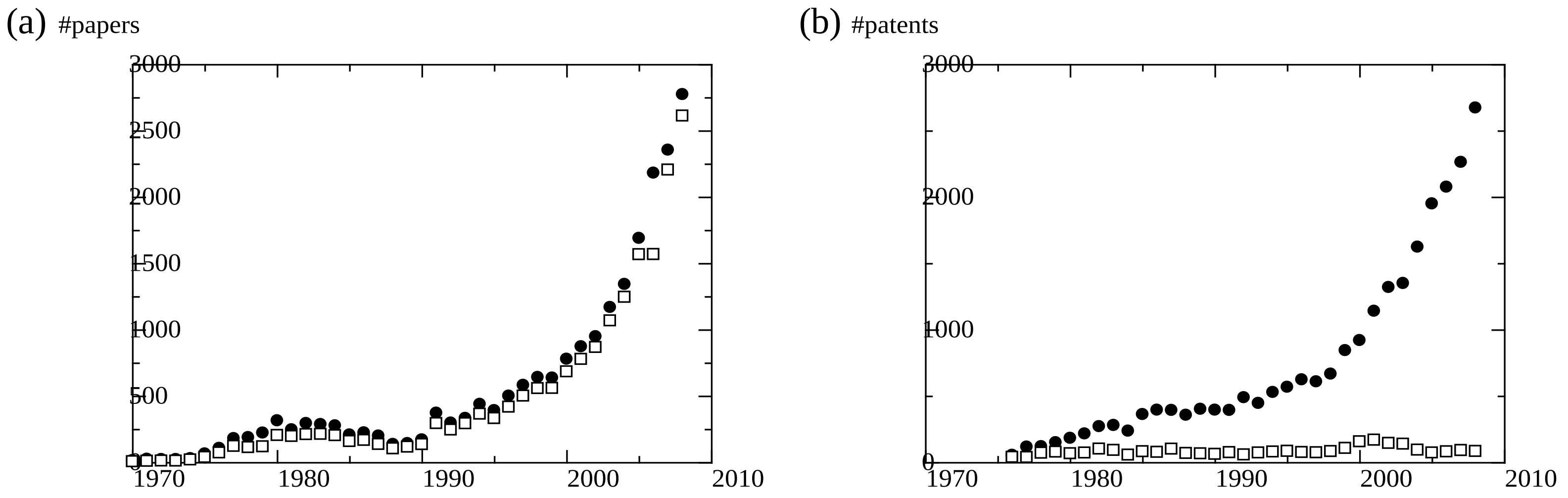


(1) Search by given query
(2) Creation of citation network
(3) Extraction of Largest Connected Component
(4) Clustering with topological clustering method
(5) Visualization
(6) Topic extraction
(7) Comparison of topics and extraction the gaps

Only step (6) and (7) are done manually and others are automatically.

A case: Solar cell

- Sustainable energy and renewable energy have been widely accepted as a key concept for our common future (Hennicke & Fischedick, 2006).
- One of the emerging and large research domains was solar cell among Energy and Fuels categories (Kajikawa et. al, 2008).
  - A solar cell or photovoltaic cell is a device that converts solar energy into electricity by the photovoltaic effect.

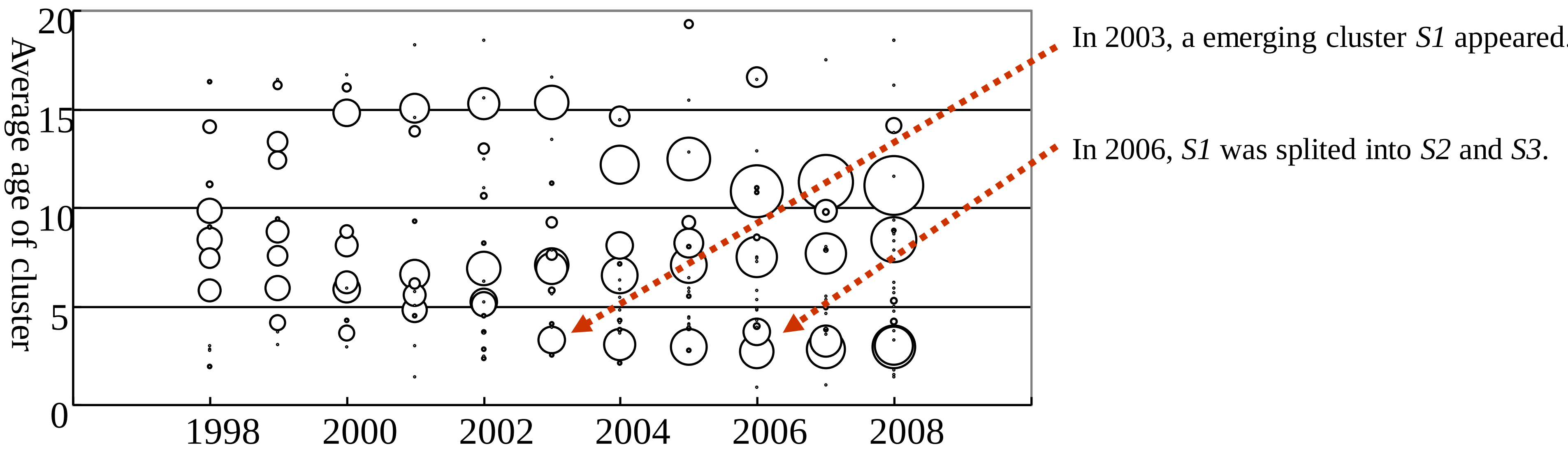


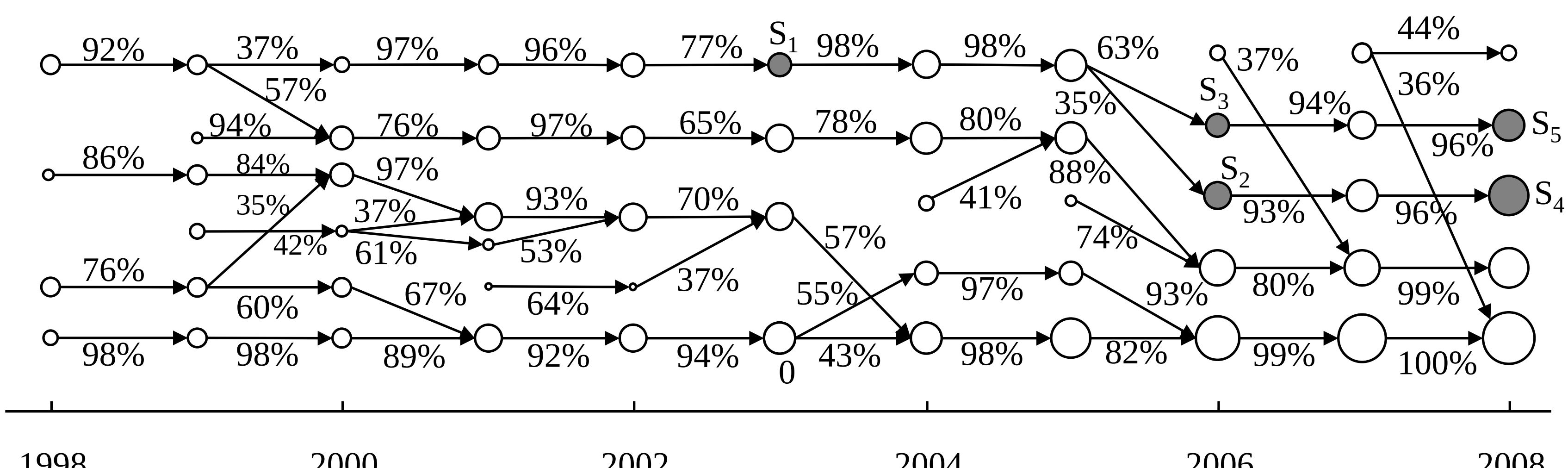
●: #documents matched by query
□: #documents in the largest component in 2008.

3. Result

Result (Science Layer) > Clustering result

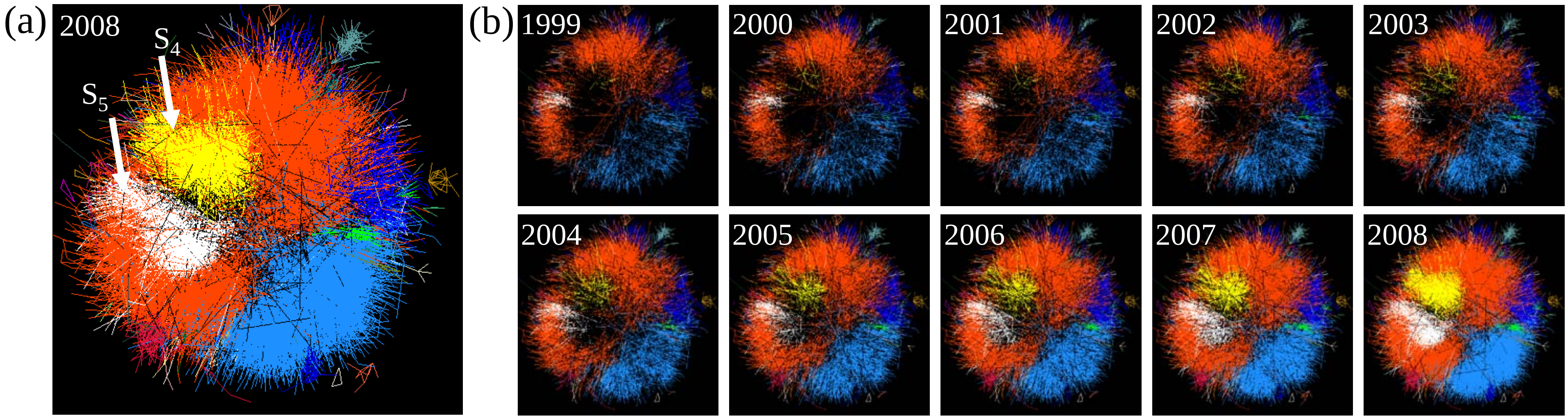
- An emerging cluster S1 appeared in 2003.
- S1 evolved and was divided into S2 and S3 at 2006 and S4 and S5 were the outcome of evolution of S2 and S3 respectively.





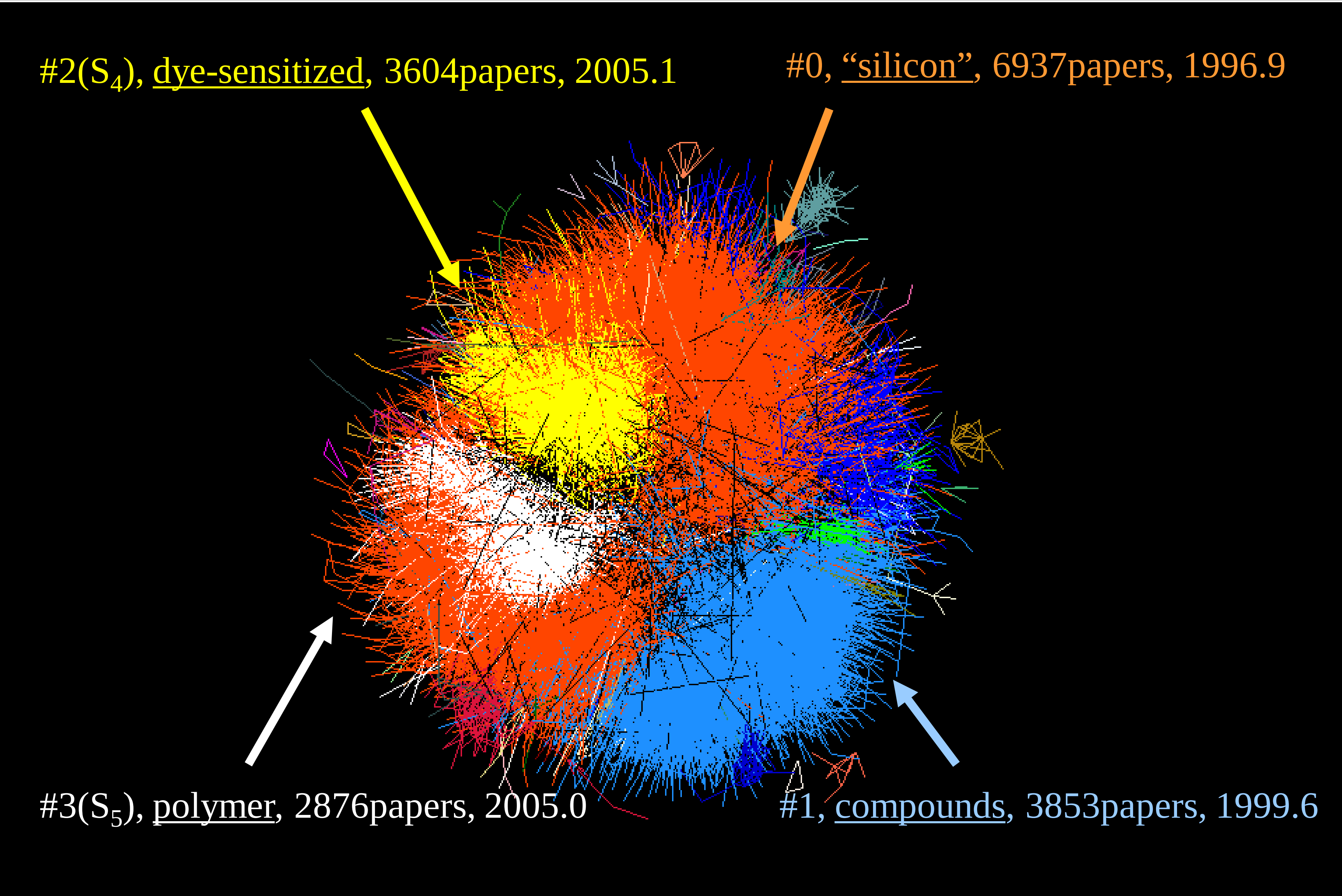
Result (Science Layer) > evolution of each cluster

- Yellow cluster (S4) and white one (S5) are emerging one.



The position of each citation is calculated by using the citation data from 1970 to final year and then fixed in other figures

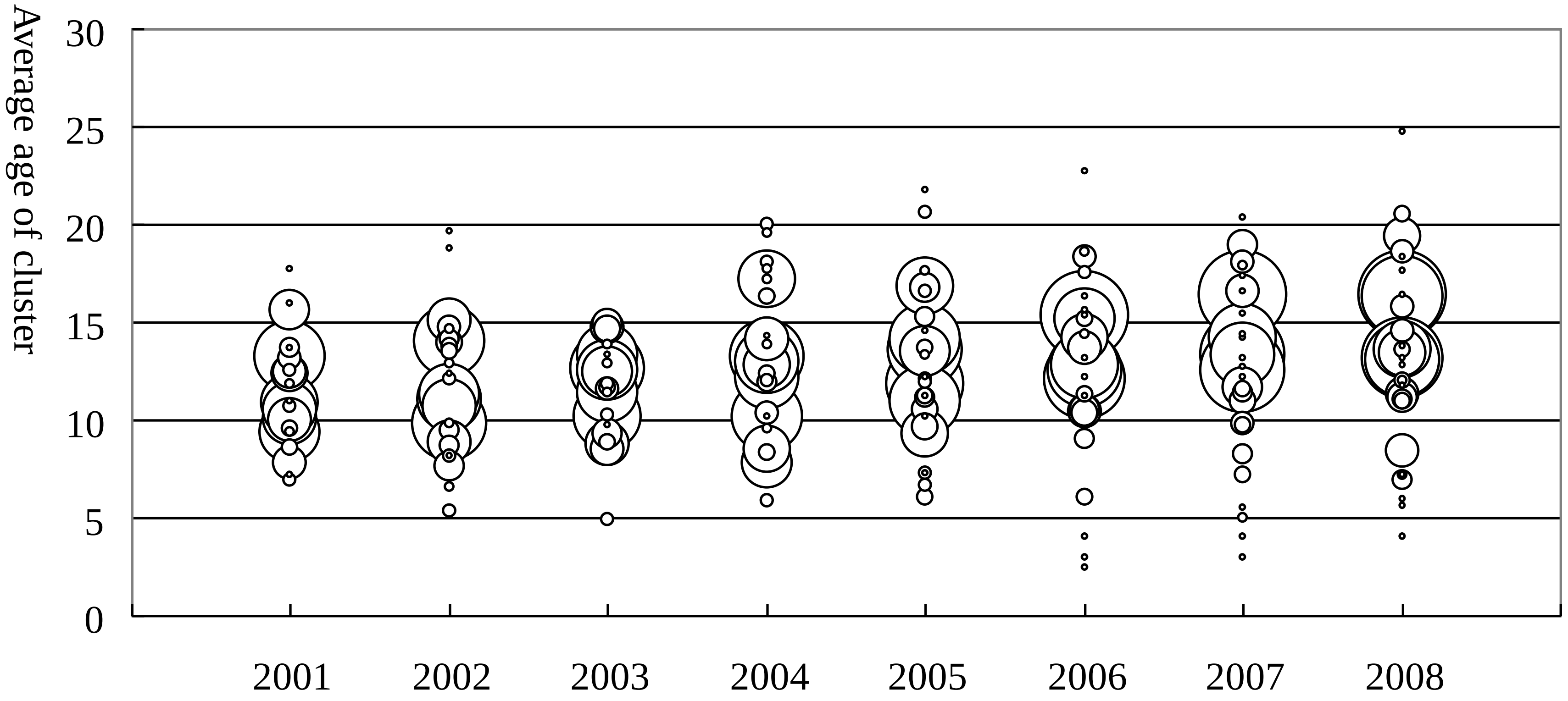
Result (Science Layer) > visualization in 2008



Only clusters whose #papers ≧ 1000 were named.

Result (Technology Layer) > Clustering result

- Patents were divided into a lot of small clusters.

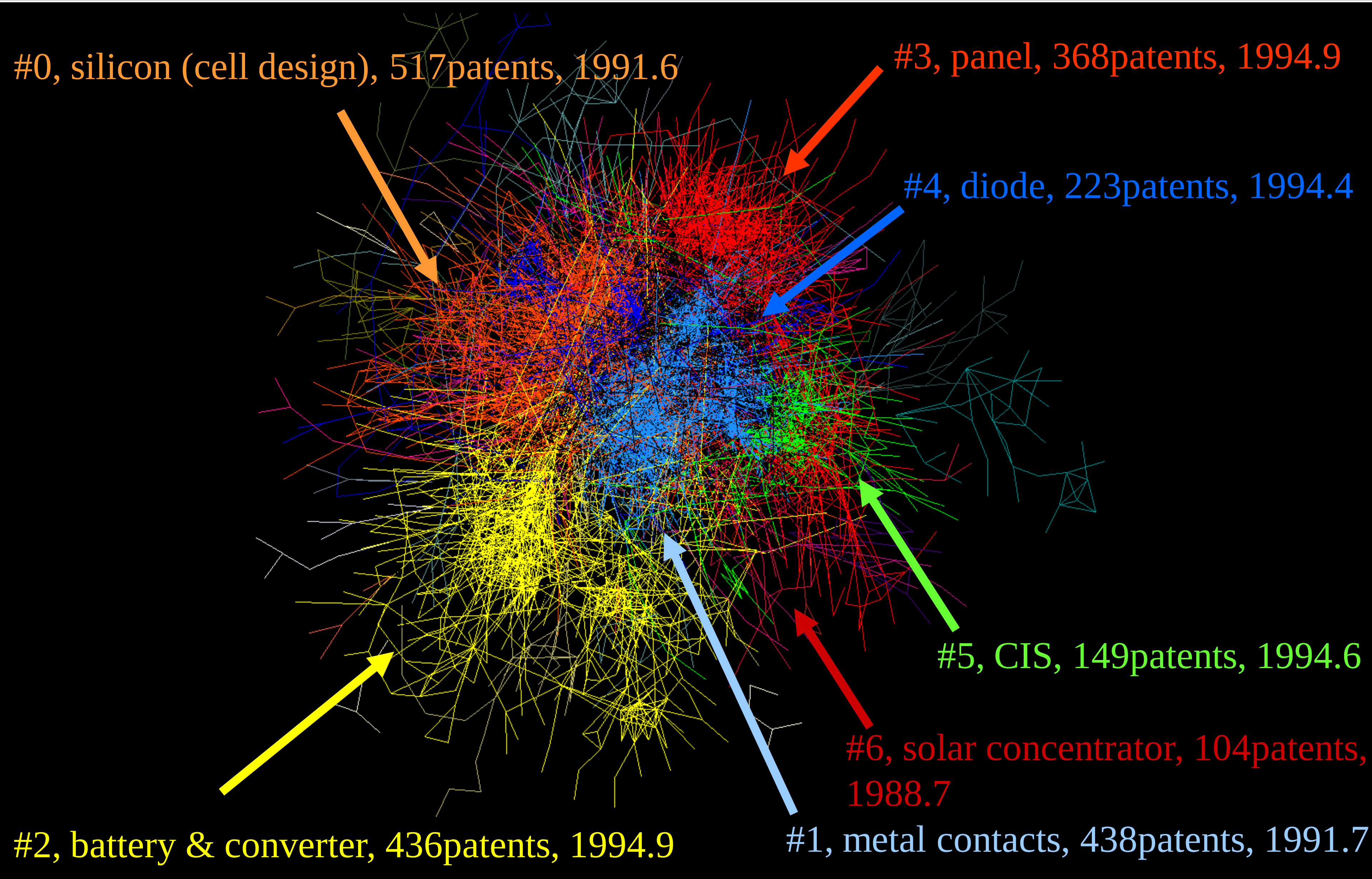


Result (Technology Layer) > Clusters & Major Applicants

id	#Patents	ave. year	topics	applicants (top5)
0	517	1991.6	silicon (cell design)	Canon Kabushiki Kaisha(51), Siemens Aktiengesellschaft(33), RCA Corporation(18), Sharp Kabushiki Kaisha(17), Atlantic Richfield Company(11),
1	438	1991.7	metal contacts	Mobil Solar Energy Corporation(24), The Boeing Company(24), Massachusetts Institute of Technology(17), Evergreen Solar, Inc.(13), The United States of America as represented by the Administrator of the National Aeronautics and Space Administration(12),
2	436	1994.9	battery & converter	Canon Kabushiki Kaisha(49), Sanyo Electric Co., Ltd.(11), Sharp Kabushiki Kaisha(8), The Boeing Company(5), RCA Corporation(5),
3	368	1994.9	panel	Canon Kabushiki Kaisha(94), Hughes Electronics Corporation(10), TRW Inc.(10), RCA Corporation(9), Sanyo Electric Co., Ltd.(8),
4	223	1994.4	diode	Sharp Kabushiki Kaisha(13), Texas Instruments Incorporated(12), Mitsubishi Denki Kabushiki Kaisha(11), Emcore Corporation(7), The United States of America as represented by the United States Department of Energy(6),
5	149	1994.6	CIS	Matsushita Electric Industrial Co., Ltd.(15), Midwest Research Institute(6), Atlantic Richfield Company(6), The Boeing Company(5), Siemens Aktiengesellschaft(5),
6	104	1988.7	solar concentrator	RCA Corporation(5), Hughes Aircraft Company(4), The United States of America as represented by the Administrator of the National Aeronautics and Space Administration(3), Hughes Electronics Corporation(2), Mobil Tyco Solar Energy Corporation(2),

Only clusters whose #patents ≧ 100 were named.

Result (Technology Layer) > visualization in 2008

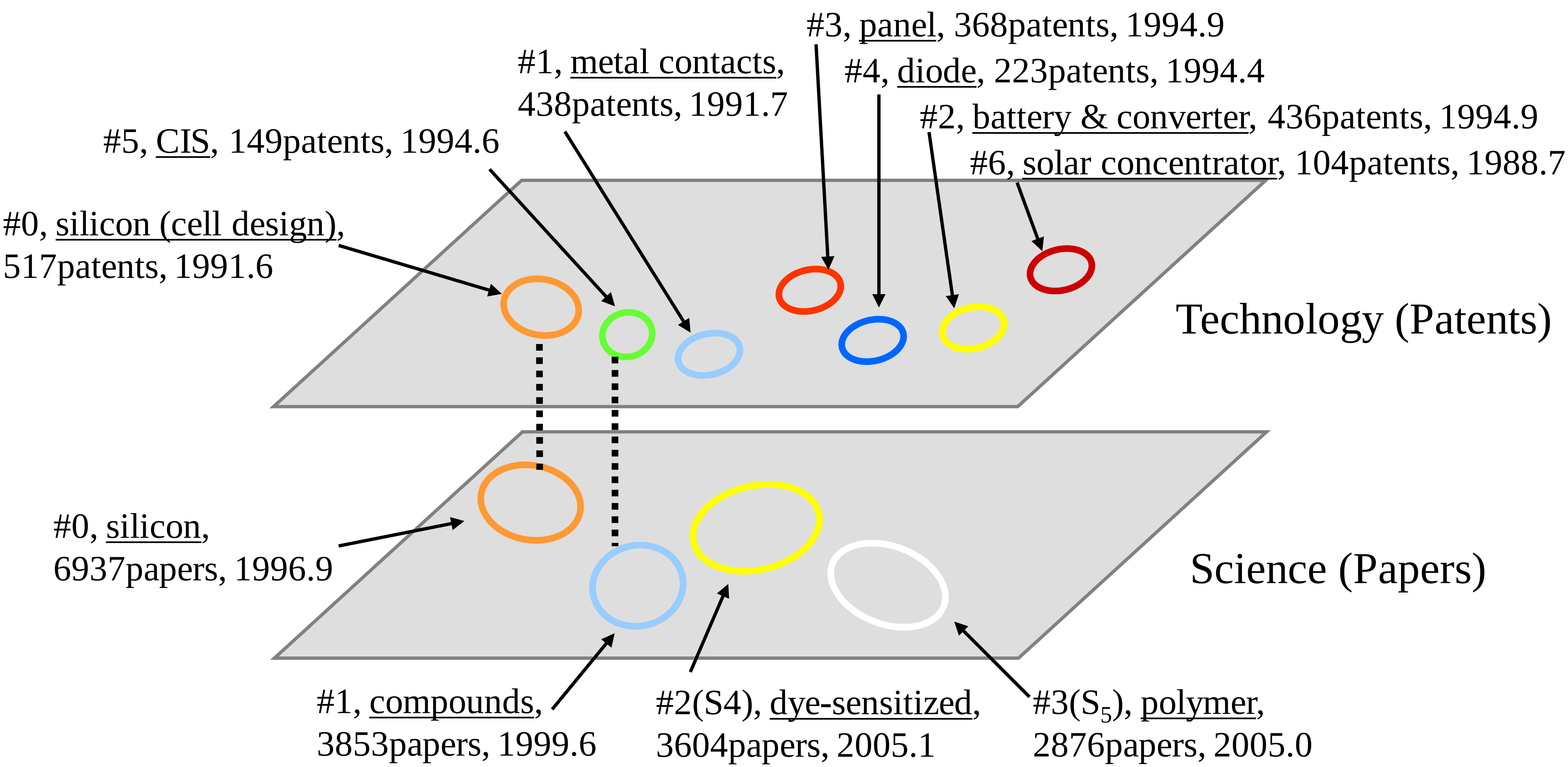


Only clusters whose #patents ≧ 100 were named.

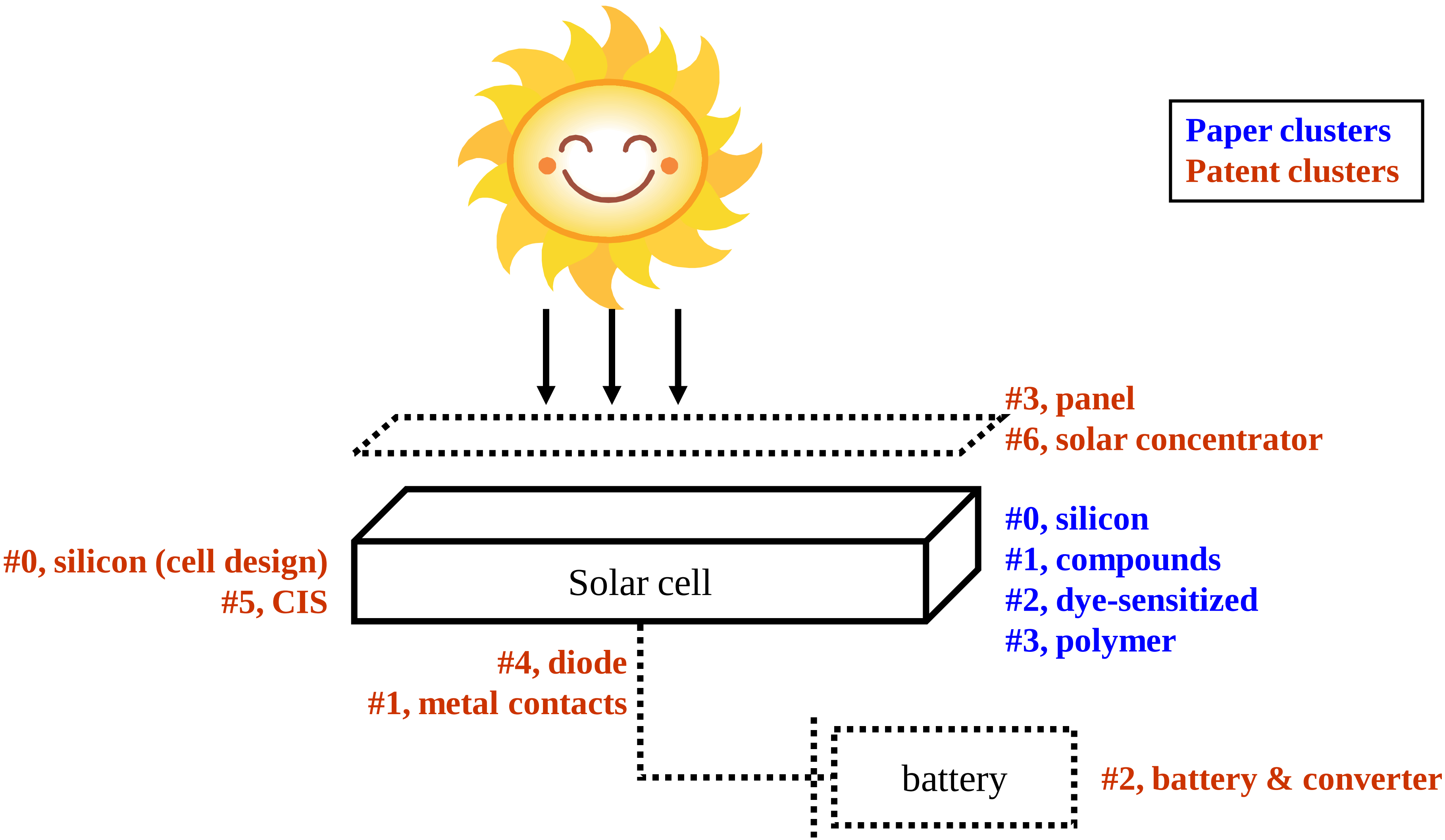
4. Discussion

Discussion

- There were mainly four(4) and seven(7) clusters in science and technology layer respectively at 2008.
- There seemed significant chances for dye-sensitized and polymer related patents.
- Applied technologies existed in technology layer but not in science layer.



Gap between Science and Technology in Solar Cell



Chance discovery

Photovoltaics (PV) Roadmap toward 2030 (PV2030) by NEDO in 2004.

- In PV2030, the target is the crystalline Si solar cell, which has the highest market share of PV, and also the high-efficiency GaAs-based or CuInSe2 (CIS) solar cell.
- Dye-sensitized cells were the subject of discussion as seed research after 2010.

Comparison and remarks

- In the case of solar cell, as written in the roadmap, there seems to exist the gap between science and technology, as dye-sensitized and polymer related solar cell.
- Our approach could, at least, provide a new viewpoint to develop a strategy of MOT to decide which domain to focus and invest.

5. Conclusion

Conclusion

- Analysis of the gap between science and technology enables both R&D managers and policy makers to develop a strategy of MOT.

- In this research, we extracted the gaps by analyzing the structure of citation networks of both science (academic publications) and patents.
  - The citation networks of papers and patents were divided into clusters and topics were assigned to each cluster manually.
  - The similarity of topics between a papers' cluster and patents' cluster was compared.

- Our results matched the roadmap and was confirmed as reasonable by experts to work as a tool for MOT.

Thank you for listening!

Appendix

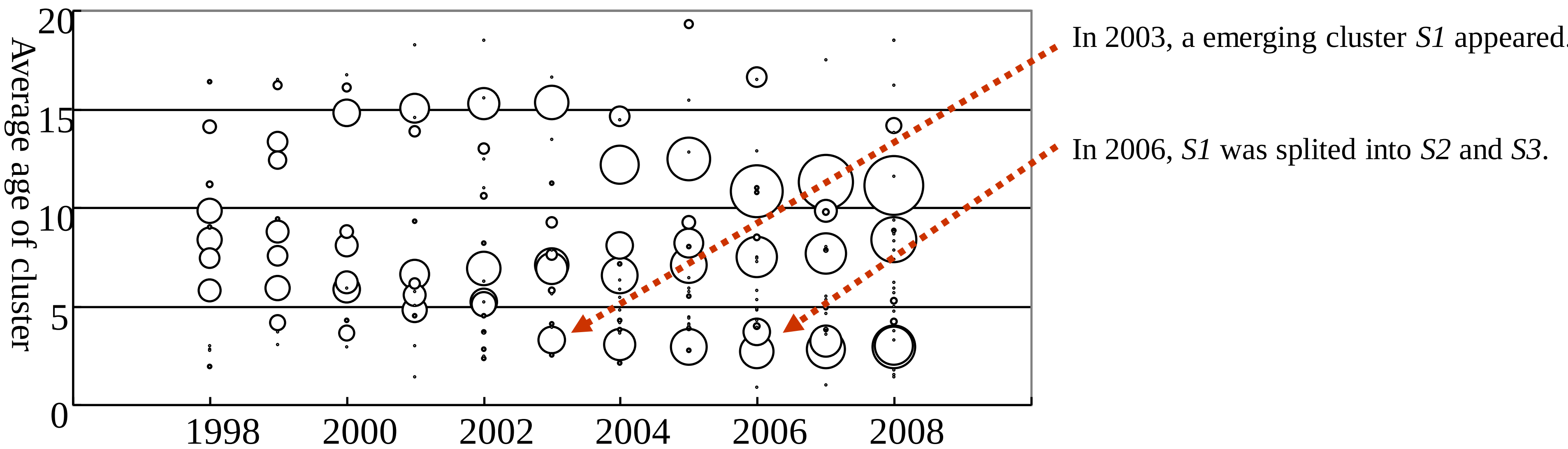
To answer a question "It might be too late to find a chance in 2009":

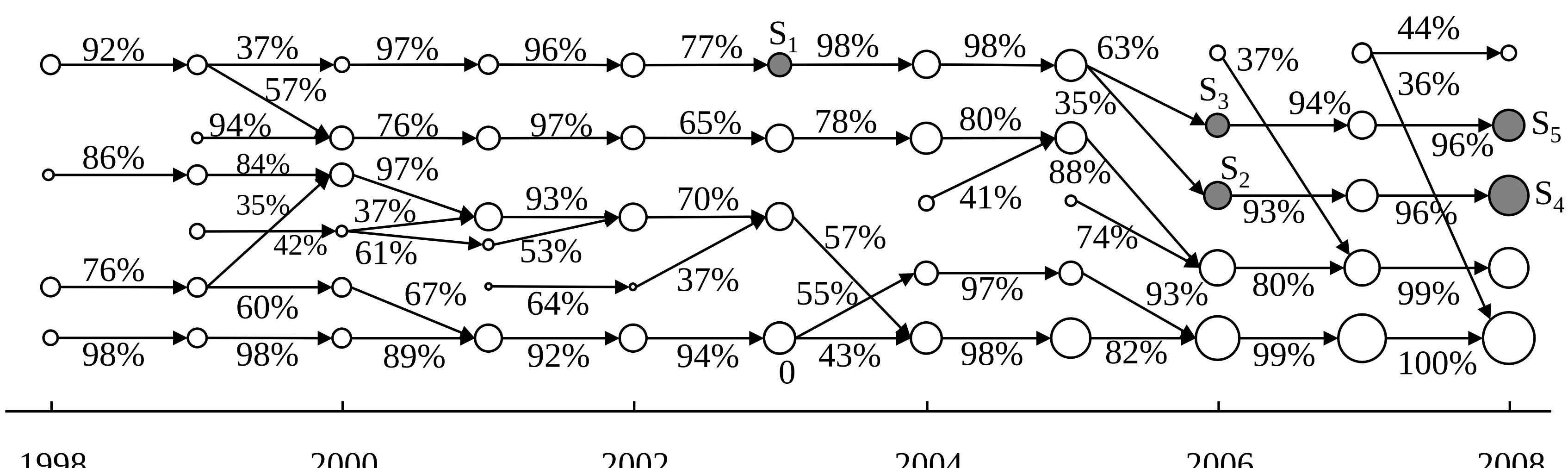
Historical Analysis

- In science layer,
  - "Dye-sensitized" cluster emerged in 2003.
  - "Polymer" one did in 2006.

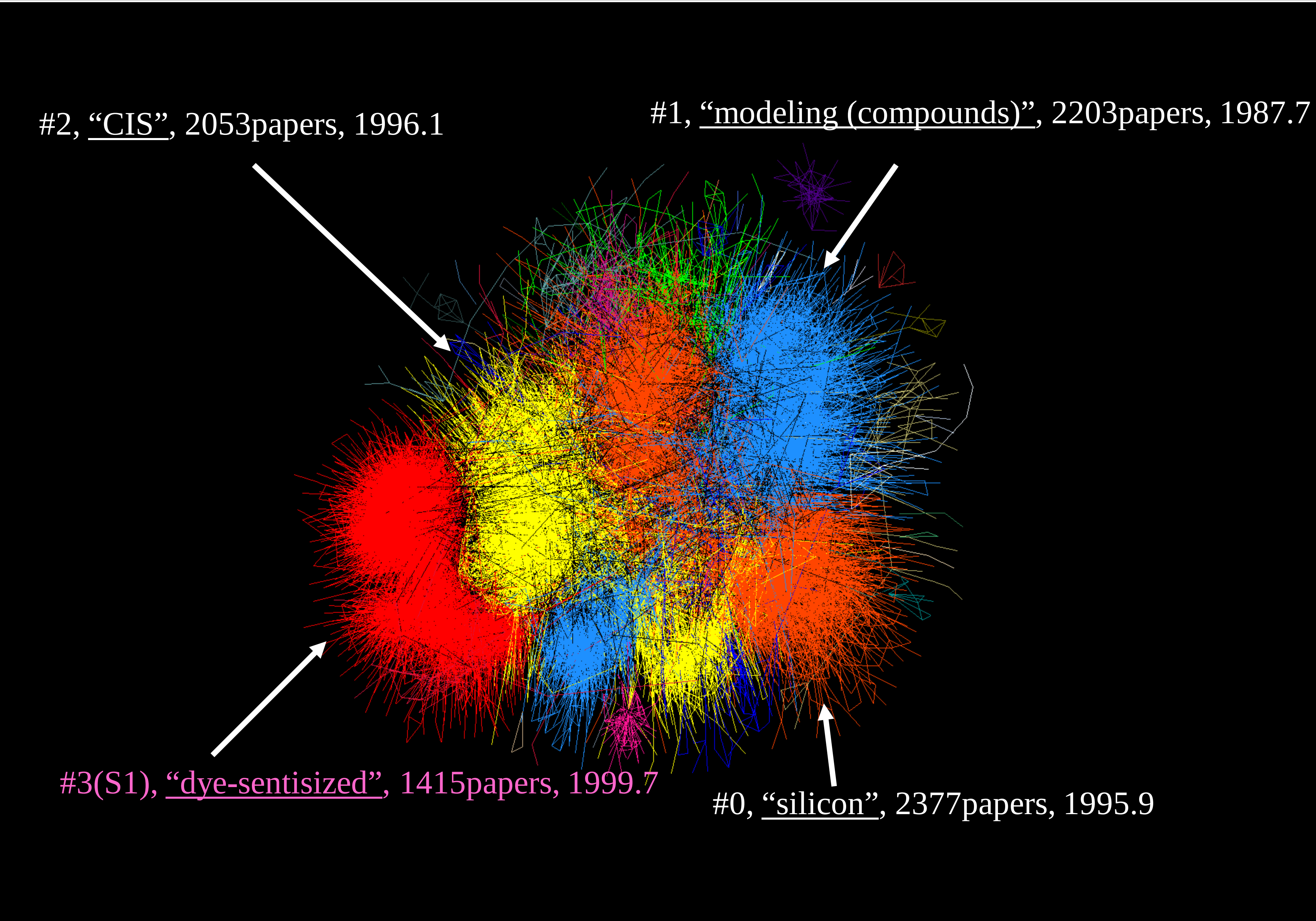
Result (Science Layer) > Clustering result

- An emerging cluster S1 appeared in 2003.
- S1 evolved and was divided into S2 and S3 at 2006 and S4 and S5 were the outcome of evolution of S2 and S3 respectively.

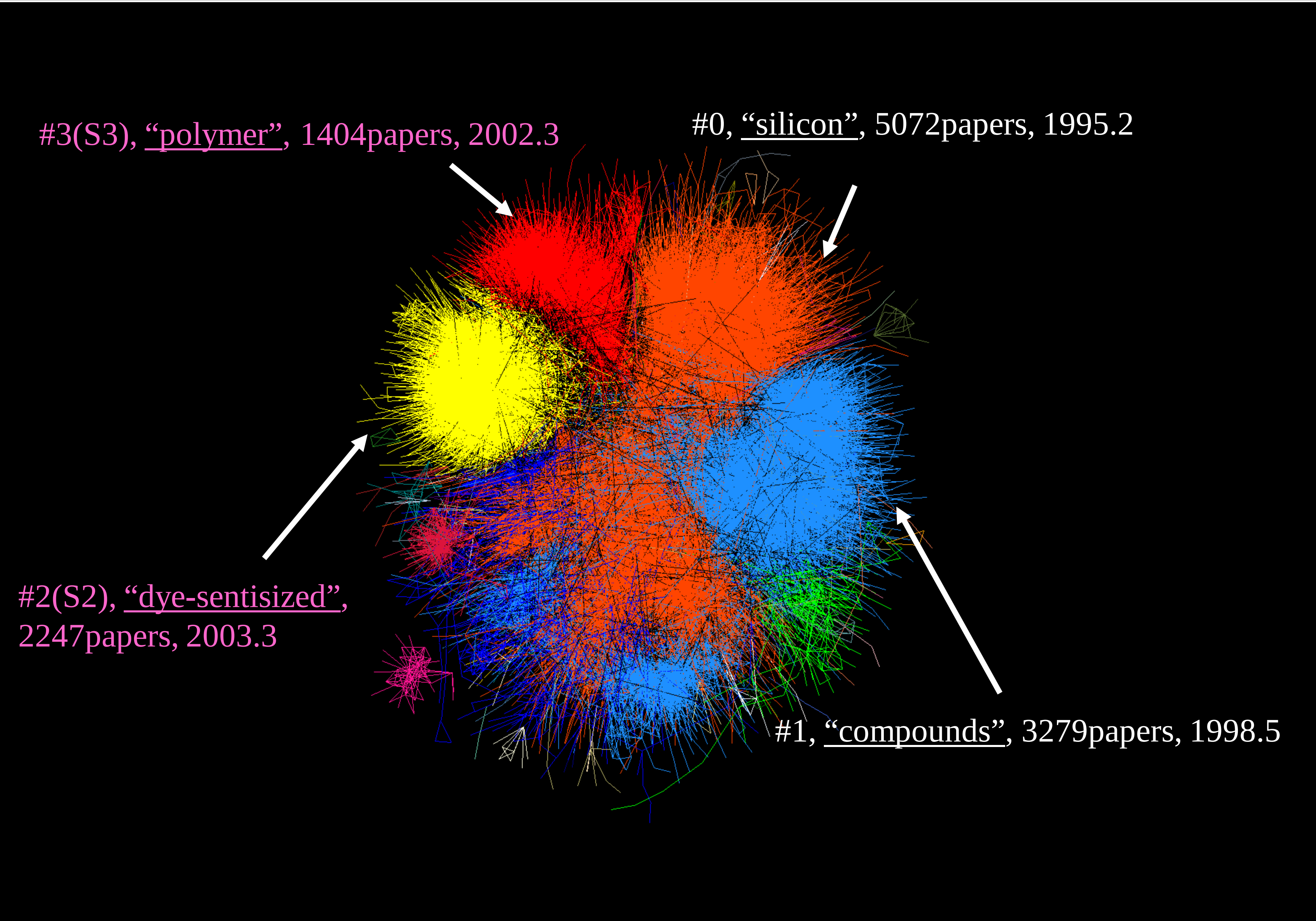




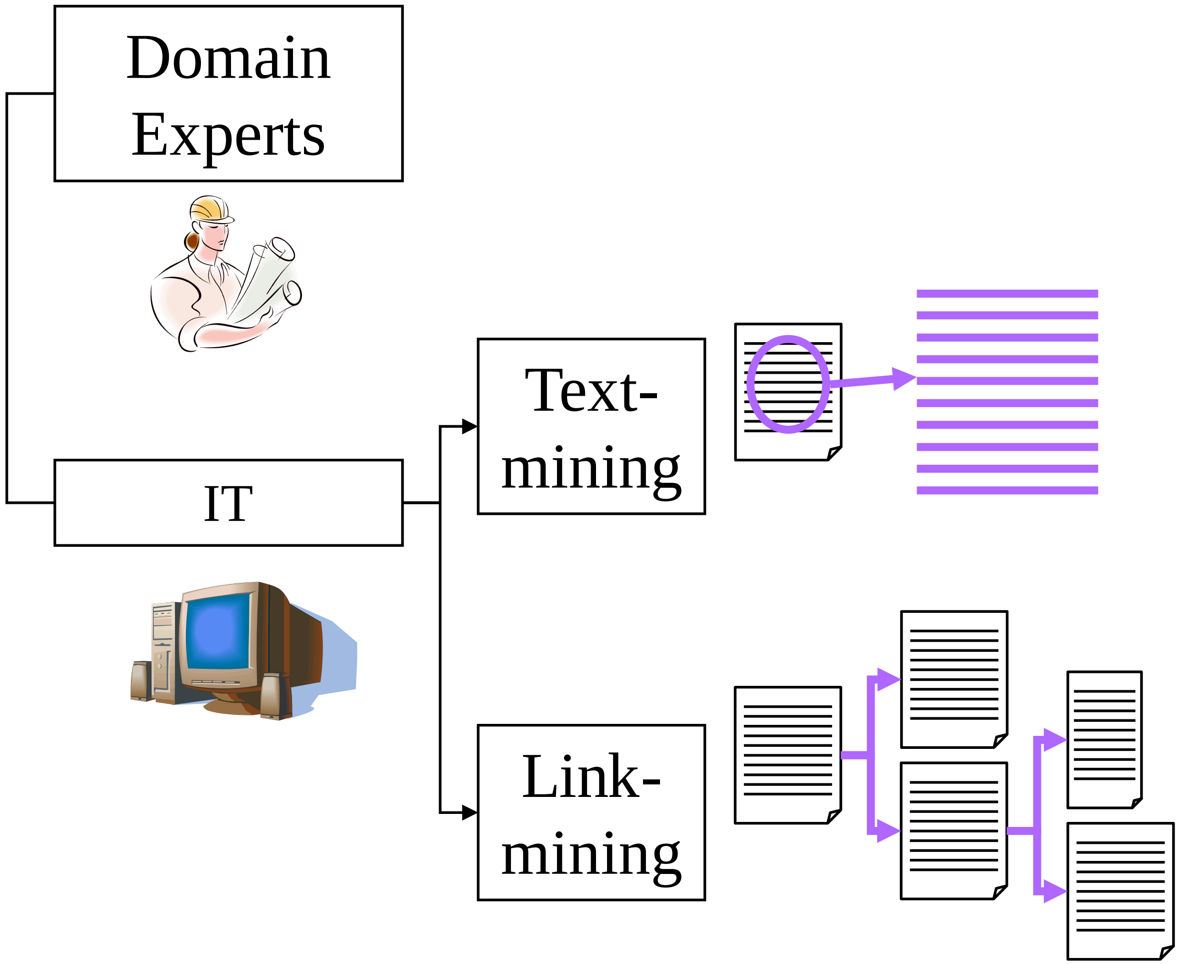
Results > visualization in 2003



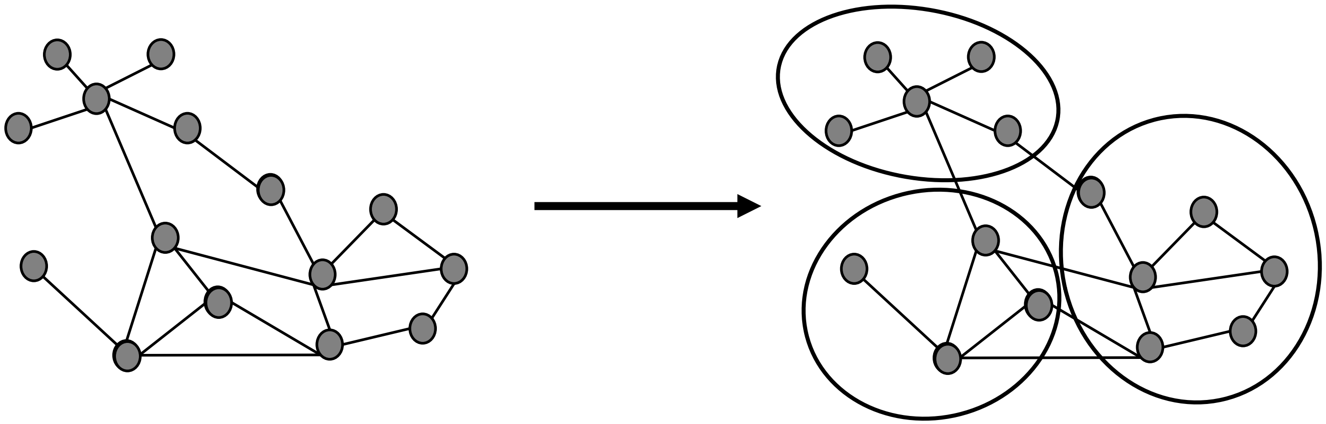
Results > visualization in 2006



How to summarize documents (in both science and technology layer)?



With topological clustering, papers could be divided into clusters.



Papers are divided into clusters as more intra-cluster edges and less inter-cluster ones.

Modularity Q

$$Q = \sum \left(e_{ii} - a_i^{\,2} \right)$$

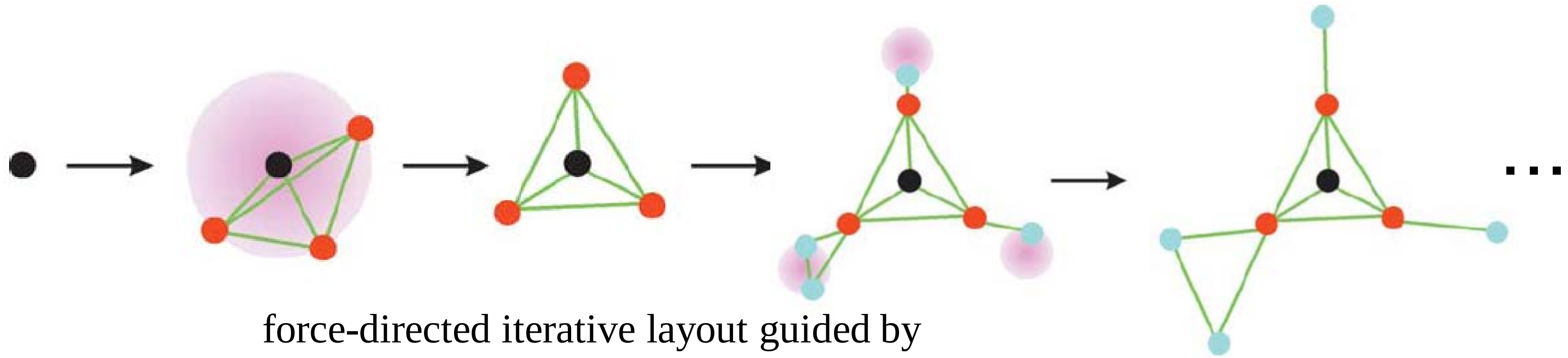
- e_{ij} : the fraction of the edges in the network that connect nodes in cluster i to those in cluster j.
- $a_i \equiv \sum_j e_{ij}$: the fraction of the edges in the network that connect nodes in cluster i.

Algorithm: maximizing Q

1. Each node is regarded as a cluster.
2. Merging cluster with following rules:
   - The pair of cluster i and j which increase ΔQ most is selected.
   - Cluster i and j are merged.
   - Continue until ΔQ becomes minus.

Newman, M. E. J. (2004). Physical Review E, 69, 066133.
Clauset, A., Newman, M. E. J. & Moore, C. (2004). Physical Review E, 70, 066111.

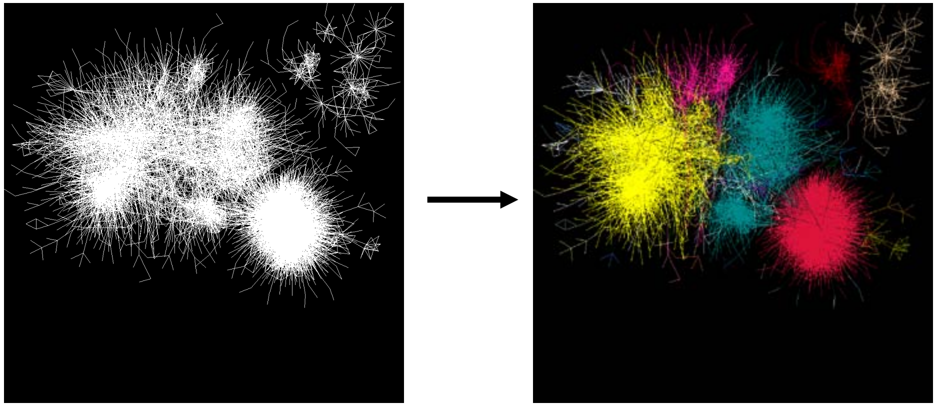
Large Graph Layout (LGL)



force-directed iterative layout guided by
a minimal spanning tree of the network

Adai et al. J. Mol. Biol. (2004).

Coloring



Visualized with intra-cluster links as the same color