Identifying Strategic Research Partners through Data-Driven Analysis of the Renewable Energy Technologies

2025. 10. 01.

Jehyun Lee[1],
Yedam Kim[2], Jihoo Jung[1], Woonho Baek[1], Jae Kyoung Yoo[1], Sangjin Choi[1, *]

[1]Korea Institute of Energy Research, Daejeon, Republic of Korea

[2]Polytech Sorbonne, Paris, France

KIER 한국에너지기술연구원 KOREA INSTITUTE OF ENERGY RESEARCH

MOST VALUABLE PROFESSIONAL Microsoft ALUMNI

Stanford University

Stanford University Libraries | Social Science Data Collection

(HCMST)

Abstract:

- This new survey, How Couples Meet and Stay Together 2017, features a fresh set of 3,510 survey respondents, with no overlap in subjects from the original HCMST survey which was first fielded in 2009. The 2017 subjects were followed up in 2020 (2,107 subjects) and in 2022 (1,722) subjects.
- HCMST 2017 featured new questions about subjects' use of phone apps like Tinder and Grindr for dating and meeting partners.
- HCMST 2020 and 2022 featured the usual HCMST relationship questions plus a lot of new questions about the impacts of the coronavirus pandemic.

Principal Investigator:

Michael J. Rosenfeld

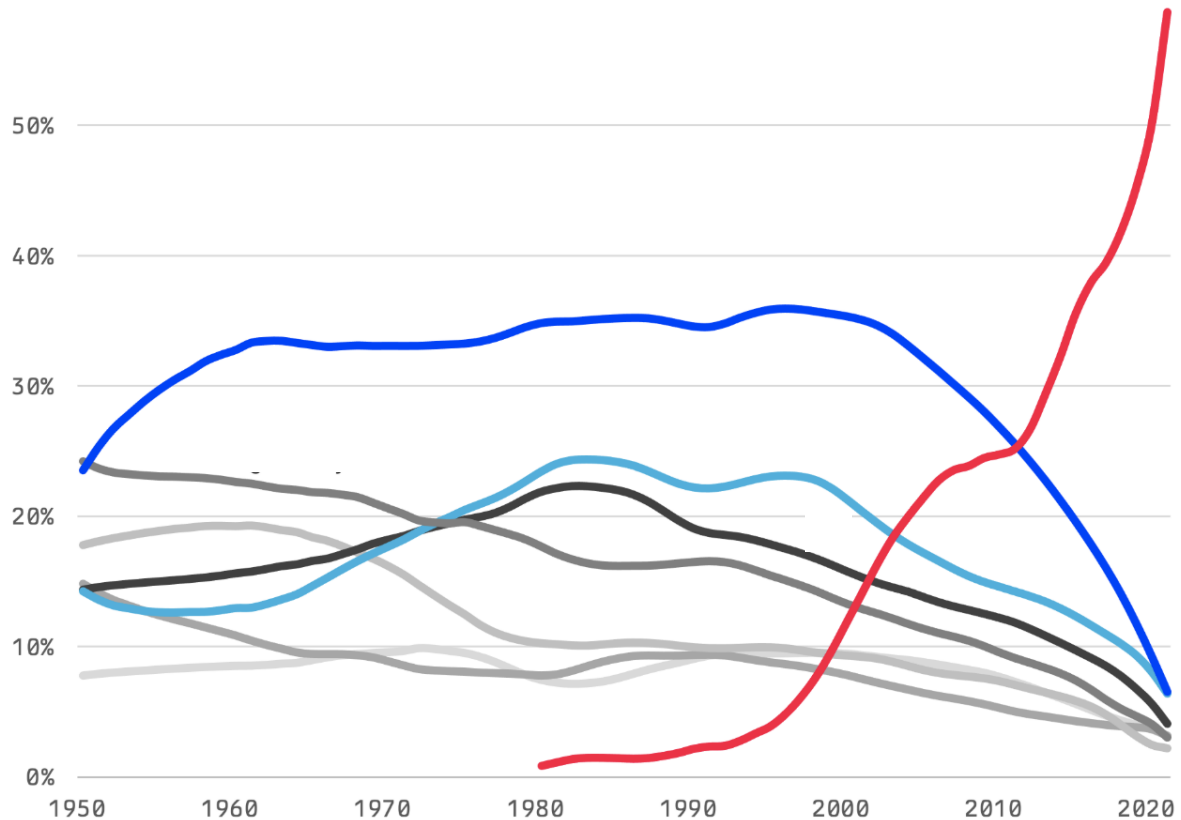
Reuben J. Thomas

Sonia Hausen

Funding Agency:

United Parcel Service Endowment at Stanford University

US National Science Foundation



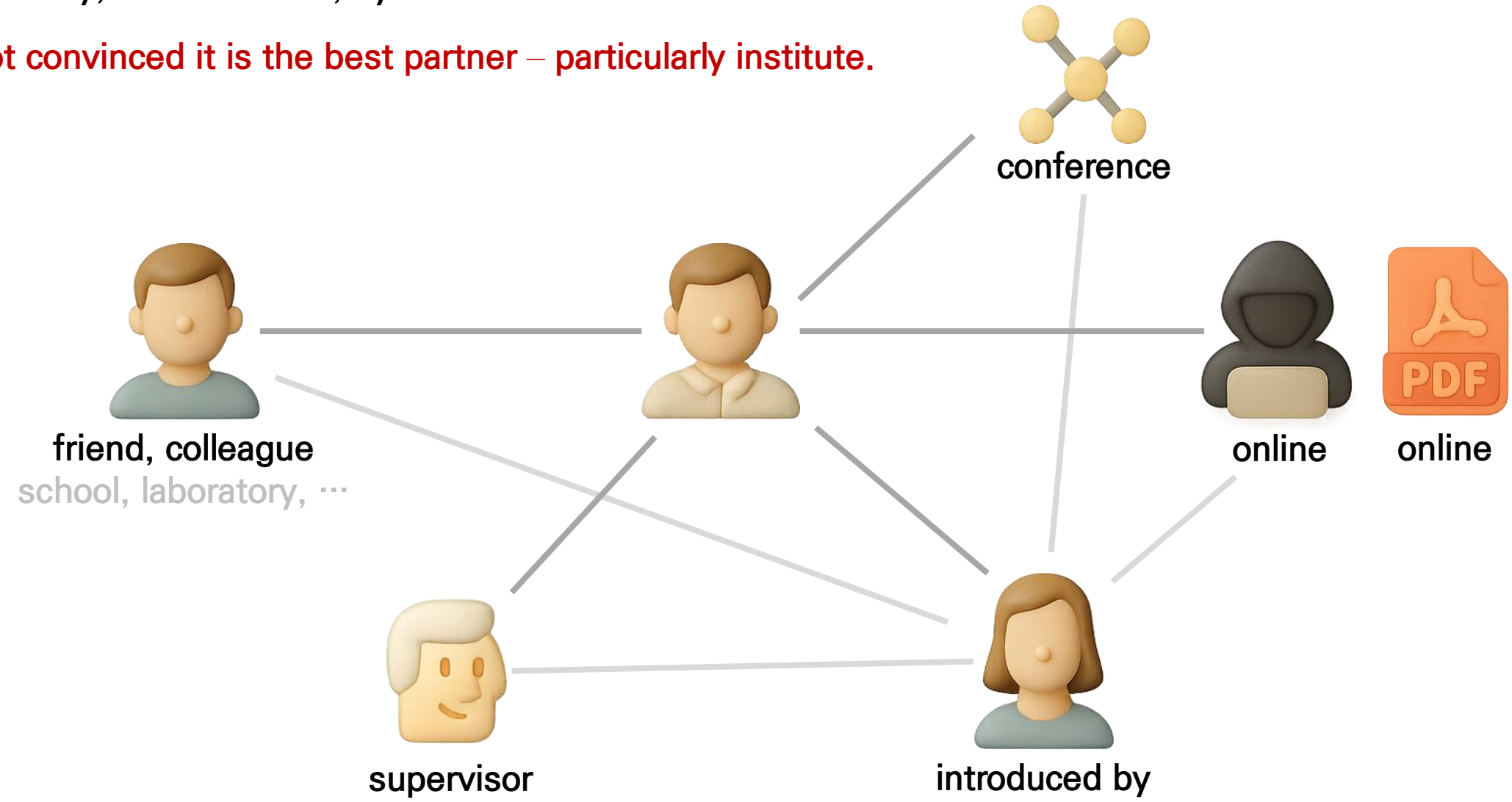
Source: "How Couples Meet and Stay Together": a longitudinal study of social life in the US by M. J. Rosenfeld, Reuben J. Thomas, and Sonia Hausen. Analysis of original survey data (n=6,519); "bars & restaurants" category cleaned to not double count couples who first met online.

MARRIAGE PACT

https://data.stanford.edu/hcmst
https://www.pnas.org/doi/full/10.1073/pnas.1908630116
https://slowrevealgraphs.com/2025/05/19/how-couples-meet-in-the-us/

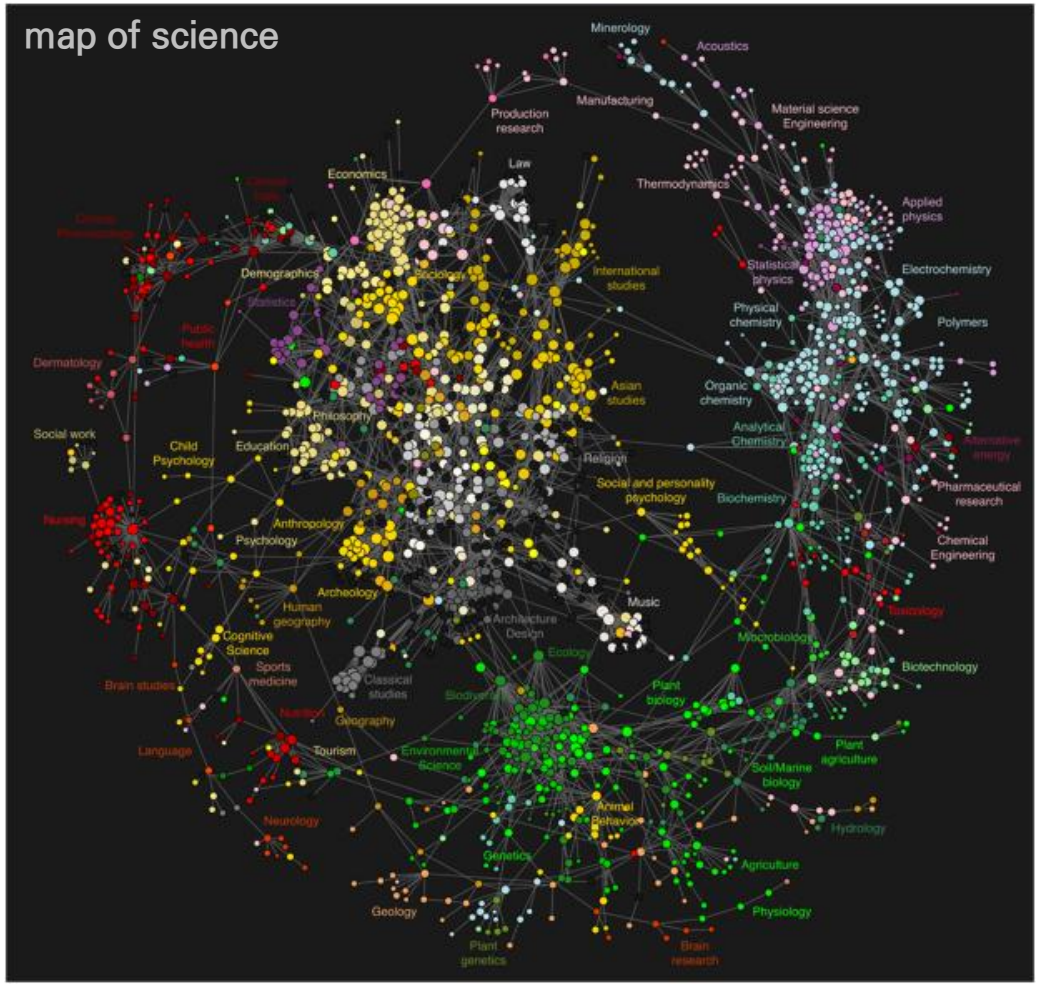
How Researchers Meet Their Collaborators?

- Basically, based on fate, by chance
- Not convinced it is the best partner – particularly institute.



Identifying Strategic Research Partners

- **Topic alignment** the degree to which collaborators share interest in the same research topic.
  - Among them, find the best one – best competitor or best collaborator
- **Repetitive job**
  - Build a system!



institute A

institute B

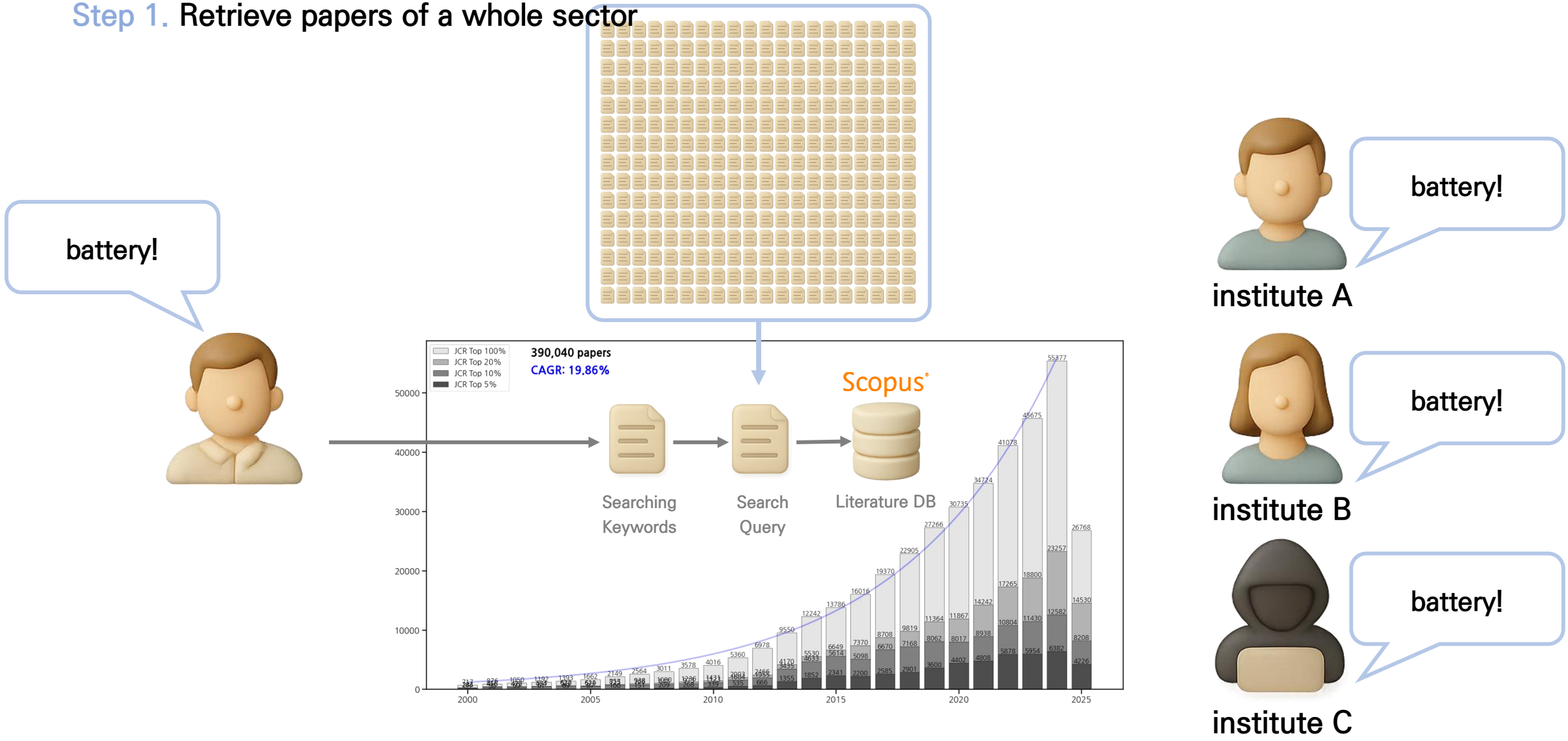
institute C

10.1371/journal.pone.0004803

2025.10.01. RD20, Jehyun LEE. jehyun.lee@gmail.com

Identifying Strategic Research Partners by Data-Driven Analysis

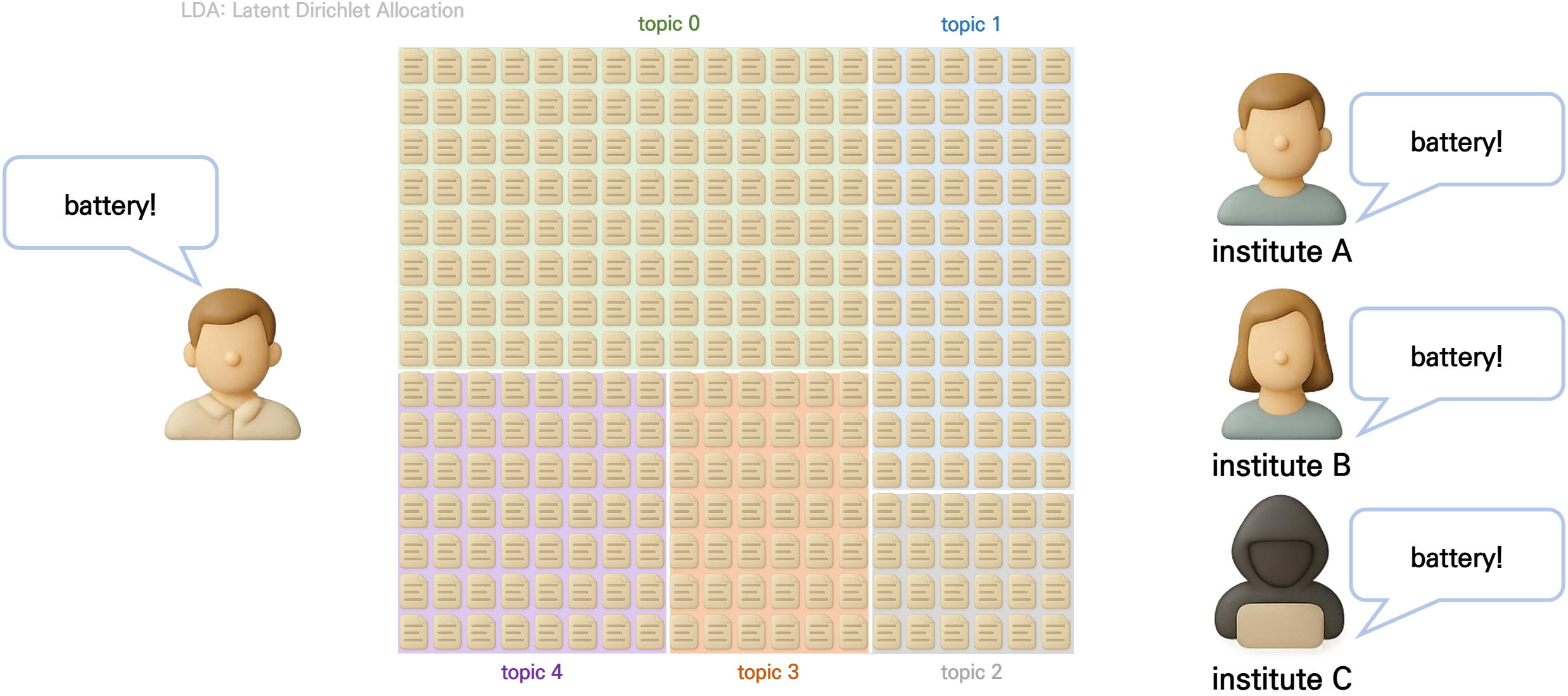
Step 1. Retrieve papers of a whole sector



Identifying Strategic Research Partners by Data-Driven Analysis

Step 2. **Clustering papers by topics** – Actually, a paper is a set of multiple topics with weights.

LDA: Latent Dirichlet Allocation



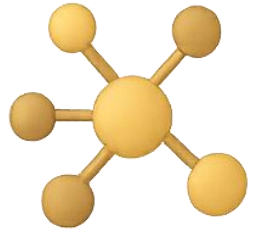
LDA Latent Dirichlet Allocation

- A journey to discover hidden topics in documents

A document = Mixture of topics

Technology 70%

Economics 30%



A topic = Mixture of words

Technology = {AI, semiconductor, data, ⋯}

Economics = {interest rate, market, investment, ⋯}

Step 1

Initialization

Randomly assign each word in every document to one of the K topics (here, K=20)

Step 2

Iterative Reassignment

For each word, reassign its topic based on two criteria, until no longer changes significantly.

1.document–topic relationship: how prevalent is this topic in this document?
2.word–topic relationship: how often is this word associated with this topic?

Step 3

Extract Result

Based on the final stable assignments, calculate the definite document–topic and topic–word distributions.

Step 4

Topic Naming

Utilizing LLM with a proper prompt, name topics with the given words list.

2025.10.01. RD20, Jehyun LEE. jehyun.lee@gmail.com

LDA Latent Dirichlet Allocation

a paper

nature reviews materials

Explore content · About the journal · Publish with us · Subscribe

nature › nature reviews materials › review articles › article

Review Article | Published: 20 March 2025

Understanding materials failure mechanisms for the optimization of lithium-ion battery recycling

Mengting Zheng, Ya You & Jun Lu

Nature Reviews Materials **10**, 355–368 (2025) | Cite this article

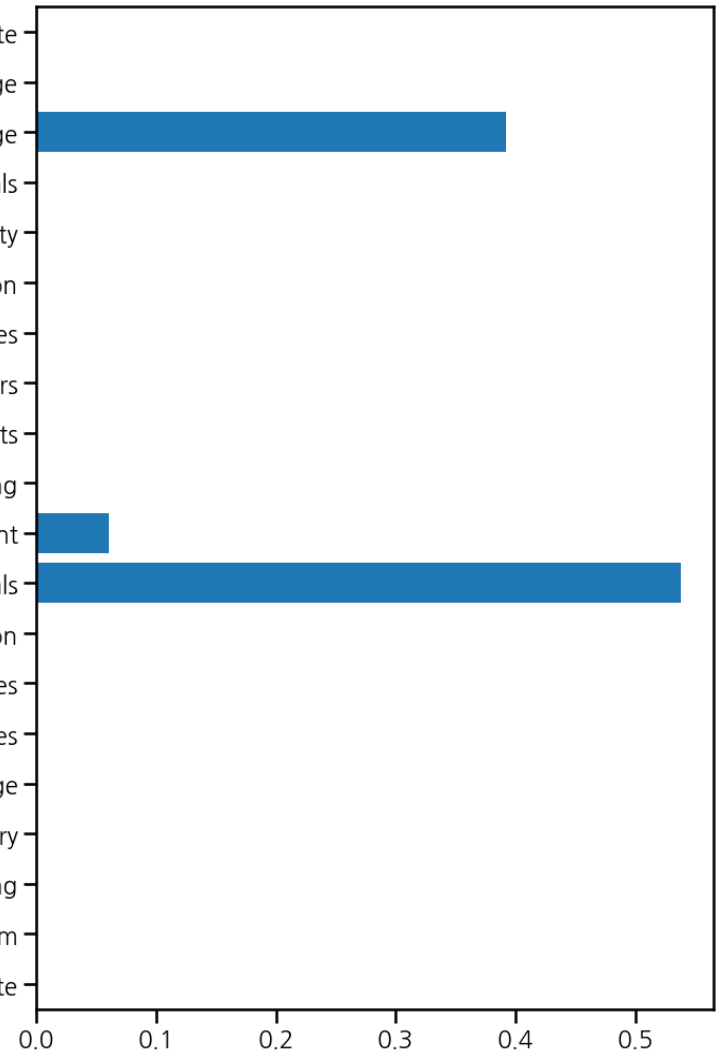
5860 Accesses | **22** Citations | **3** Altmetric | Metrics

Abstract

The pace of electrification is surging, and recycling is key towards a circular battery life cycle. However, as the usage of lithium-ion batteries (LIBs) expands in modern technologies and ever more complex elements are incorporated, complicated degradation behaviours are introduced, posing challenges for recycling. Metallurgy-based material extraction methods are independent of the complexity of materials decay but at the cost of compromised economic and environmental sustainability. Although direct regeneration is expected to reduce the environmental impact of recycling and improve its economic benefits, it cannot properly deal with failure at different scales and parameters. To effectively manage the growing stream of spent LIBs, strategies on multiple fronts are imperative. Recent developments in recycling mechanisms have highlighted the importance of understanding battery failure mechanisms to achieve environmentally friendly and sustainable recycling practices. In this Review, failure mechanisms in state-of-the-art LIBs are discussed from the particle scale to the cell scale, offering insights for navigating recycling efforts. Recent advancements in material extraction and direct regeneration are summarized, and perspectives on the most pressing challenges for recycling, optimization of recycling processes, and recycling strategies for next-generation batteries are offered.

LDA

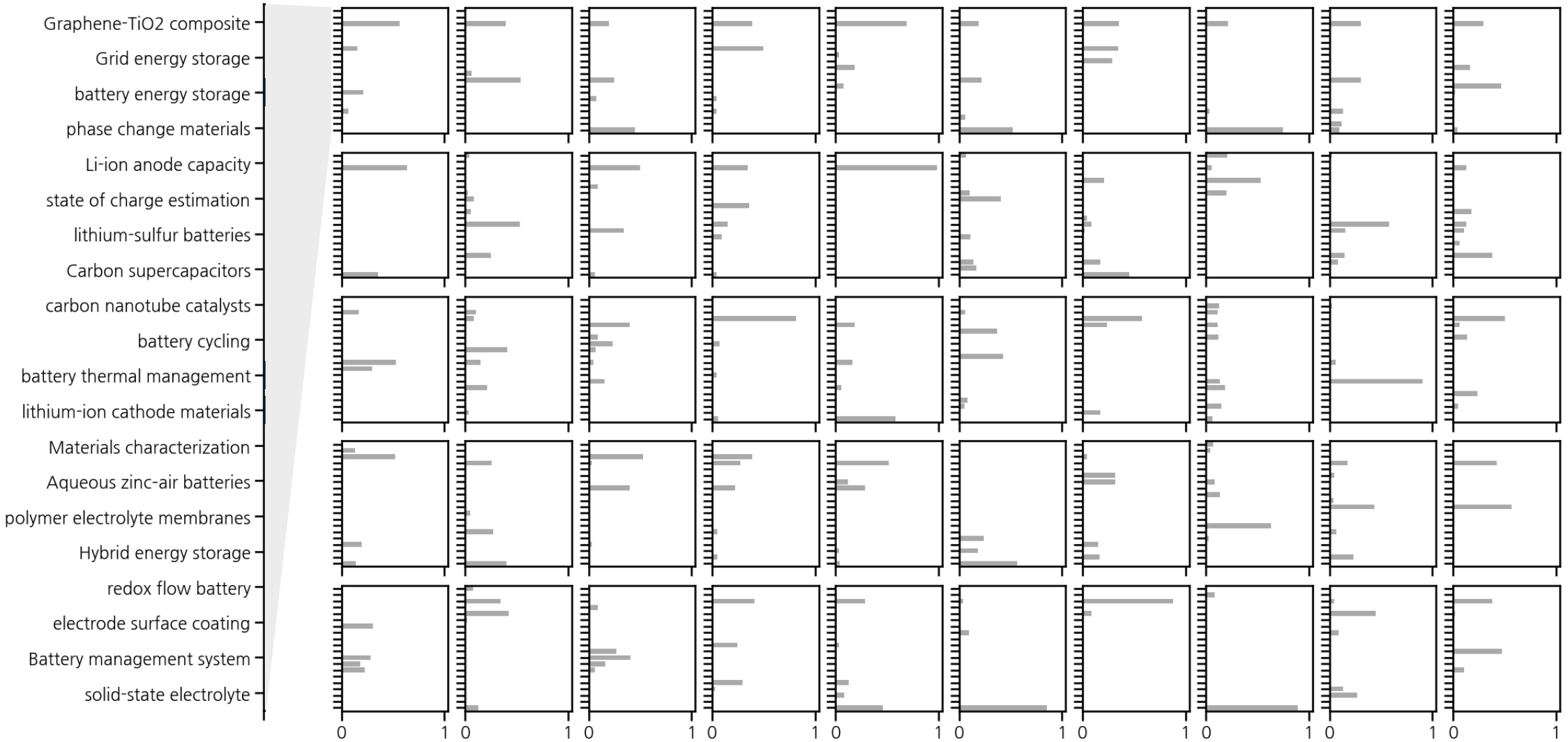
topic name	weight
Graphene-TiO2 composite	
Grid energy storage	
battery energy storage	
phase change materials	
Li-ion anode capacity	
state of charge estimation	
lithium-sulfur batteries	
Carbon supercapacitors	
carbon nanotube catalysts	
battery cycling	
battery thermal management	
lithium-ion cathode materials	
Materials characterization	
Aqueous zinc-air batteries	
polymer electrolyte membranes	
Hybrid energy storage	
redox flow battery	
electrode surface coating	
Battery management system	
solid-state electrolyte	



10.1038/s41578-025-00783-5

LDA Latent Dirichlet Allocation

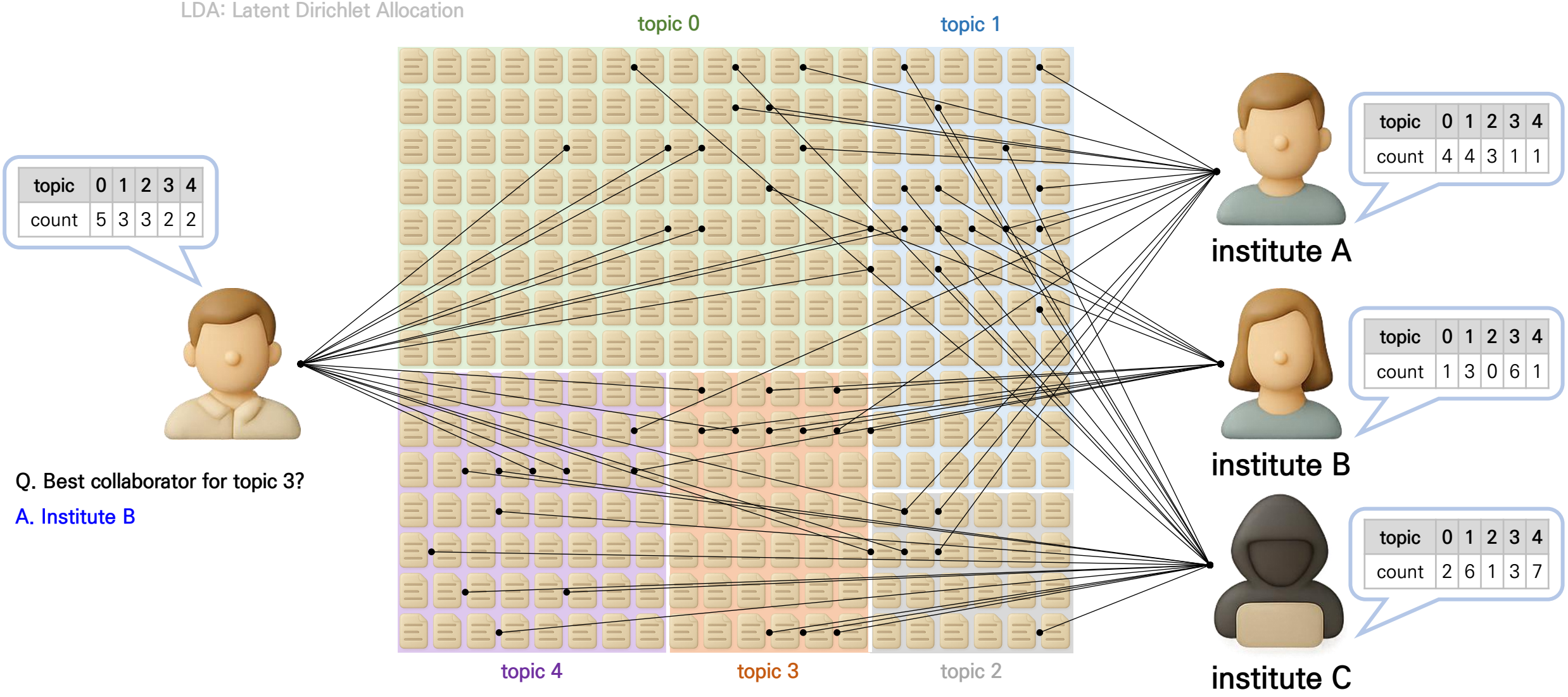
- 50 Battery papers



Identifying Strategic Research Partners by Data-Driven Analysis

Step 2. Clustering papers by topics. Then, best authors of a topic are elucidated.

LDA: Latent Dirichlet Allocation

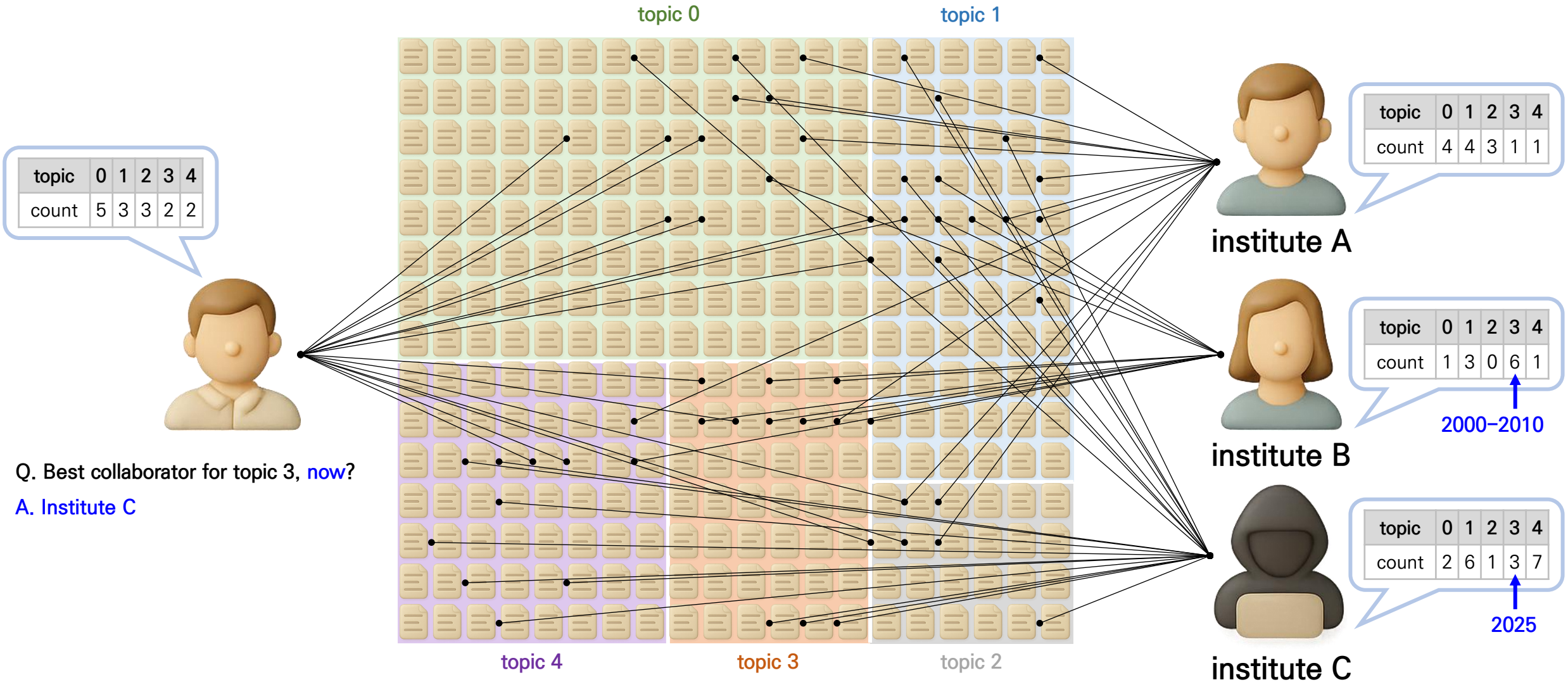


Q. Best collaborator for topic 3?

A. Institute B

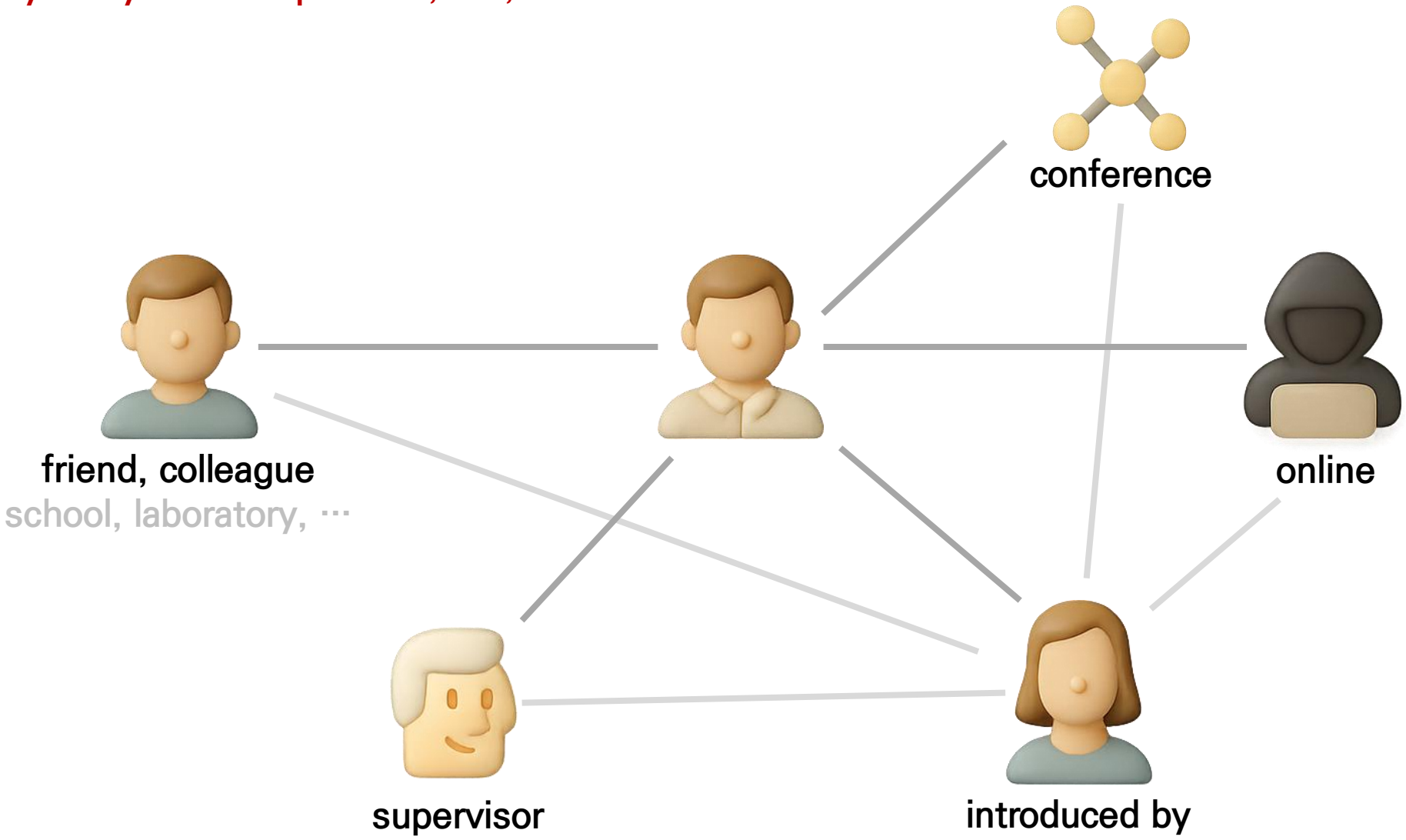
Identifying Strategic Research Partners by Data-Driven Analysis

Step 3. Analyze by features and desires ① year, ② nationality, ③ topic hierarchy, ④ co-authorship, …



You may claim "We already know all important players!"

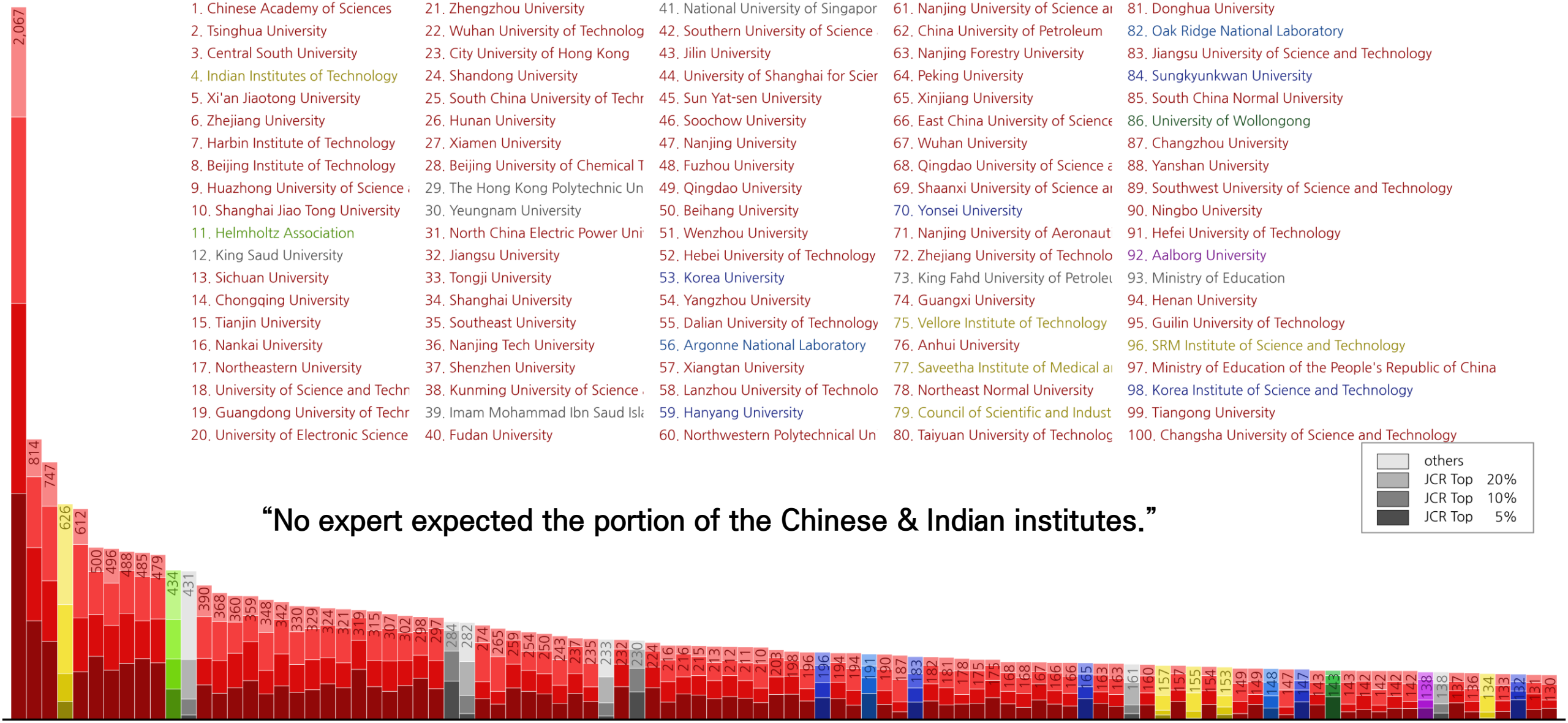
- Really? Maybe some persons, but, institutes?



2025.10.01. RD20, Jehyun LEE. jehyun.lee@gmail.com

Top 100 institutions by No. of Publications (2024)

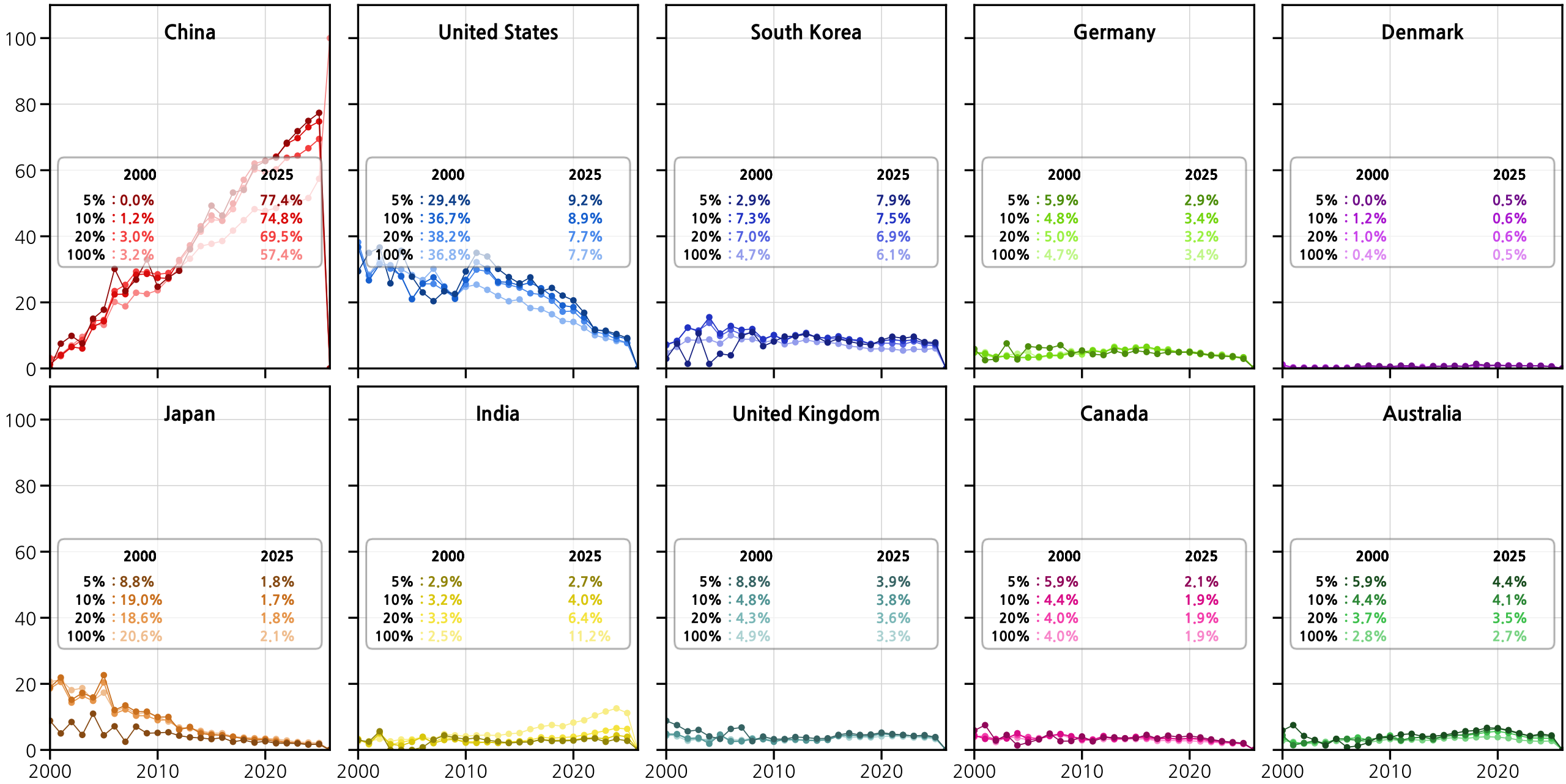
1. Chinese Academy of Sciences
2. Tsinghua University
3. Central South University
4. Indian Institutes of Technology
5. Xi'an Jiaotong University
6. Zhejiang University
7. Harbin Institute of Technology
8. Beijing Institute of Technology
9. Huazhong University of Science
10. Shanghai Jiao Tong University
11. Helmholtz Association
12. King Saud University
13. Sichuan University
14. Chongqing University
15. Tianjin University
16. Nankai University
17. Northeastern University
18. University of Science and Techn
19. Guangdong University of Techr
20. University of Electronic Science
21. Zhengzhou University
22. Wuhan University of Technolog
23. City University of Hong Kong
24. Shandong University
25. South China University of Techr
26. Hunan University
27. Xiamen University
28. Beijing University of Chemical T
29. The Hong Kong Polytechnic Un
30. Yeungnam University
31. North China Electric Power Uni
32. Jiangsu University
33. Tongji University
34. Shanghai University
35. Southeast University
36. Nanjing Tech University
37. Shenzhen University
38. Kunming University of Science
39. Imam Mohammad Ibn Saud Isl
40. Fudan University
41. National University of Singapor
42. Southern University of Science
43. Jilin University
44. University of Shanghai for Scier
45. Sun Yat-sen University
46. Soochow University
47. Nanjing University
48. Fuzhou University
49. Qingdao University
50. Beihang University
51. Wenzhou University
52. Hebei University of Technology
53. Korea University
54. Yangzhou University
55. Dalian University of Technology
56. Argonne National Laboratory
57. Xiangtan University
58. Lanzhou University of Technolo
59. Hanyang University
60. Northwestern Polytechnical Un
61. Nanjing University of Science ar
62. China University of Petroleum
63. Nanjing Forestry University
64. Peking University
65. Xinjiang University
66. East China University of Science
67. Wuhan University
68. Qingdao University of Science a
69. Shaanxi University of Science ar
70. Yonsei University
71. Nanjing University of Aeronaut
72. Zhejiang University of Technolo
73. King Fahd University of Petrole
74. Guangxi University
75. Vellore Institute of Technology
76. Anhui University
77. Saveetha Institute of Medical ar
78. Northeast Normal University
79. Council of Scientific and Indust
80. Taiyuan University of Technolog
81. Donghua University
82. Oak Ridge National Laboratory
83. Jiangsu University of Science and Technology
84. Sungkyunkwan University
85. South China Normal University
86. University of Wollongong
87. Changzhou University
88. Yanshan University
89. Southwest University of Science and Technology
90. Ningbo University
91. Hefei University of Technology
92. Aalborg University
93. Ministry of Education
94. Henan University
95. Guilin University of Technology
96. SRM Institute of Science and Technology
97. Ministry of Education of the People's Republic of China
98. Korea Institute of Science and Technology
99. Tiangong University
100. Changsha University of Science and Technology



"No expert expected the portion of the Chinese & Indian institutes."

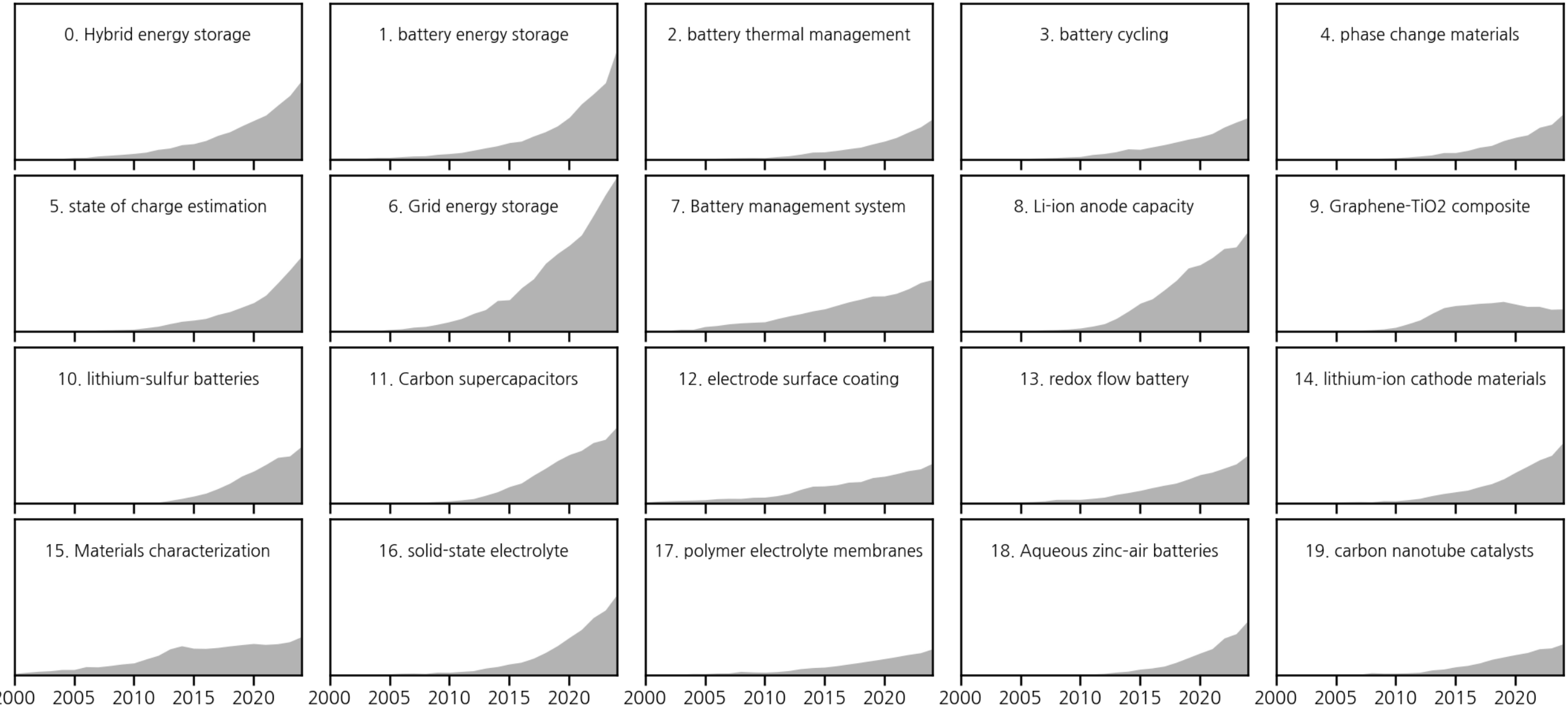
2025.10.01. RD20, Jehyun LEE. jehyun.lee@gmail.com

“No expert expected 77% occupation of top 5% publications by Chinese institutes, and less than 10% occupation of all publications by United States.”

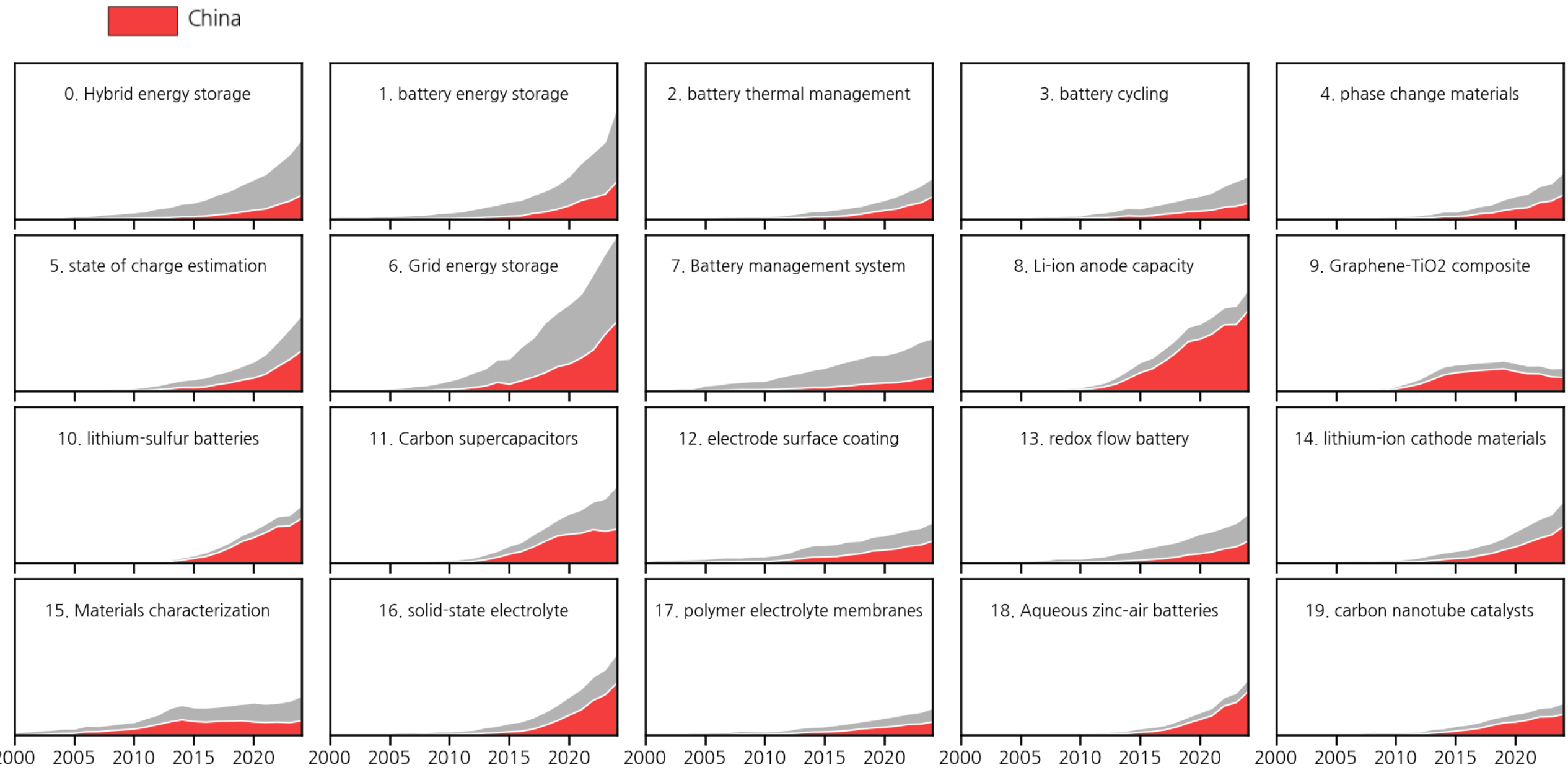


topic vs time

- world

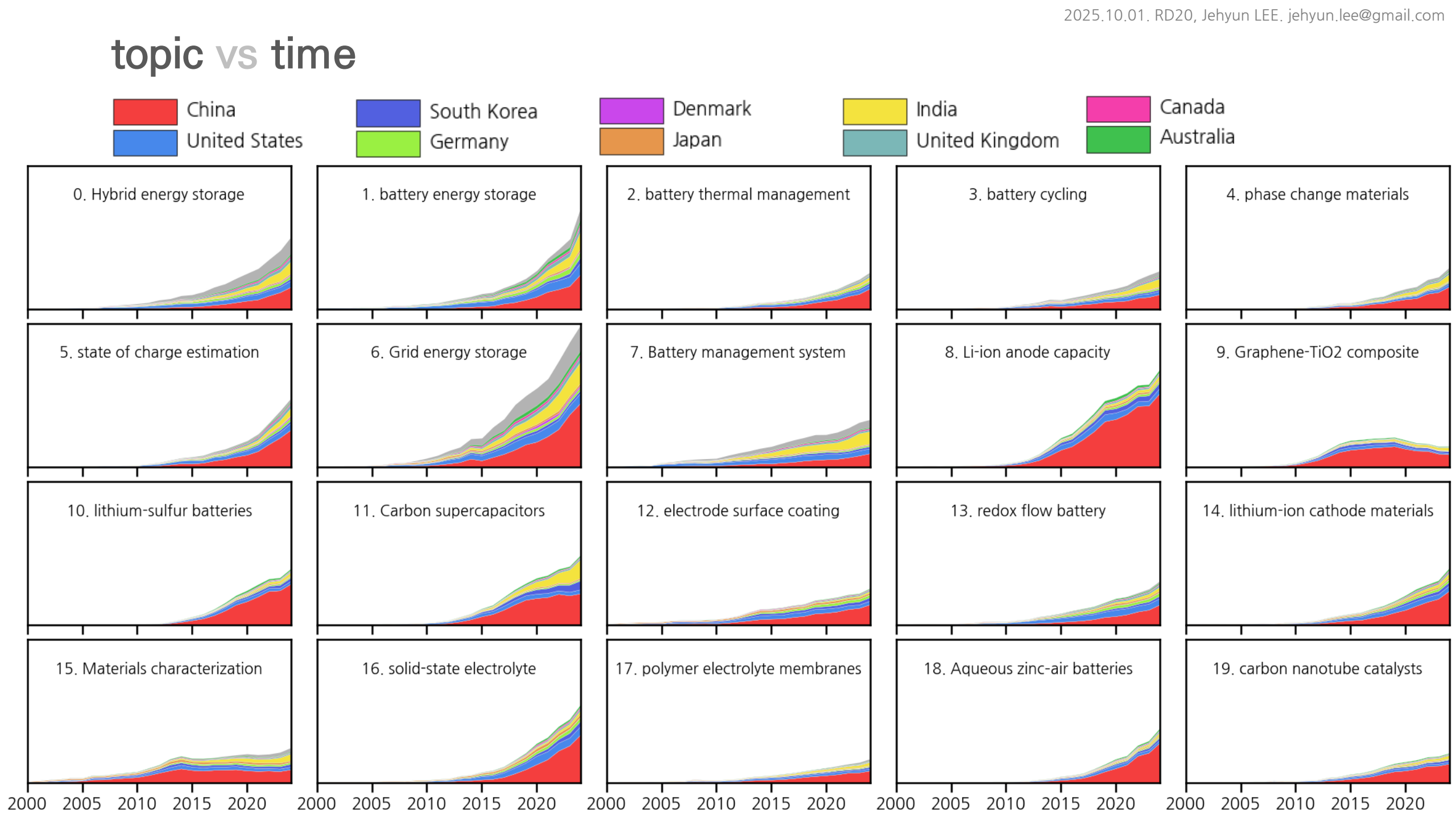


topic vs time



2025.10.01. RD20, Jehyun LEE. jehyun.lee@gmail.com

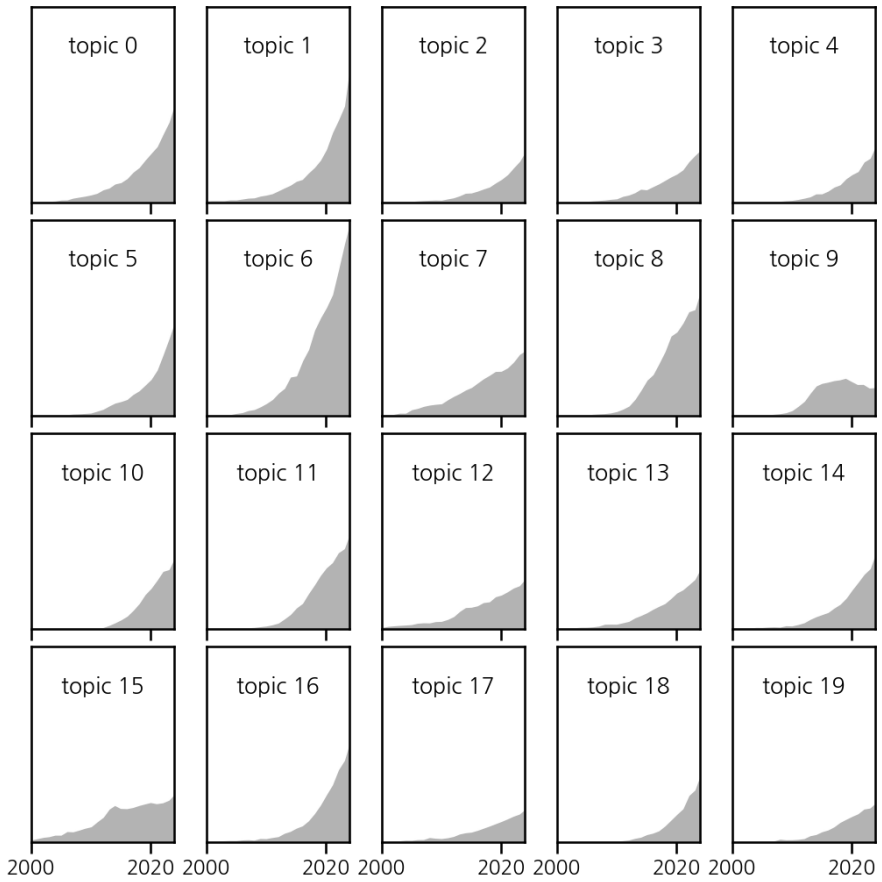
topic vs time



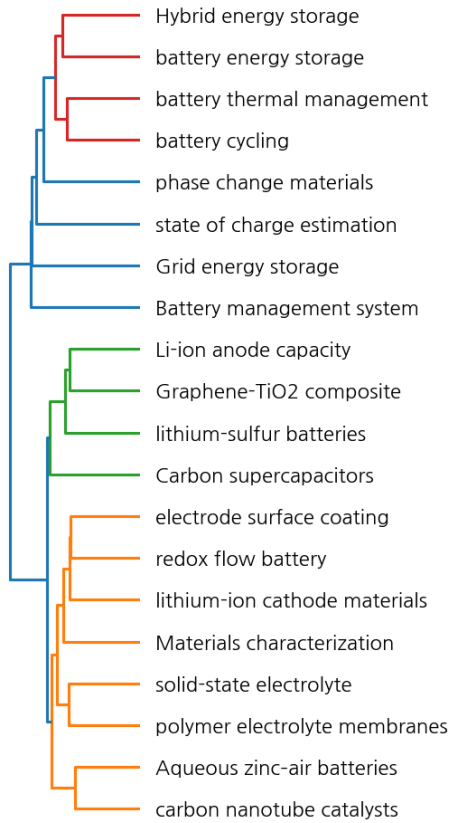
2025.10.01. RD20, Jehyun LEE. jehyun.lee@gmail.com

topic vs time matrix

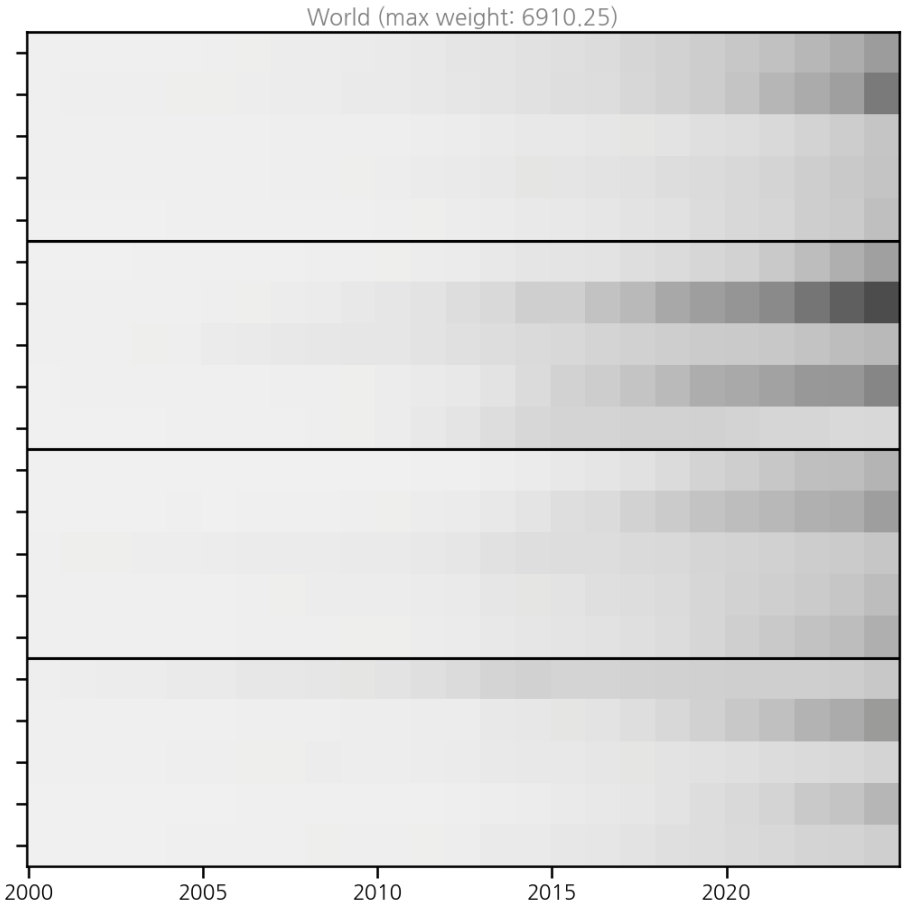
stackplot



hierarchical clustering

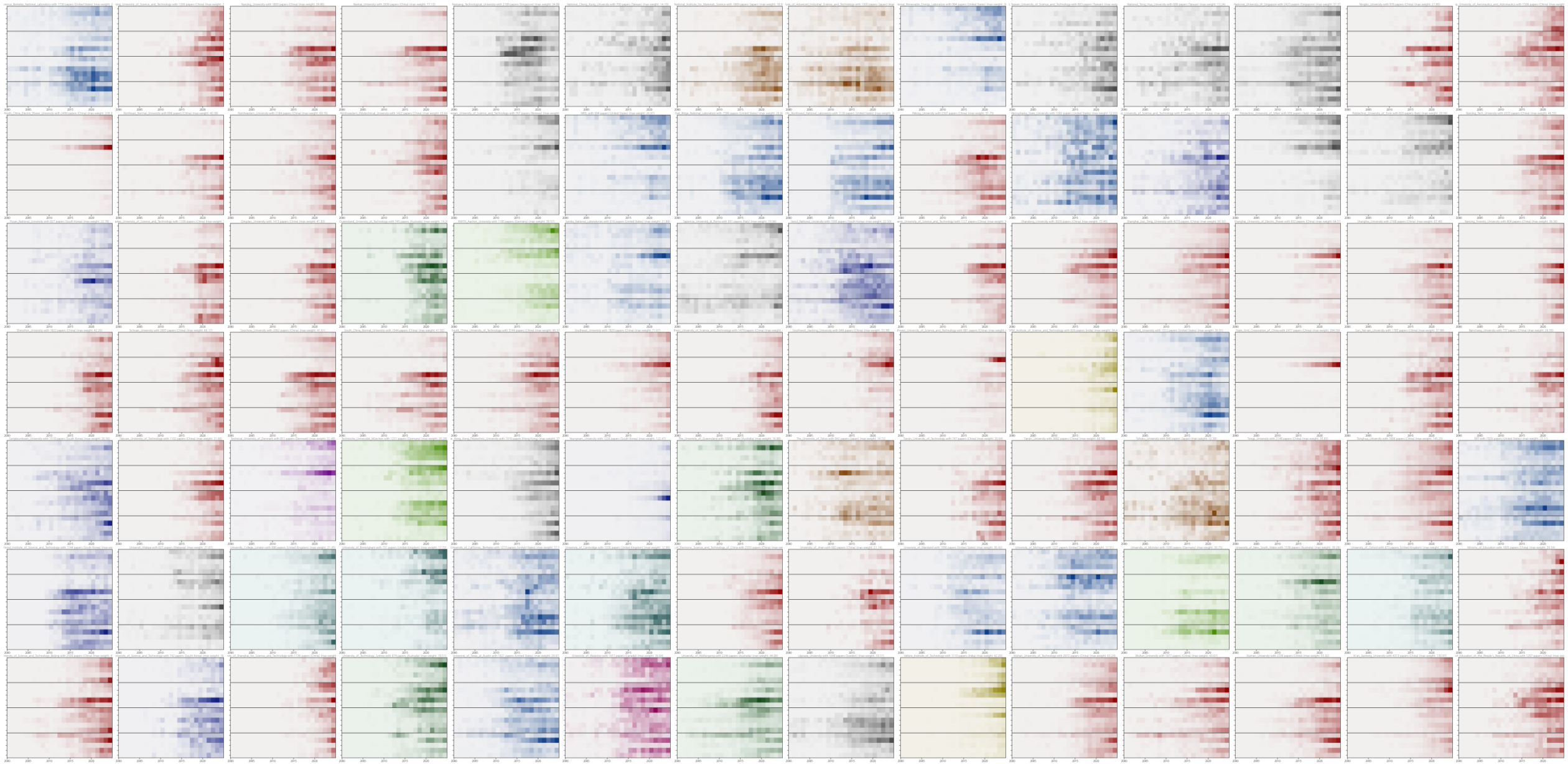


heatmap



2025.10.01. RD20, Jehyun LEE. jehyun.lee@gmail.com

topic vs time matrix of institute = fingerprint



2025.10.01. RD20, Jehyun LEE. jehyun.lee@gmail.com

topic vs time matrix of institute → index

$X_i \in \mathbb{R}^{25\times 20}$ topic data of an institution $Z_i = \frac{X_i - \mu_w}{\sigma_w}$ world based Z-score

IOI institute orientation index

- Cosine similarity to the world matrix
- alignment of the institute's spatio-temporal pattern with the world

$$\frac{\langle Z_i, Z_c \rangle}{\|Z_i\|_F \|Z_w\|_F} \in [-1,1]$$

IEI institute energy index

- Frobenius norm of raw data
- total magnitude/activity across years x features

$$\|Z_i\|_F$$

PS projection share

- Fraction of world "energy" that lies along the institution's direction
- $\approx \cos^2\theta$

$$\frac{\langle Z_w, \hat{Z}_i \rangle^2}{\|Z_w\|_F^2} = \cos^2\theta \in [0,1]$$

FSI feature salience index

- Log over/under-representation of each feature vs the world share
- 〉0 : over-represented

$$s_i(f) = \sum_t X_i(t,f), \quad p_i(f) = \frac{s_i(f)}{\sum_f s_i(f)}, \quad \mathrm{FSI}_i(f) = \log \frac{p_i(f)}{p_w(f)}$$

ID institute distinctiveness

- 1-IOI
- how different the country's pattern is from the world

TAF time alignment by feature

- per-feature cosine of institute vs world time series
- timing co-movement (-1..1)

IIF institute interaction fingerprint

- Feature-feature Gram matrix on normalized data
- coupling/co-variation structure

$$G_i = \frac{1}{T} Z_i^T Z_i \quad \text{(FxF Gram)}$$

- △IIF : difference from the world's coupling matrix
  stronger/weaker than global

$$\Delta G_i = G_i - G_w$$

OVL overlap similarity

- Cosine similarity between two institutes' matrices
- extent of co-activation/overlap

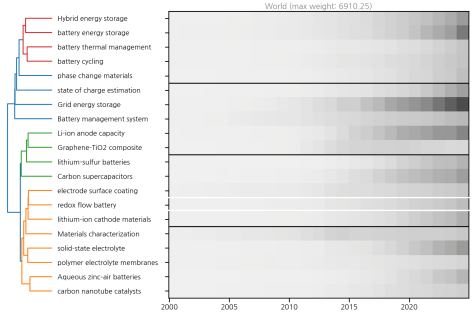
$$\frac{\langle Z_i, Z_j \rangle}{\|Z_i\|_F \|Z_j\|_F}$$

Rolling variants

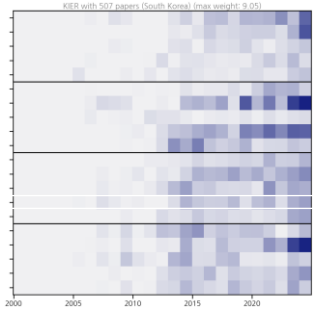
- windows versions (IOI/energy/TAF) to track shifts and breakpoints

topic vs time matrix of institute → index

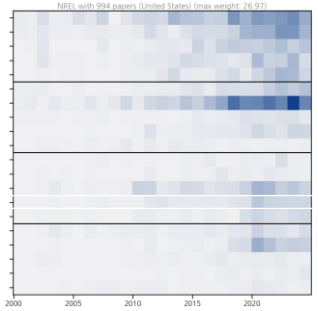
- **indices of some institutes**



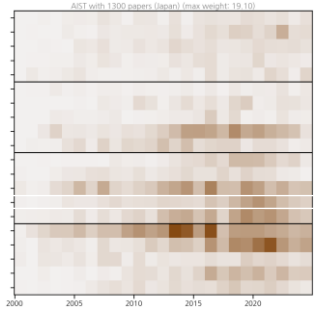
world



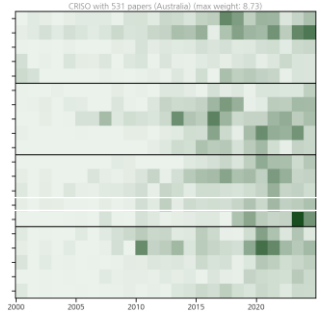
KIER South Korea



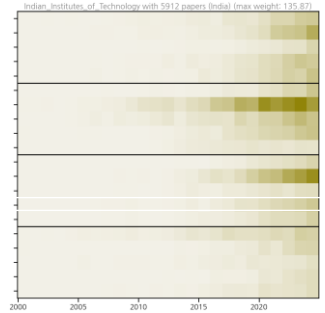
NREL United States



AIST Japan



CSIRO Australia

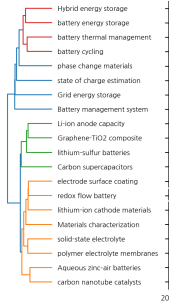


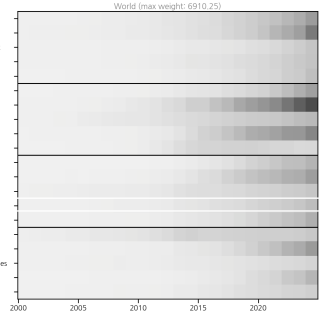
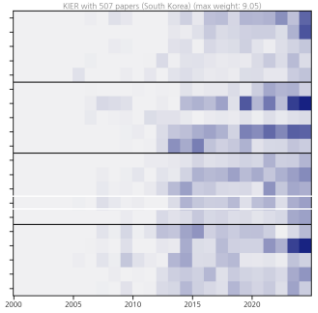
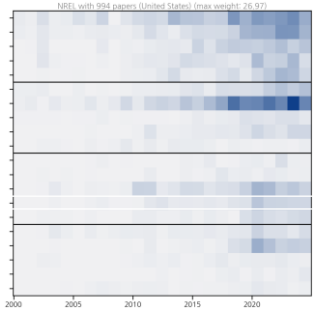
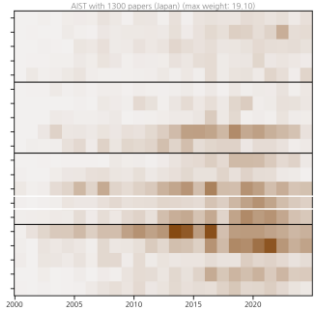
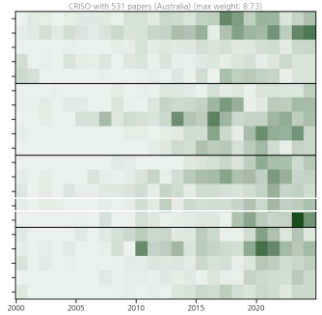
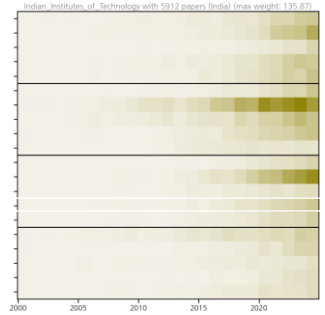
IIT India

	world	KIER South Korea	NREL United States	AIST Japan	CSIRO Australia	IIT India
IOI institution orientation index	1.0	0.00126	0.00209	0.00163	0.00099	0.01860
ID institution distinctiveness	0.0	0.99874	0.99791	0.99837	0.99901	0.98140
IEI institution energy index	26763.2	37.9	84.8	89.5	36.6	499.8
PS projection share	1.0	1.59×10^{-6}	4.37×10^{-6}	2.65×10^{-6}	9.83×10^{-7}	3.46×10^{-4}
FSI feature salience index	0.0	0.269	0.211	0.469	0.163	0.106
• Redox flow battery						
TAF time alignment by feature	1.0	0.00154	0.00256	0.00339	0.00126	0.02380
• Redox flow battery						

topic vs time matrix of institute → index → best colleague

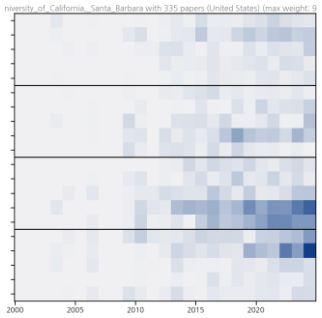
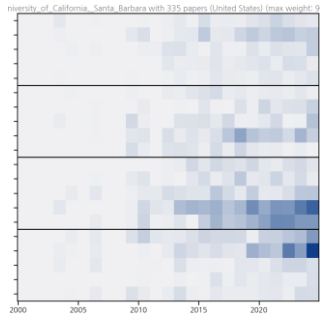
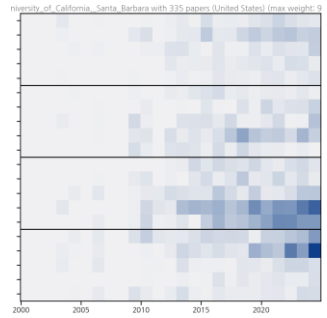
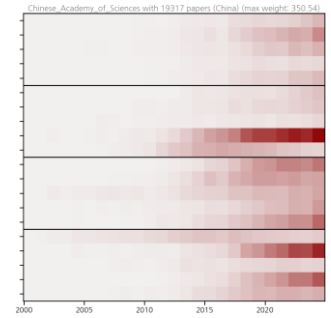
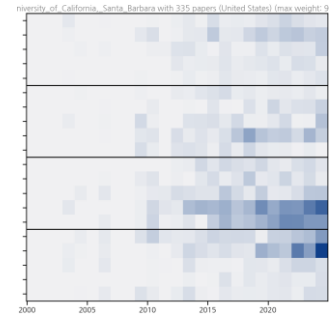
- Best colleague for Redox flow battery R&D?
- Opposite tendency for other features – expecting another collaboration



world	KIER South Korea	NREL United States	AIST Japan	CSIRO Australia	IIT India
					
UCSB 0.709	UCSB 0.710	UCSB 0.711	CAS 0.712	UCSB 0.711
CAS 0.706	CAS 0.709	CAS 0.689	UCSB 0.709	U Kentucky 0.690
U Kentucky 0.633	U Kentucky 0.634	U Kentucky 0.636	U Kentucky 0.633	Harvard U 0.666
Harvard U 0.610	Harvard U 0.611	Harvard U 0.612	Harvard U 0.610	Lomonosov Moscow 0.658
Lomonosov Moscow 0.603	Lomonosov Moscow 0.604	Lomonosov Moscow 0.604	Lomonosov Moscow 0.603	EU Synchrotoron Rad Fasc. 0.626

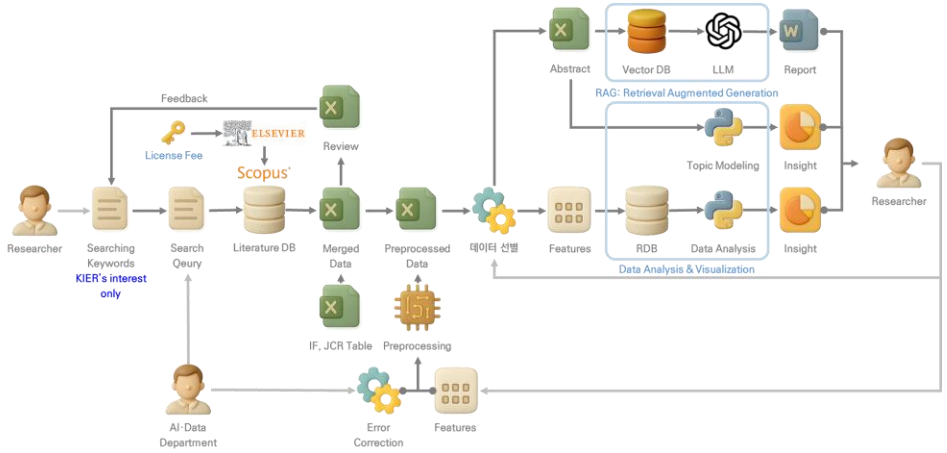
$$\text{score}(x) = \sum_i w_i A_i$$

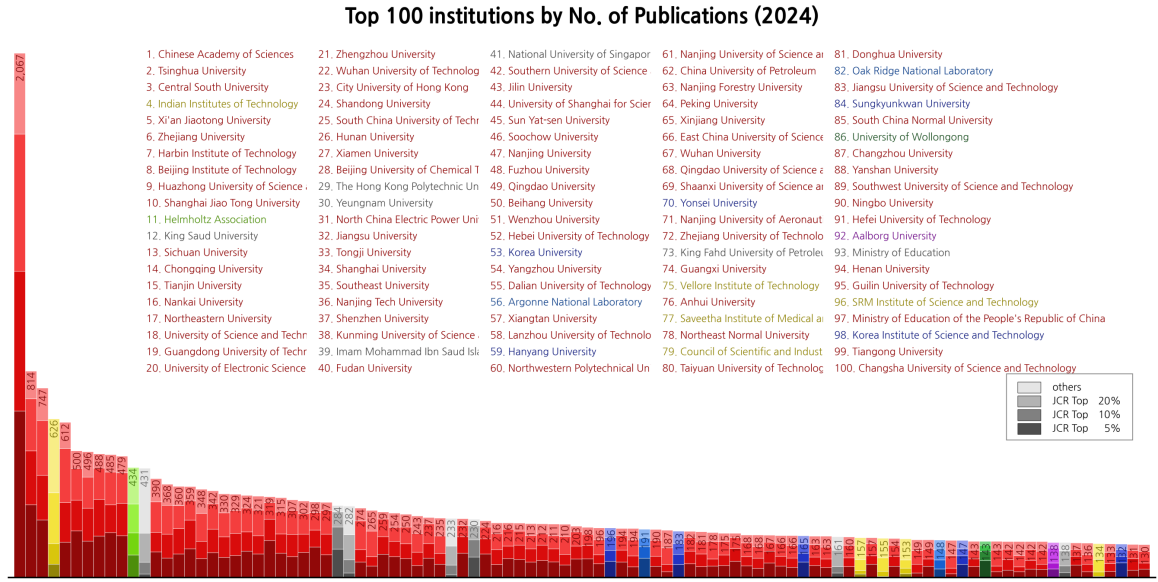
Best Colleague

UCSB United States	UCSB United States	UCSB United States	CAS China	UCSB United States
				

Conclusion

- Data-driven method can reach unknown unknowns
  - Data pipeline with preprocessing is essential





- Status of research of institutes can be quantified as a form of index
- Collabortor searching was supported by numerical methods