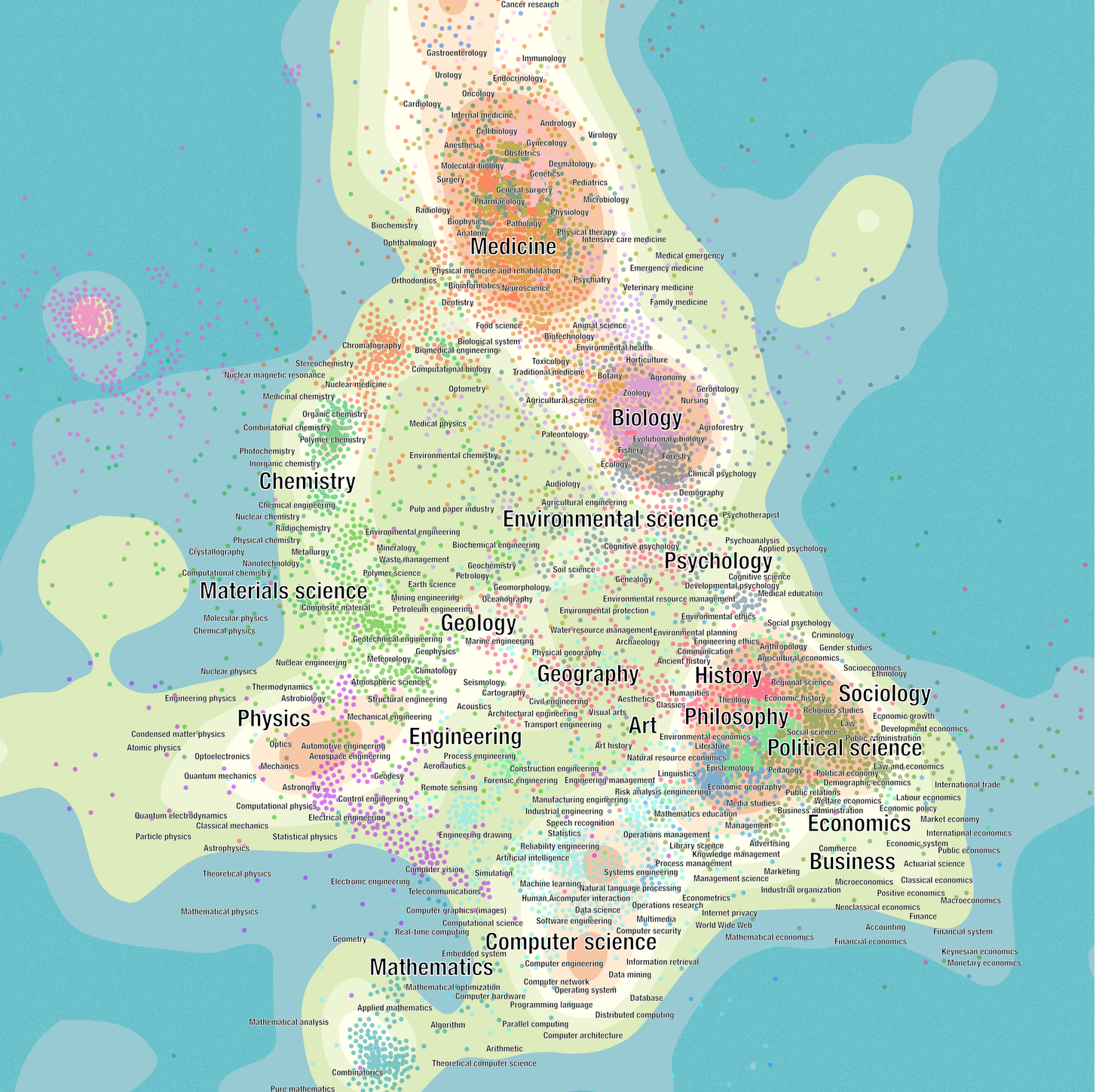
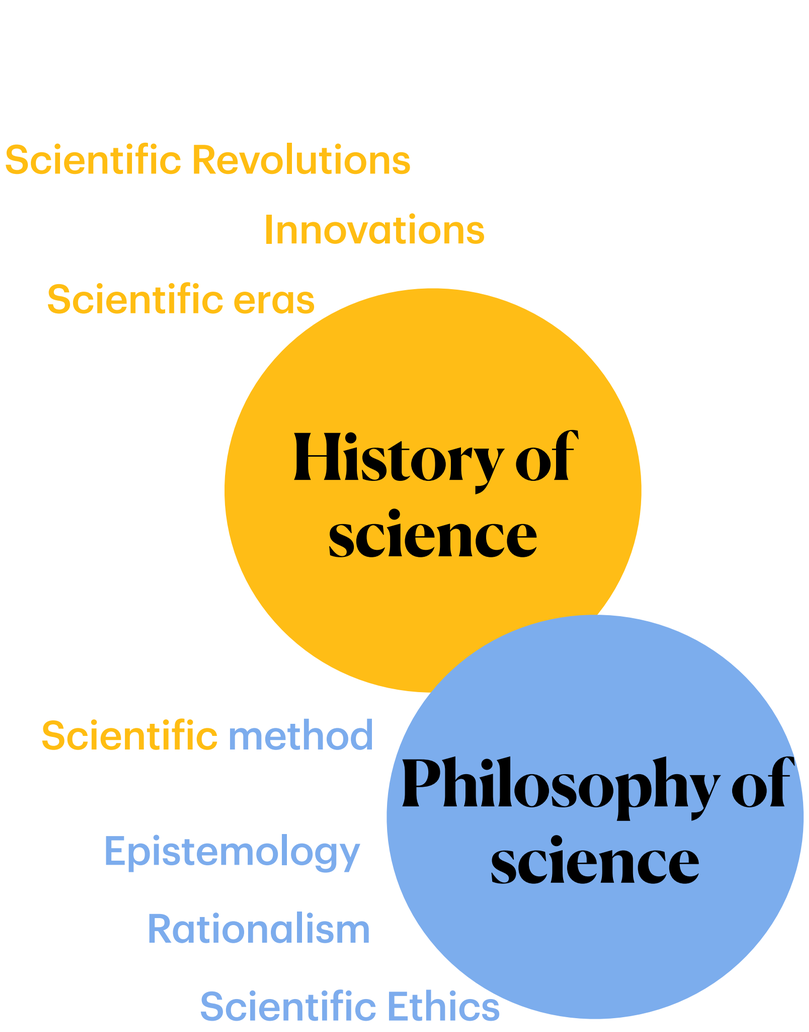


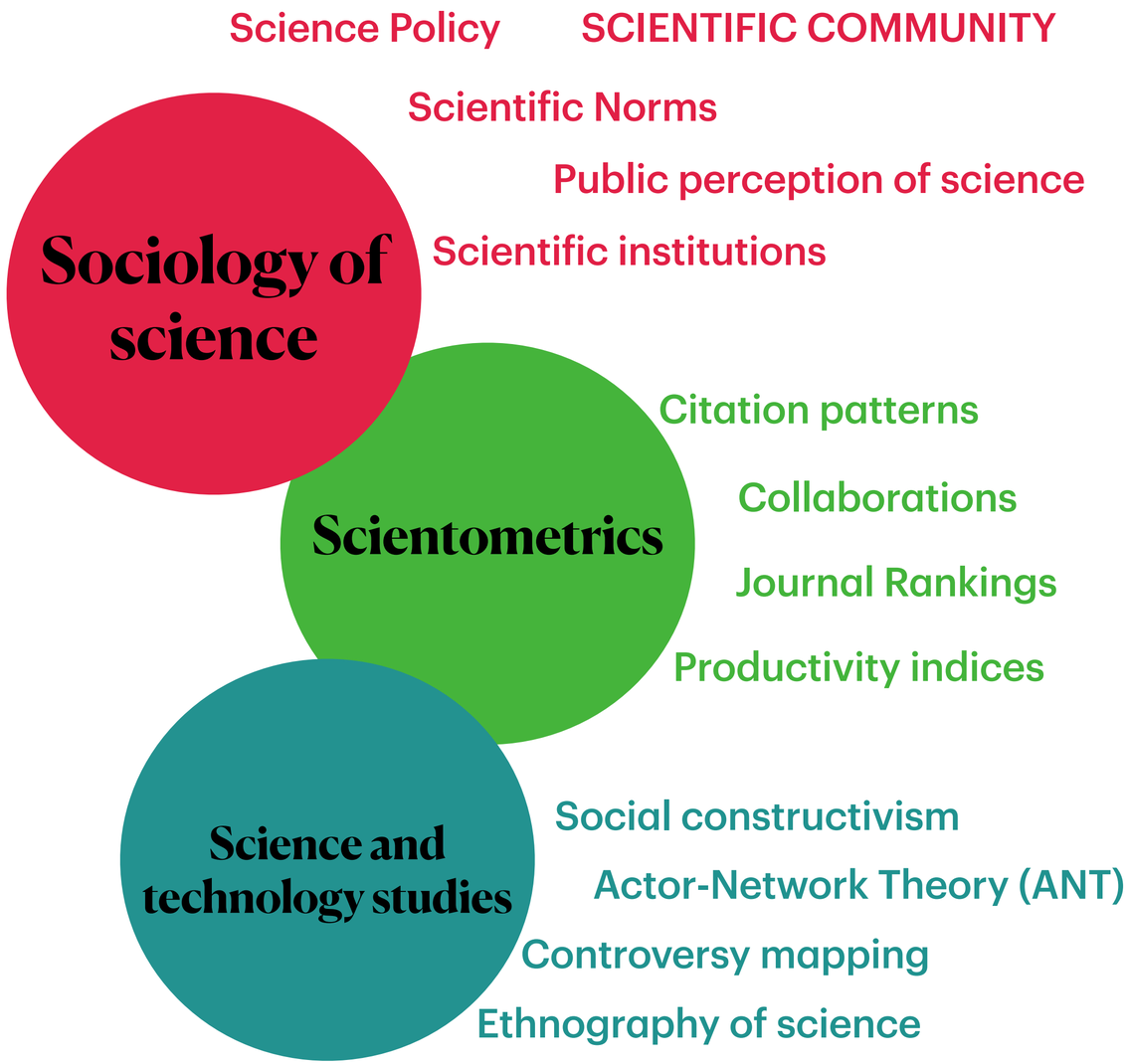
Science of science

Analyzing the impact of AI on Science

Floriana Gargiulo

Traditions of Science studies





Science of science

Scientometrics

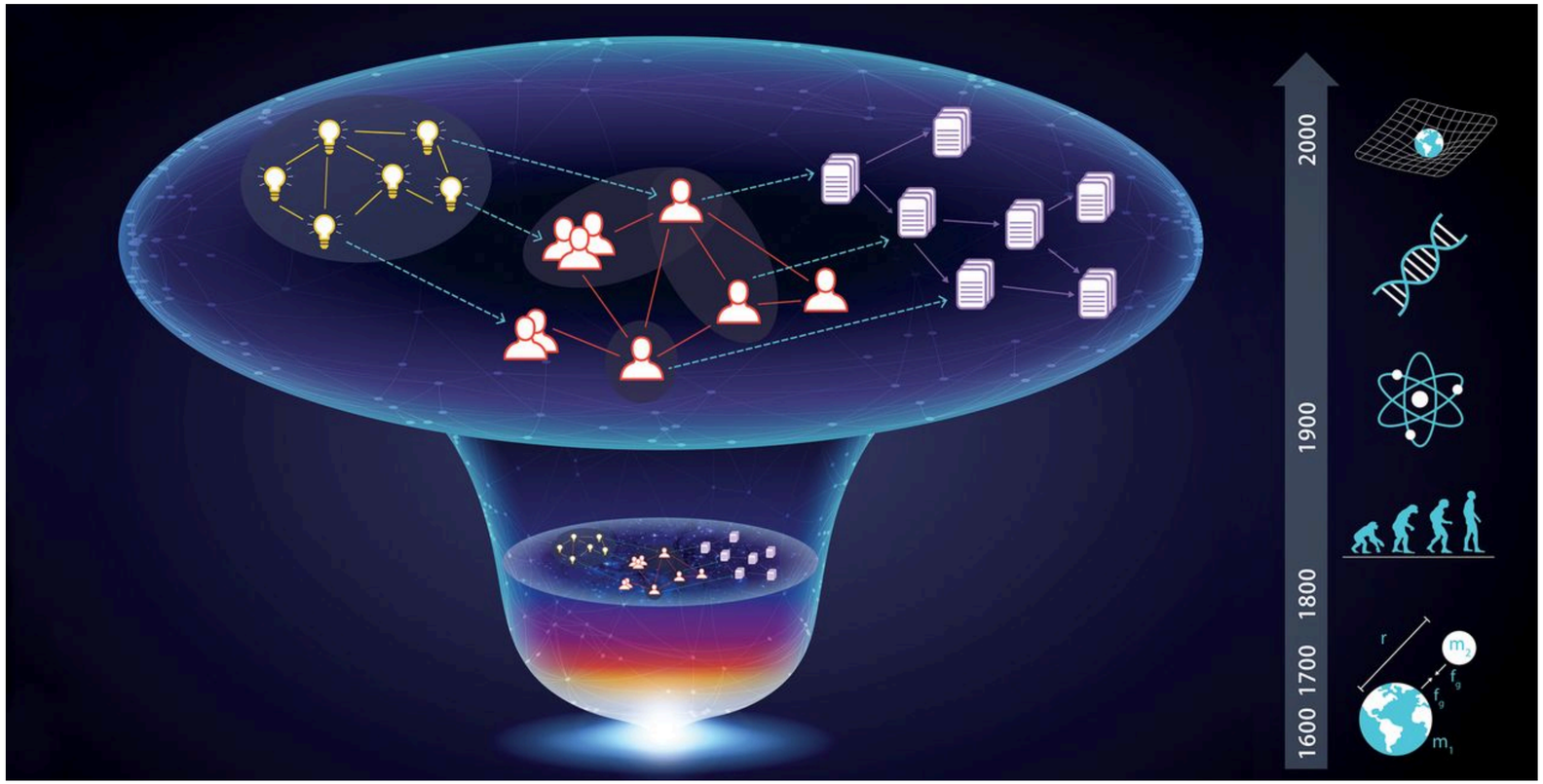
- Citation patterns
- Collaborations
- Journal Rankings
- Productivity indices

Large-scale datasets

Modeling tools

Science as a complex system

An evolving network of scientists, papers and concepts



Artificial intelligence in science

"Few fields are untouched by the machine-learning revolution, from materials science to drug exploration; quantum physics to medicine." (Nature Editorial, 2019)

"The era of direct experimentation is gone, replaced by the era of data collection[...] Science takes place within the data." (Hey, Tansley, & Tolle, 2009)

AI can be characterized as a "general method of invention," which brings benefits and challenges to the science systems. (Bianchini, Muller, & Pelletier, 2022)

Is AI changing the epistemic structure and the scientific practices?

Is AI changing the epistemic structure and the scientific practices?

Floriana Gargiulo, Sylvain Fontaine, Michel Dubois, Paola Tubaro; **A meso-scale cartography of the AI ecosystem.** *Quantitative Science Studies* 2023; doi: https://doi.org/10.1162/qss_a_00267

- What is AI?
- How AI diffuses in the scientific ecosystem?

Fontaine, S., Gargiulo, F., Dubois, M., & Tubaro, P. (2023). **Epistemic integration and social segregation of AI in neuroscience.** *arXiv preprint arXiv:2310.01046*.

- The case study of Neuroscience
- Is AI creating a separate Neuroscience sub-discipline?

What is AI?

A structured set of technical, computational and mathematical tools, developed by computer scientists and mathematicians and applied in other disciplines.

First challenge:
How to build a dataset to address our research questions?

1. Retrive a large set of keywords associated to AI exploring the Wikipedia pages associated to AI and several AI glossaries available on the web.

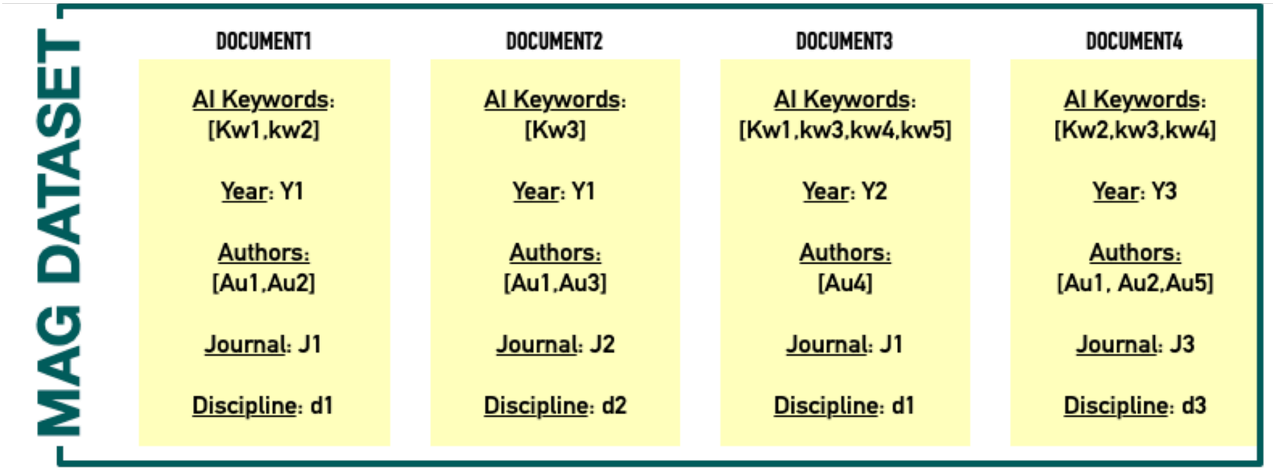
   ~600 AI concepts

2. Query on OpenAlex (before MAG) to extract all the papers containing the AI keywords.

   ~2,7 millions of papers containing one or more AI concepts

3. Assign to papers a discipline, using the journal classification of the Web of Science

   ~1,1 millions of papers containing one or more AI concepts and an associated discipline

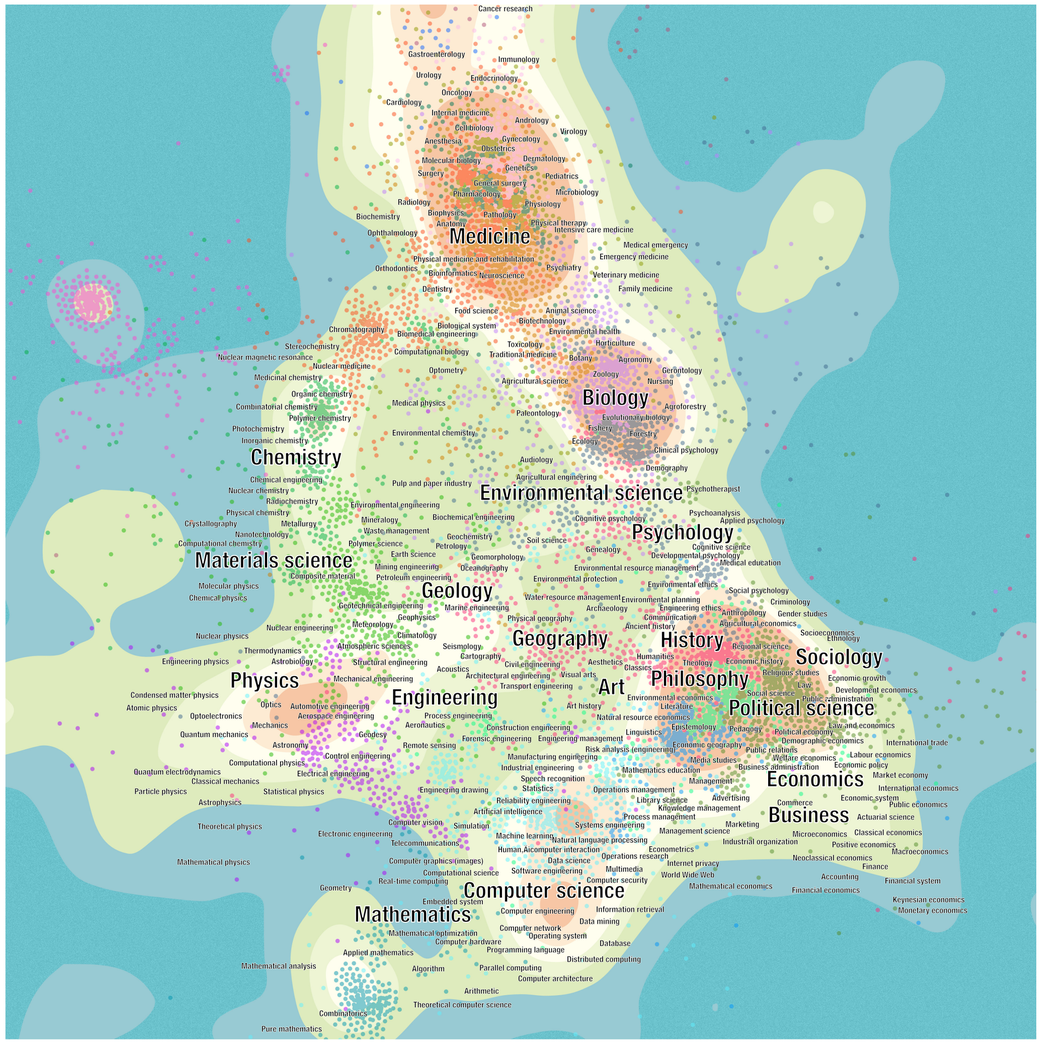


What is AI?

A structured set of technical, computational and mathematical tools, developed by computer scientists and mathematicians and applied in other disciplines.

First challenge - part 2:

How can we define a distance in the disciplinary space?



1. 10000 random samples of 100.000 papers on MAG with their references.

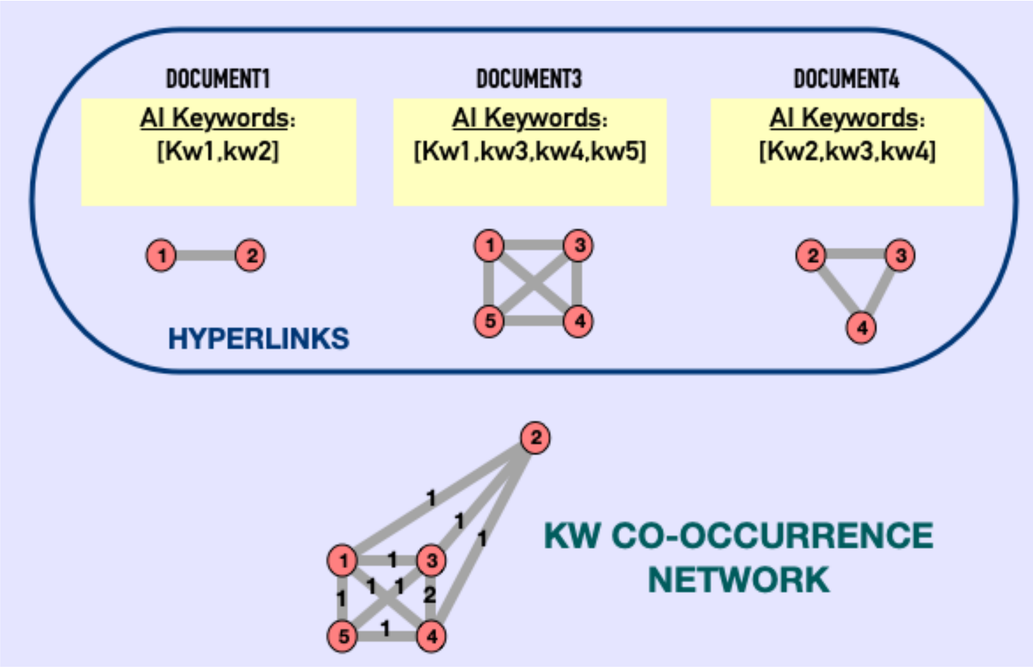
2. Build the co-citation matrix aggregated at the level of disciplines

3. The distance between disciplines is defined through the pointwise mutual information (PMI):

$$\mathrm{pmi}_{ij} = \max\left(2\log_2\left(\frac{w_{ij}}{\sum_k(w_{ik})\sum_k(w_{jk})}\right), 0\right)$$

$$D_{ij} = 1 - pmi_{ij}$$

The meso-scale structure of AI



Network partitioning in communities: detecting group of nodes that have more connections among them than with the rest of the graph.

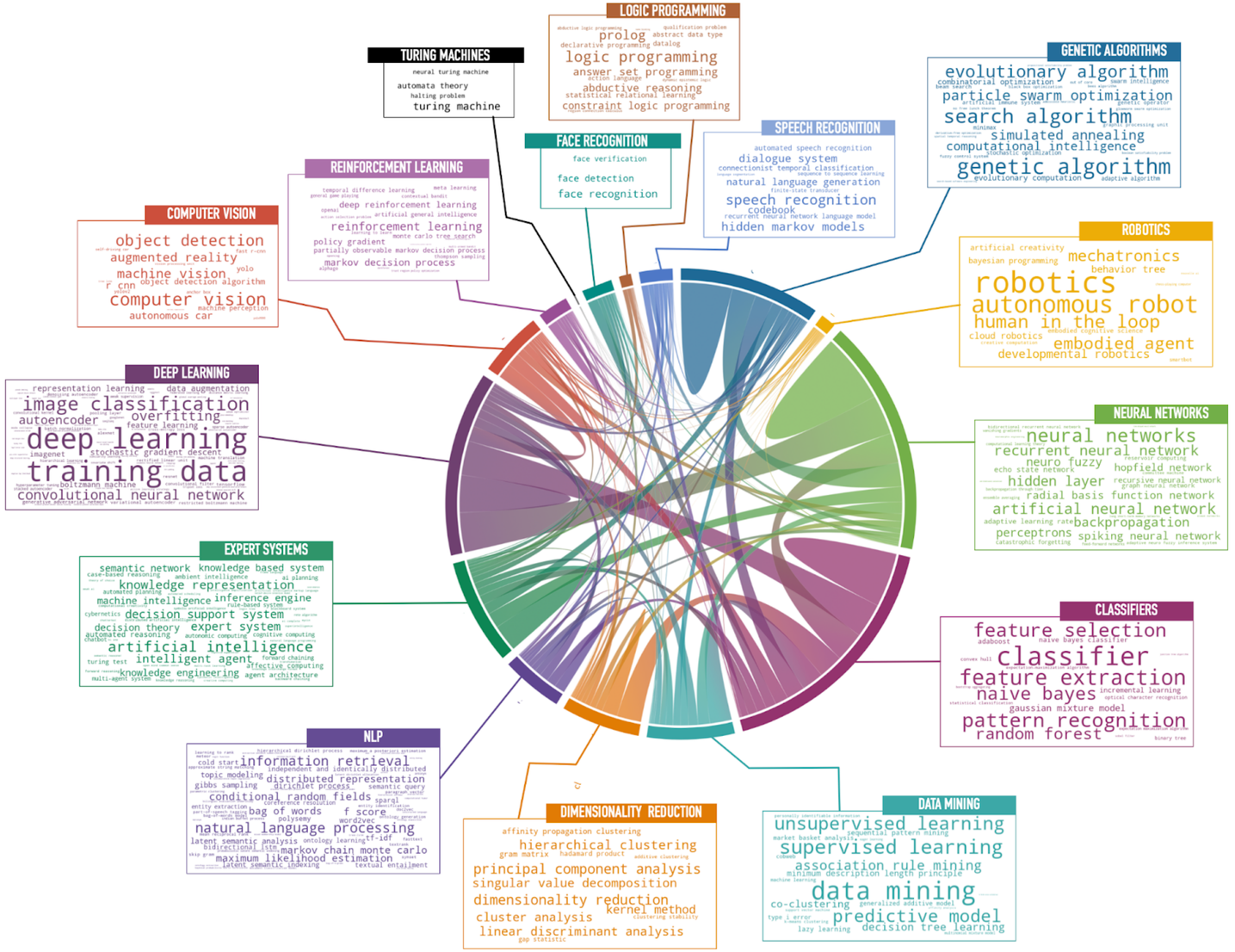
Performed using the Louvain algorithm (greedy optimization of modularity)

Technical issue:
The graph is very dense (several edges with very heterogeneous weights)

We apply a disparity filter on the original graph

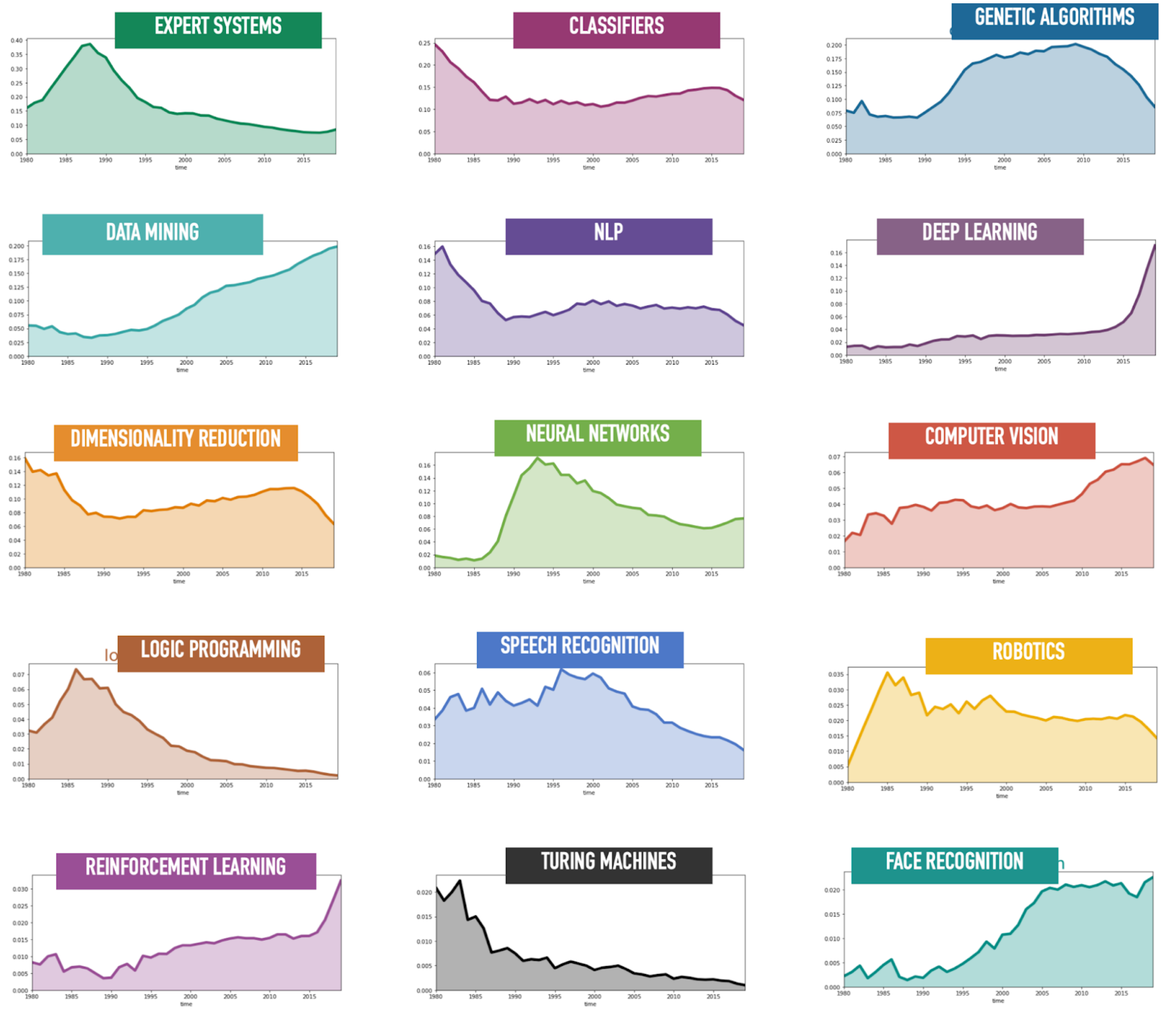
(Serrano, M. A., Boguna´, M., and Vespignani, A. (2009). Extracting the multiscale backbone of complex weighted networks. Proceedings of the national academy of sciences, 106(16):6483–6488.)

The meso-scale structure of AI



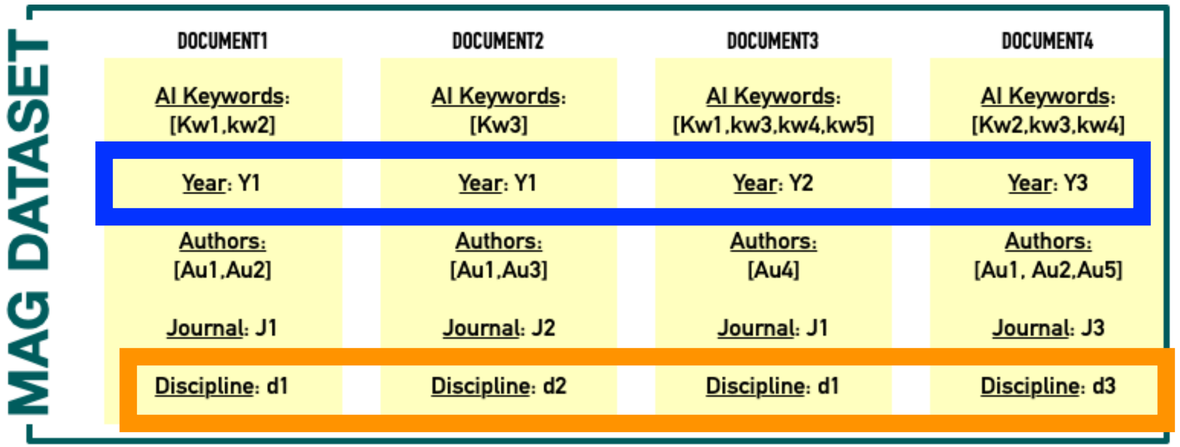
We identify **15 AI sub-categories**

The meso-scale structure of AI



We identify **15 AI sub-categories** with different temporal patterns

AI from development to application...



We define **Mathematics and Computer science** the "**originating disciplines**" of AI

For each year we take the list of disciplines in which AI keywords are used

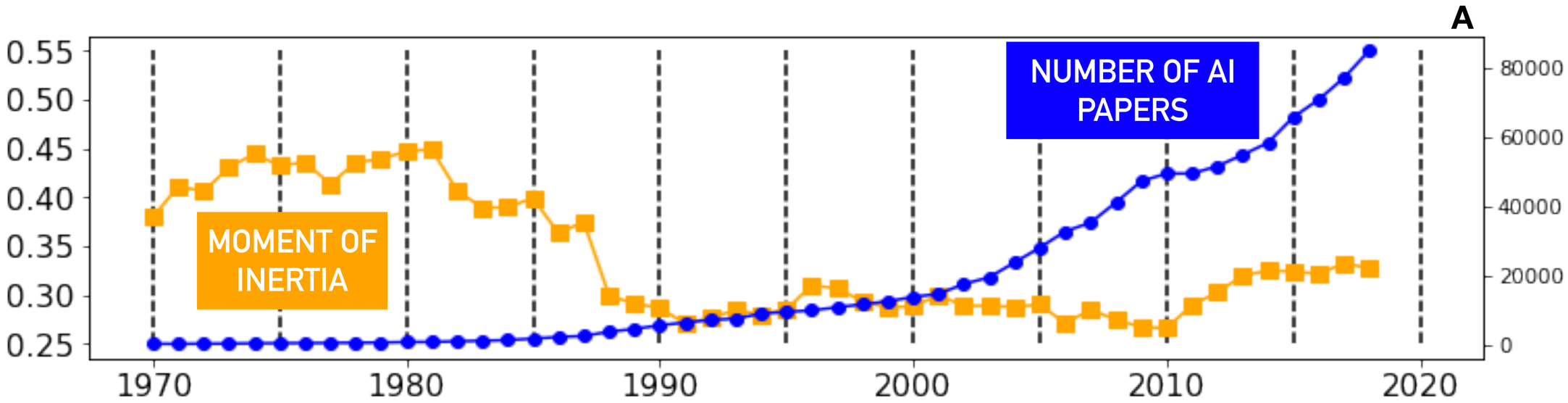
Y1: {d1: 145, d2:364,...}
Y2: {d1: 137, d2:789,...}
Y3: {d1: 56, d2:1034,...}
....

Inertia moment around originating disciplines

$$m_I = \sum_i \frac{n_i}{N_{tot}} \min(D_{i,CS}, D_{i,Math}, D_{i,Stat})^2$$

LOW MI = CONCENTRATION AROUND ORIGINATING
HIGH MI = DISCIPLINATY DISPERSION

AI from development to application...

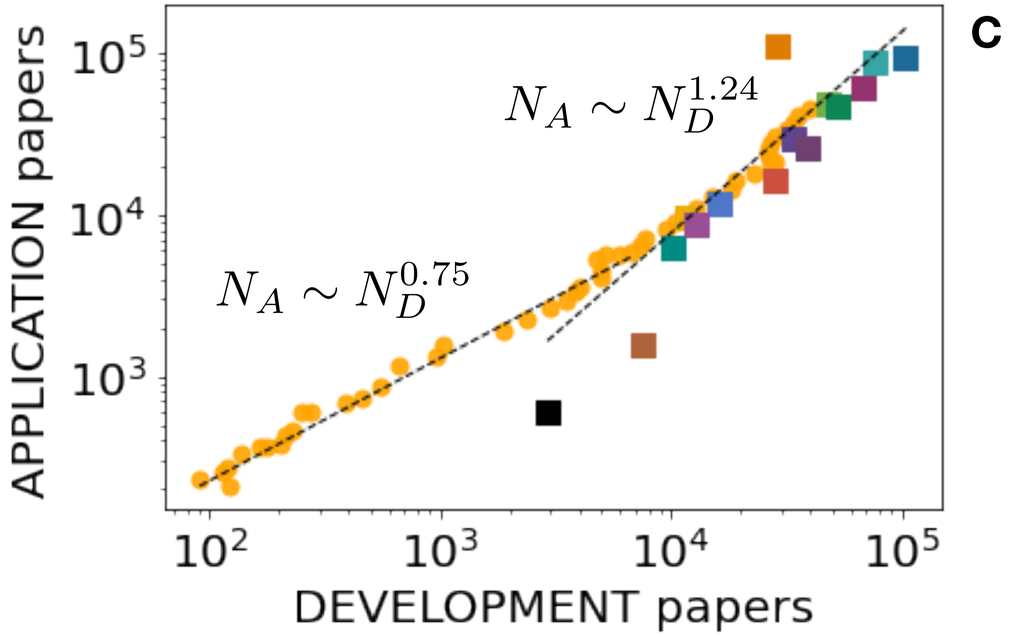


High dispersion around native disciplines

CONCENTRATION PHASE: AI is concentrated around originating disciplines (Mathematics and Computer science).

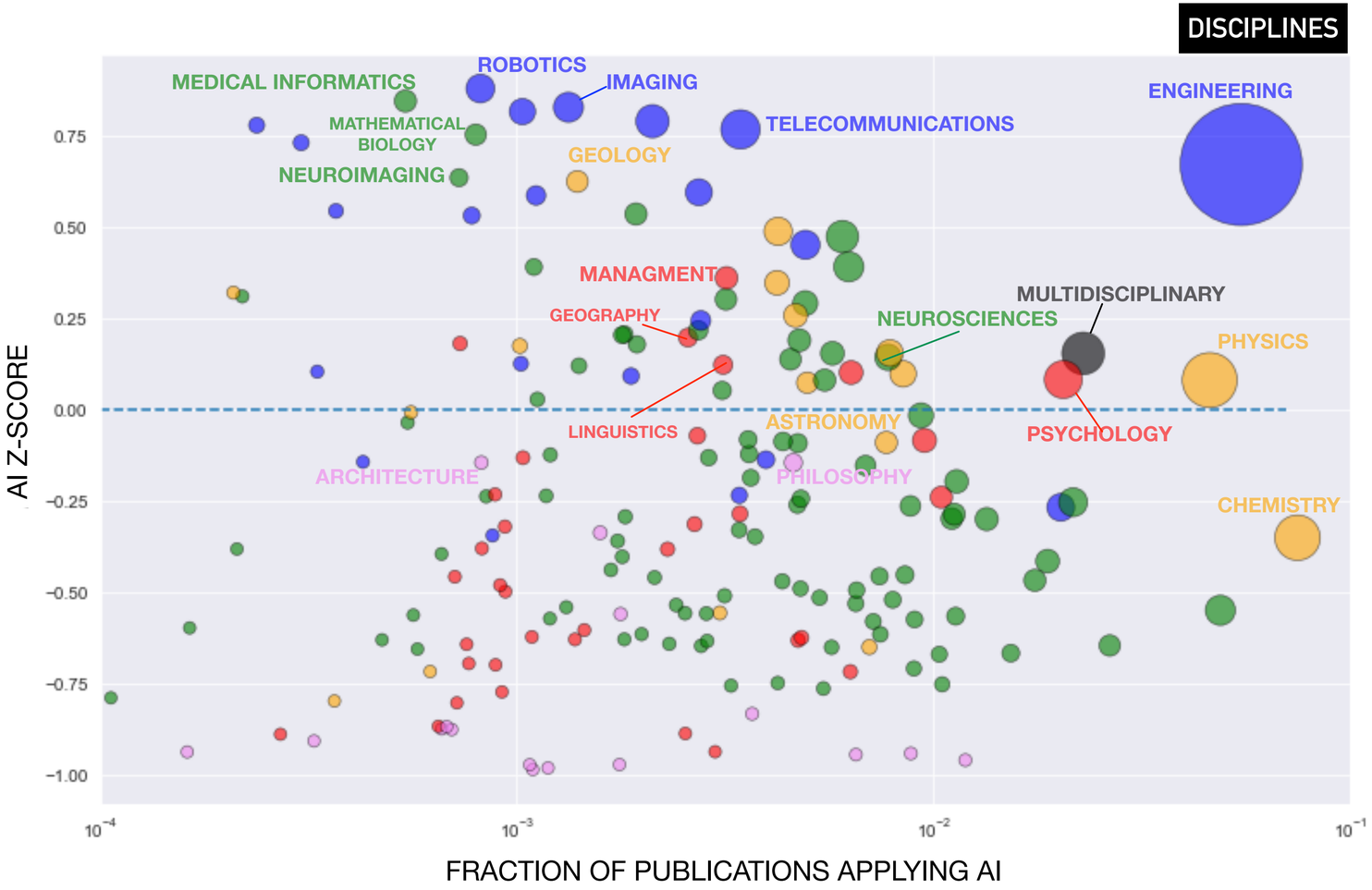
DIFFUSION PHASE:
AI diffuses from originating disciplines to application domains.

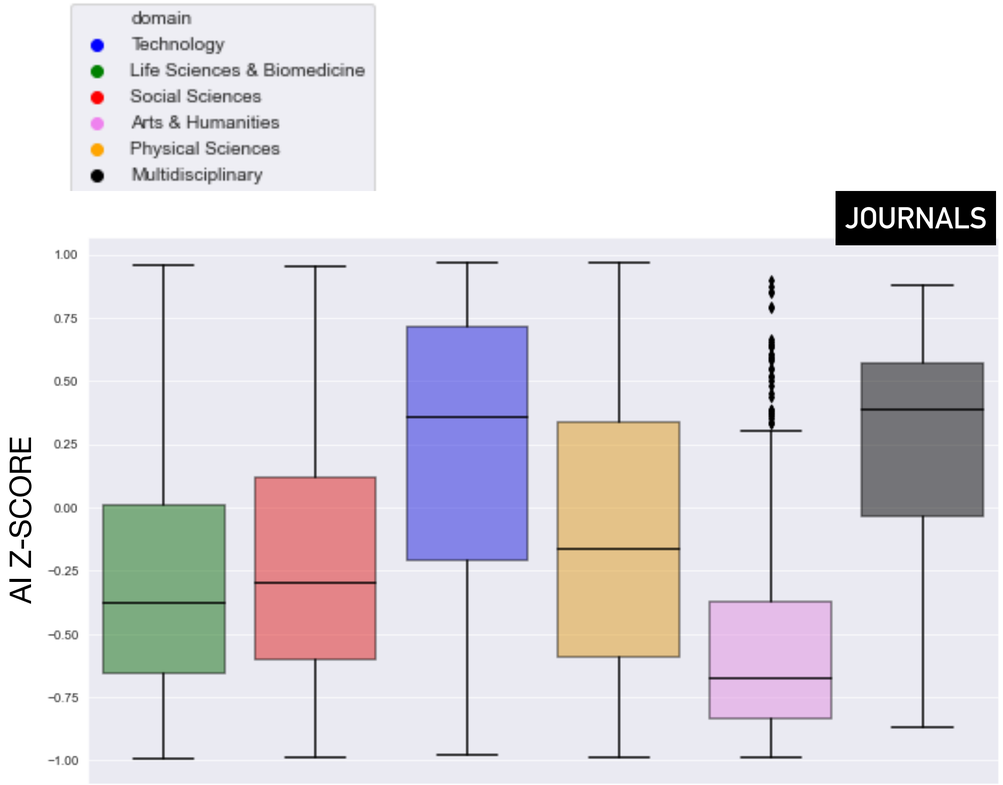
AI from development to application...



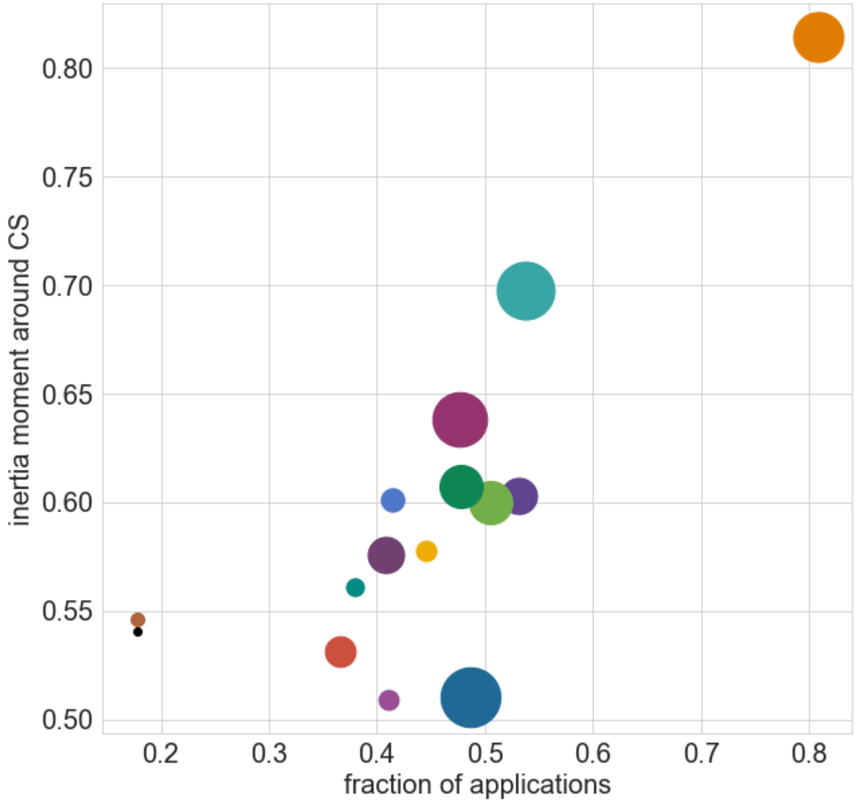
Superlinear growth of applications with development papers.

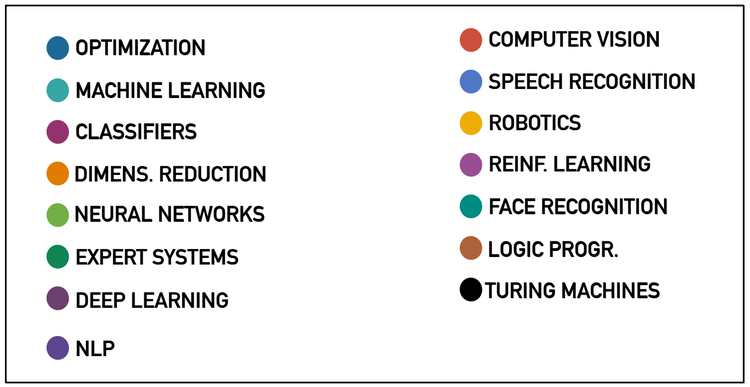
In which disciplines AI has diffused?

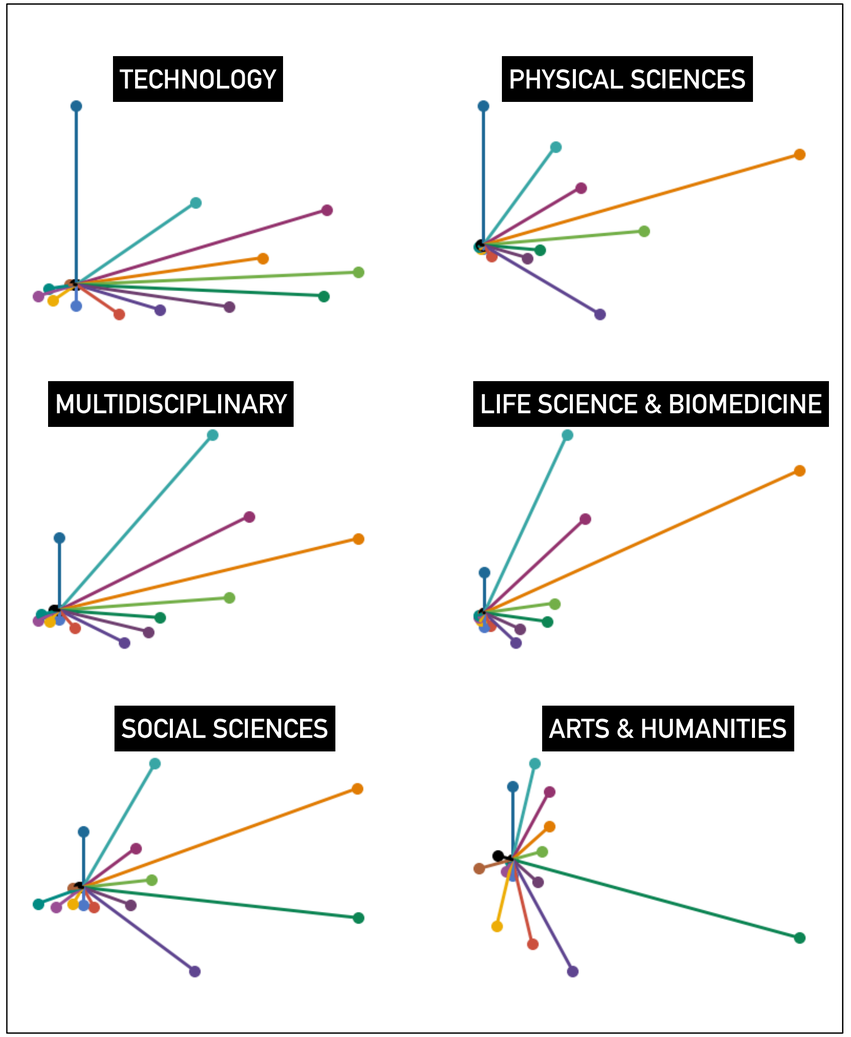




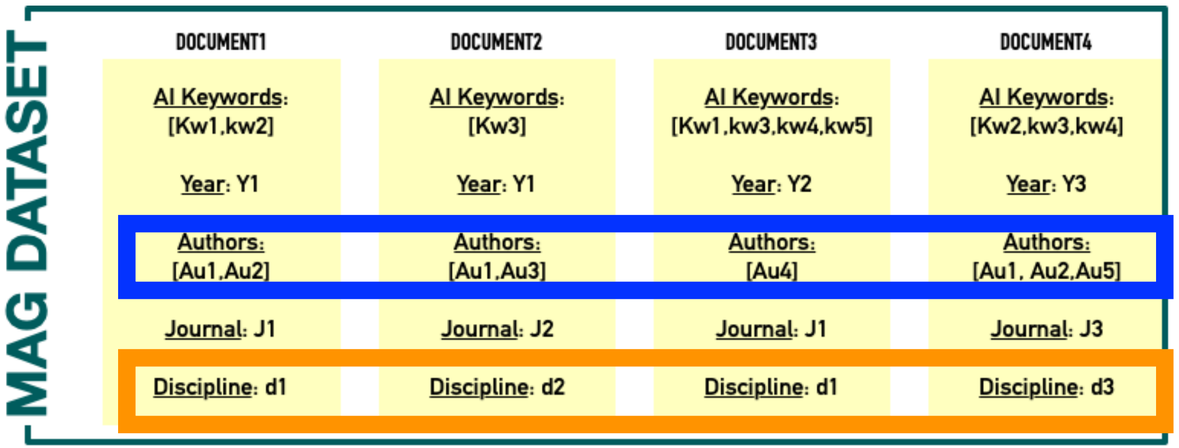
In which disciplines AI has diffused?



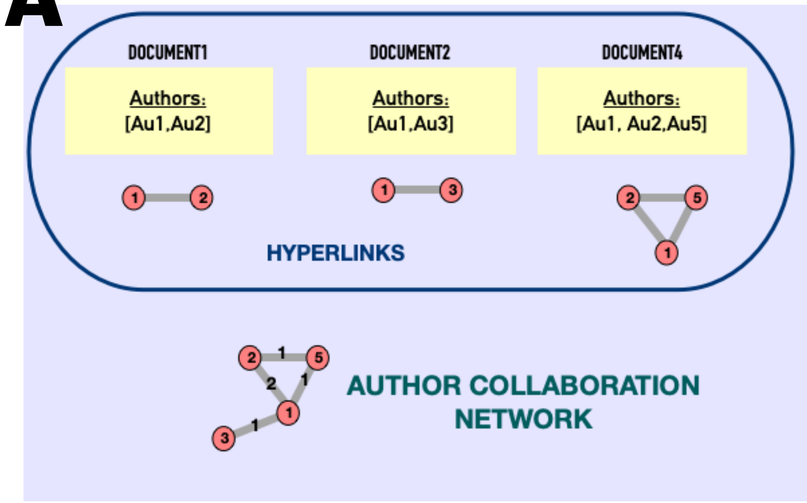




Interdisciplinary collaborations around AI



A



B

For author year we take the list of disciplines in which he/she published

Au1: {d1:4, d2:6,...}
Au2: {d1: 1, d2:6,...}
Au3: {d1: 10, d2:0,...}
....

Author interdisciplinary score:
Fraction of papers of the author in AI originating disciplines (Mathematics and CS)

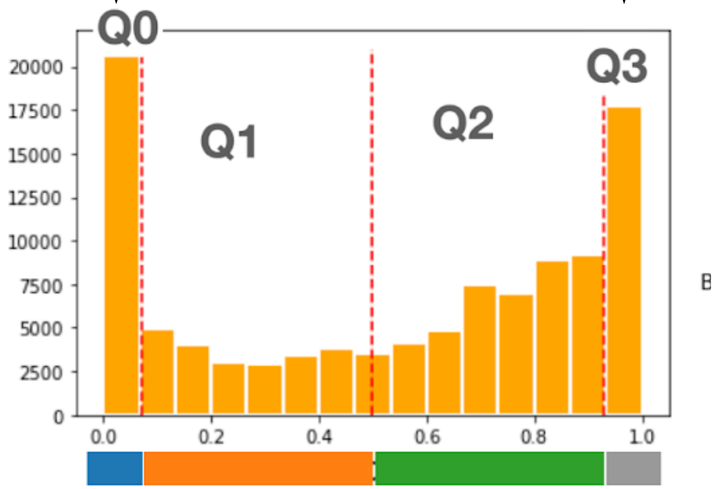
Interdisciplinary collaborations around AI

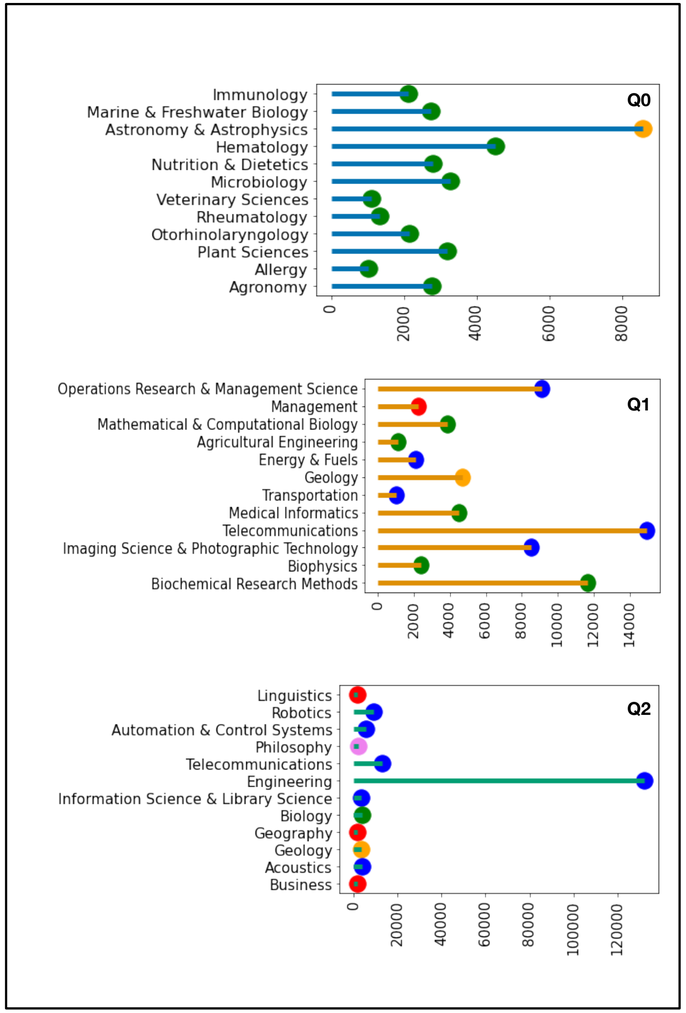
QUARTILES OF AUTHORS INTERDISCIPLINARY SCORE

Most of the authors are pure developers (only publicate in CS or Math) or pure applicators (never publish in CS or Math)

Authors NEVER publishing in AI originating disciplines

Authors ONLY publishing in AI originating disciplines

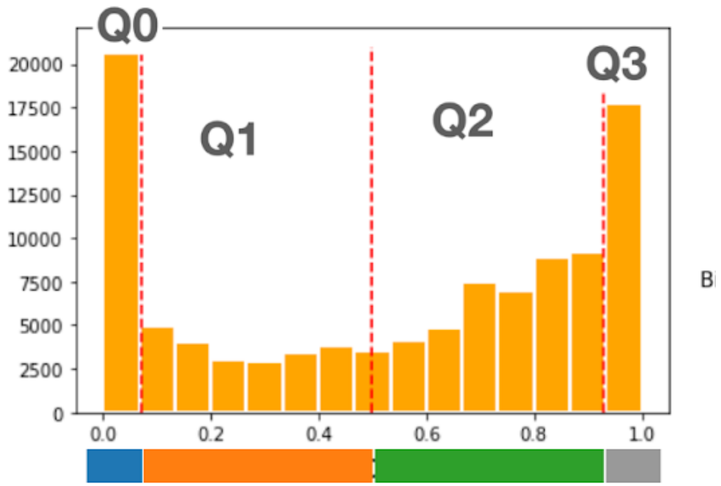


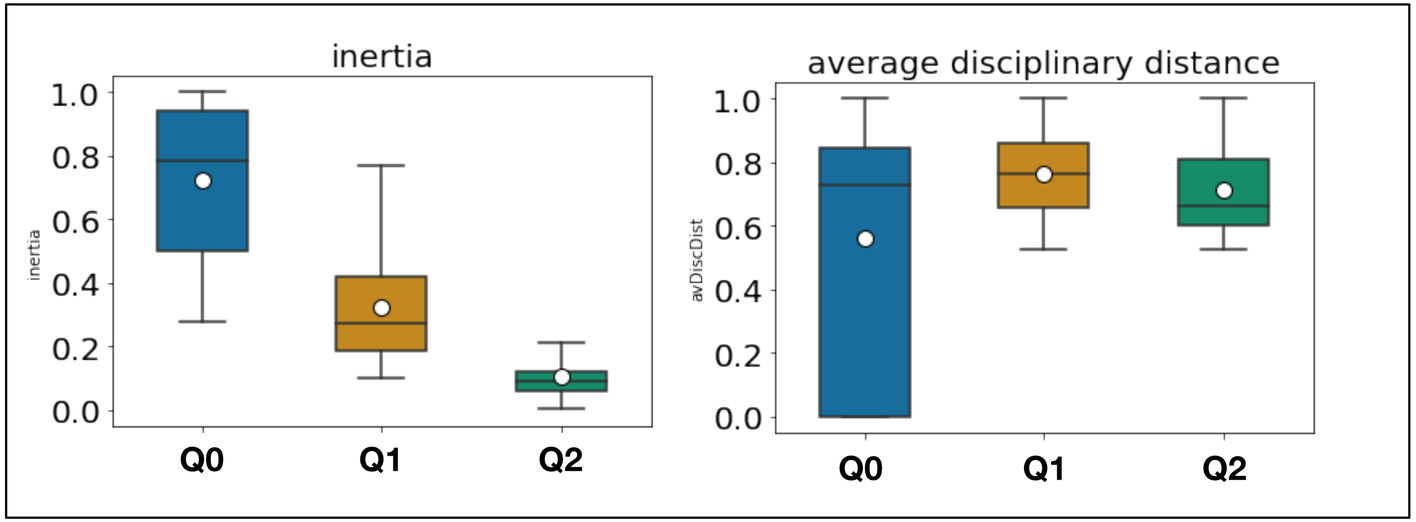


Interdisciplinary collaborations around AI

Authors NEVER publishing in AI originating disciplines

Authors ONLY publishing in AI originating disciplines



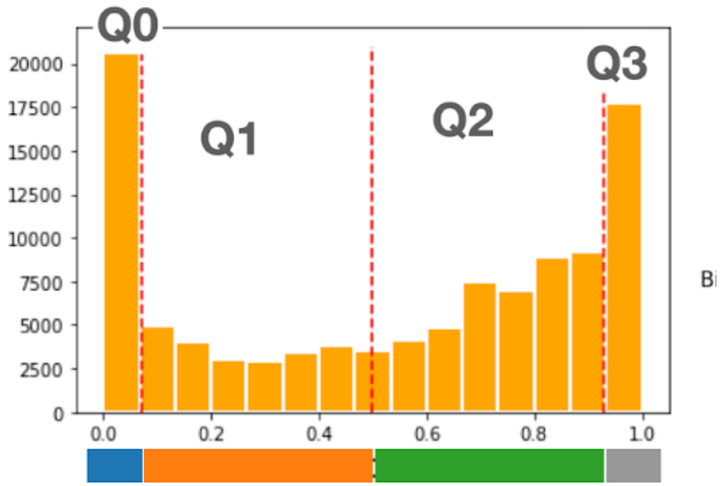


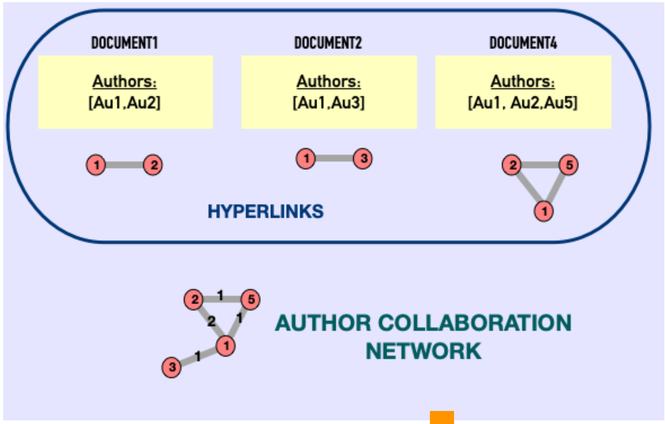
Authors in **Q0** have the highest moment of inertia around native disciplines but at the same time have a low global level of interdisciplinarity: they interact with few disciplines, quite close to each other.

Q1 includes most authors involved in interdisciplinary collaborations.

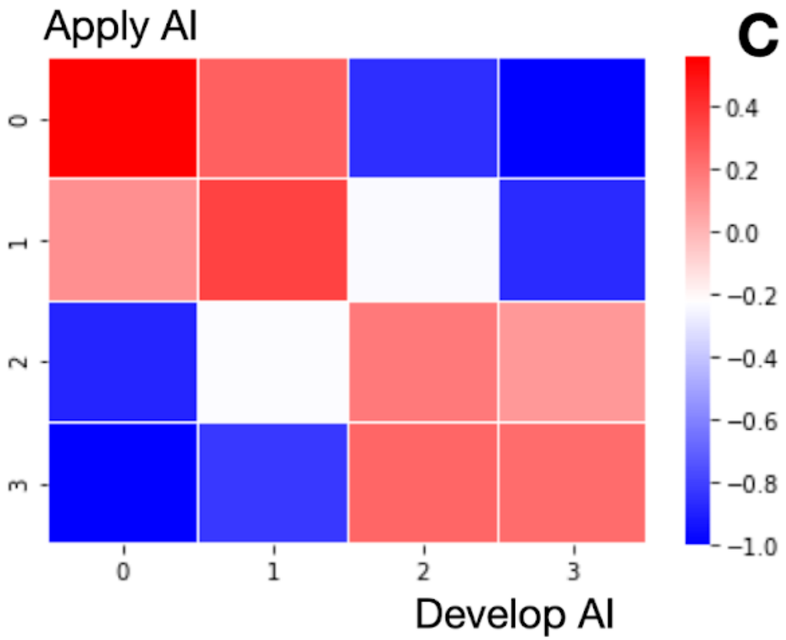
Q2 authors publish in disciplines close to the native ones (low moment of inertia) but at a quite large distance from them.

Interdisciplinary collaborations around AI





1. Aggregated at the level of AI score groups
2. Compared with multinomial expectations



Few direct collaborations.
Q1 and Q2 gradually make the passage between originating and applying disciplines

Main findings

- AI is a set of 15 interconnected sub-categories of concepts each with its different historical pattern.
- 3 phases of AI diffusion in disciplines: large participation to foundation, shrink on CS, spread in applications.
- Direct collaborations between developers and users are extremely rare.

AI in Neuroscience

The dataset

1. All the papers in OpenAlex appearing in journals labeled as Neuroscience in Web of science

2. Papers citing and cited by the Neuroscience corpus

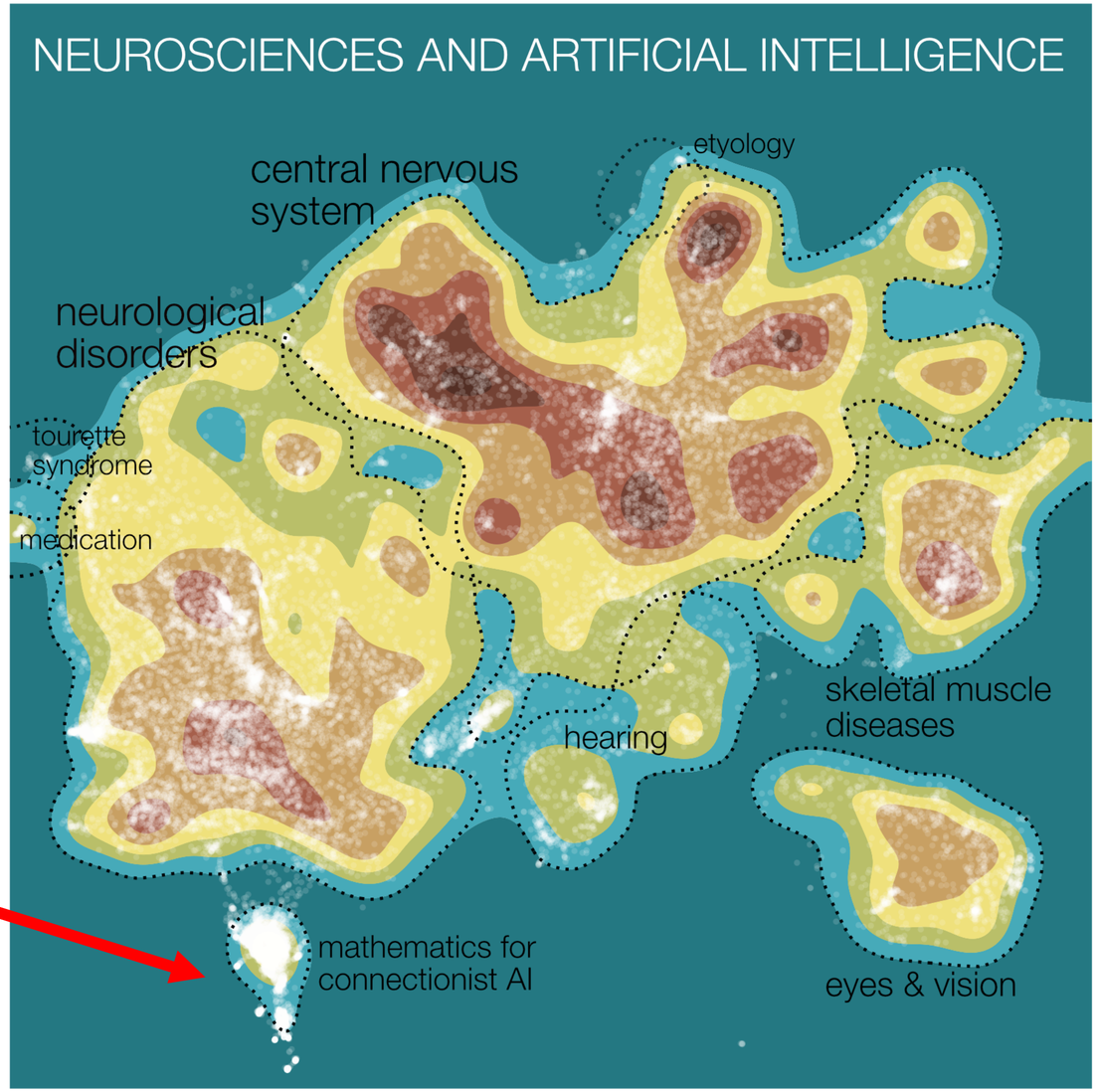
3. Two subsets:
   - papers with AI keywords (applying AI in neuroscience)
   - Papers without AI keywords

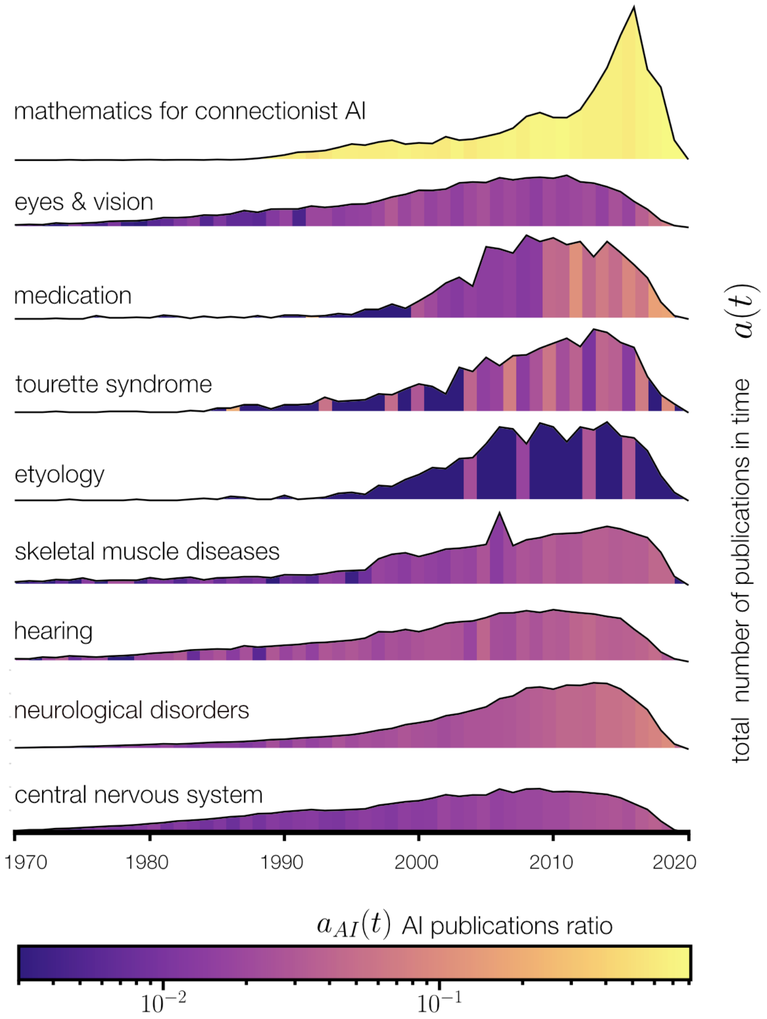
How to build the disciplinary landscape of AI?

- Contextual words embedding of each paper of our dataset using Specter (an embedding model based on SciBert) (https://github.com/allenai/specter)
- Reduce the data if needed for a first 2D spatialized visualization with UMAP
- Use HDBSCAN for local density-based clustering, which determines automatically the number of clusters (points that does not belong to particular clusters are drawn in the same color)
- Characterize the founded clusters with concepts of papers from MAG concepts

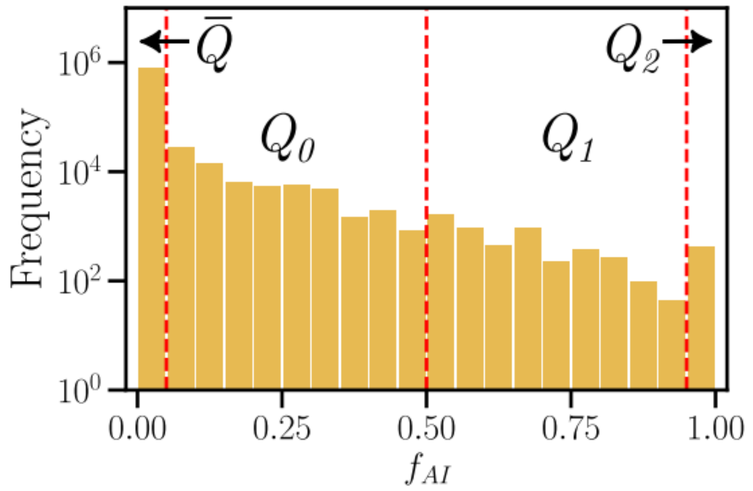
AI in Neuroscience

Concentration on the cluster "Math for AI" (Computer Science, Math and Stats) and on the vocabulary space next to it in "Neurological disorders" → **linked to the data deluge!**





AI in Neuroscience



Main findings

- Epistemic (partial) integration of AI in Neuroscience
- Social segregarion of authors practicing AI in Neuroscience