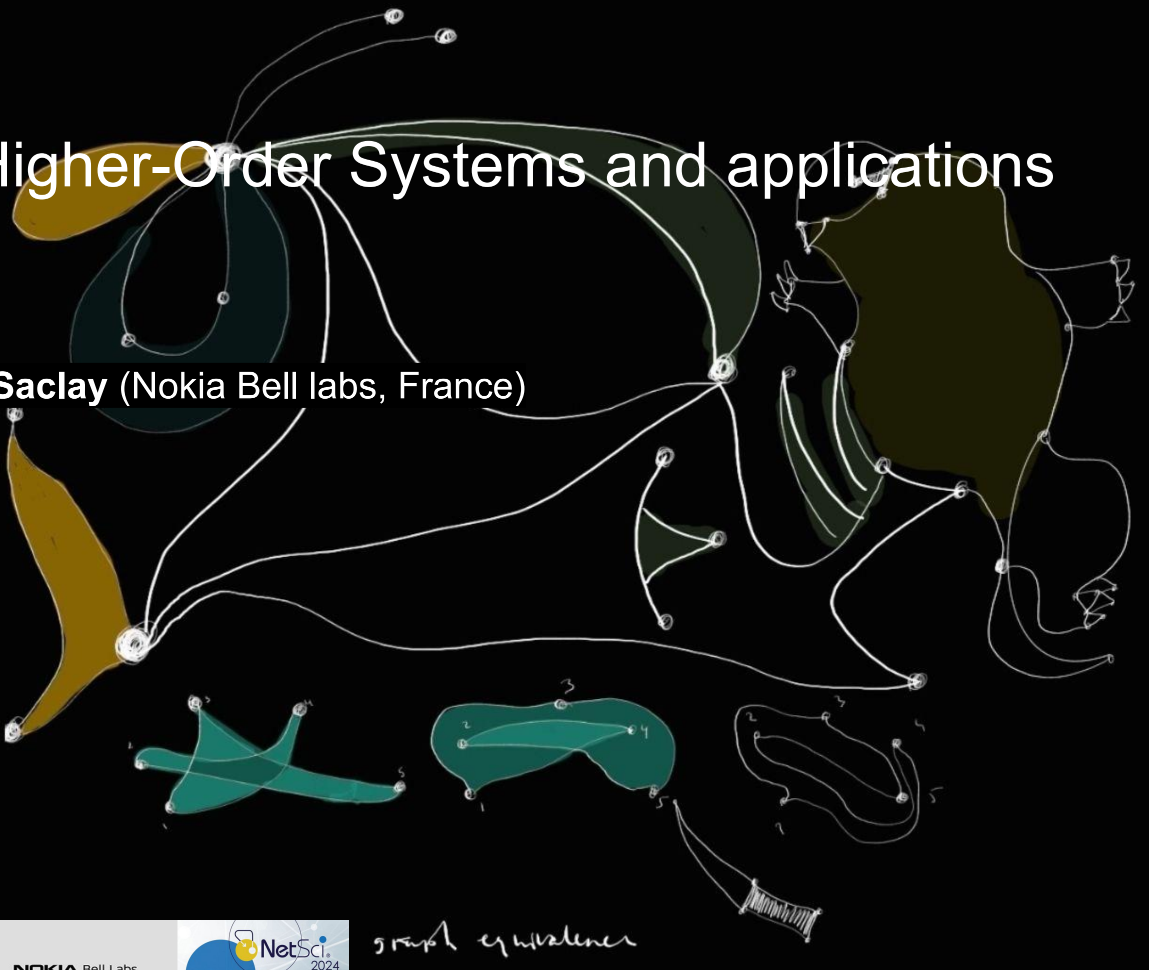
Methods for modeling Higher-Order Systems and applications to knowledge graphs

Liubov Tupikina,
Machine Learning & Systems, Paris-Saclay (Nokia Bell labs, France)
Paris Descartes, LPI, Paris, France



Thanks to NetSci community

- Science of science analysis collaboration with Marc (LPI), Hritika (Bell labs)
- CUDAN lab Estonia, conference 2023
- Hypernetwork workshop 2024
  Wolfram Institute

See also talk by Carlos Zapata, main session, NetSci 2024, Friday

How to analyze texts and knowledge landscapes?

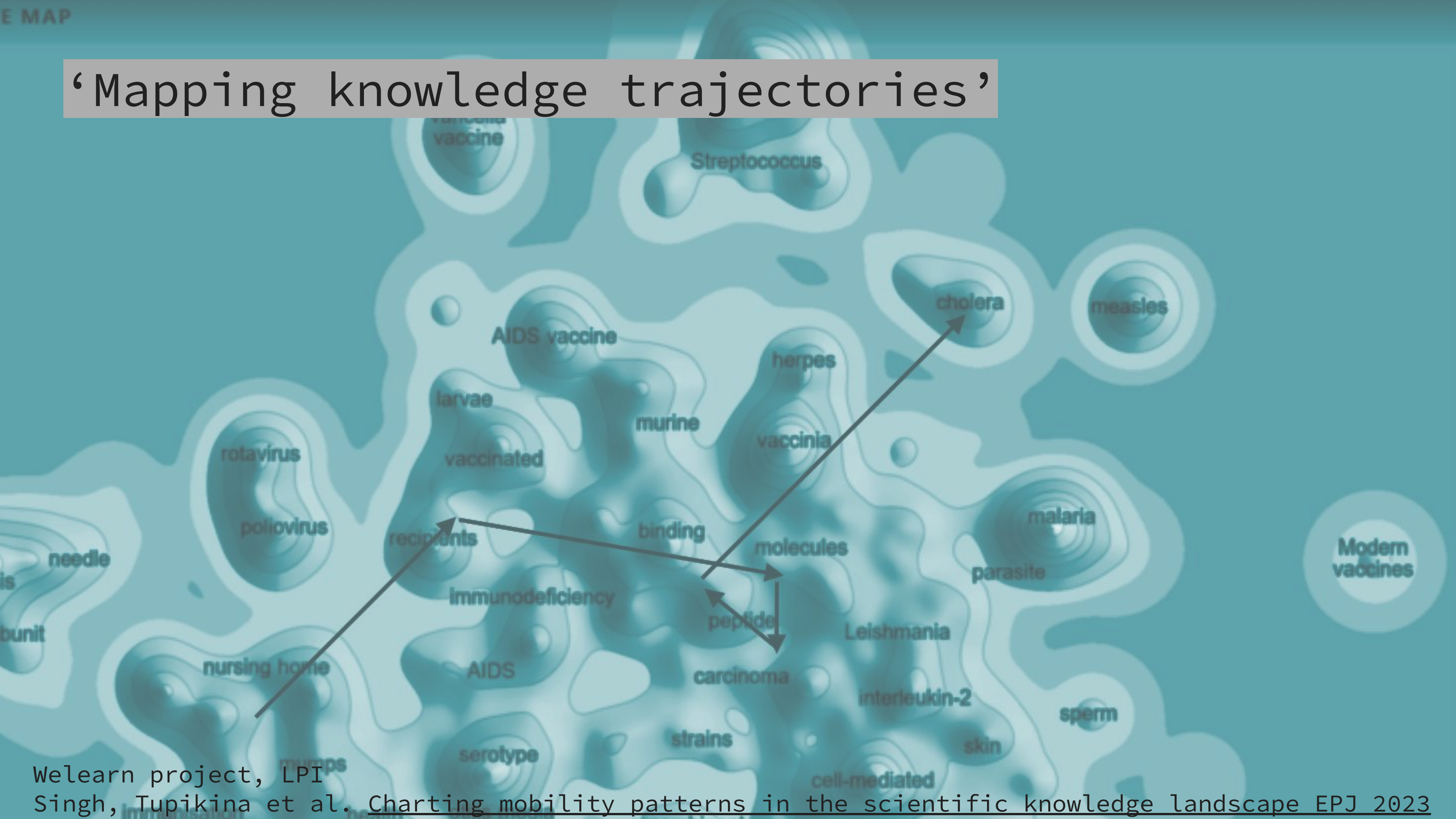
How to analyze texts and knowledge landscapes?

In this talk:

Arxiv dataset-scaping
Rise and fall of scientific field

Higher-order structure for analysis of knowledge hypergraphs
Non-associative algebraic structures

‘Mapping knowledge trajectories’



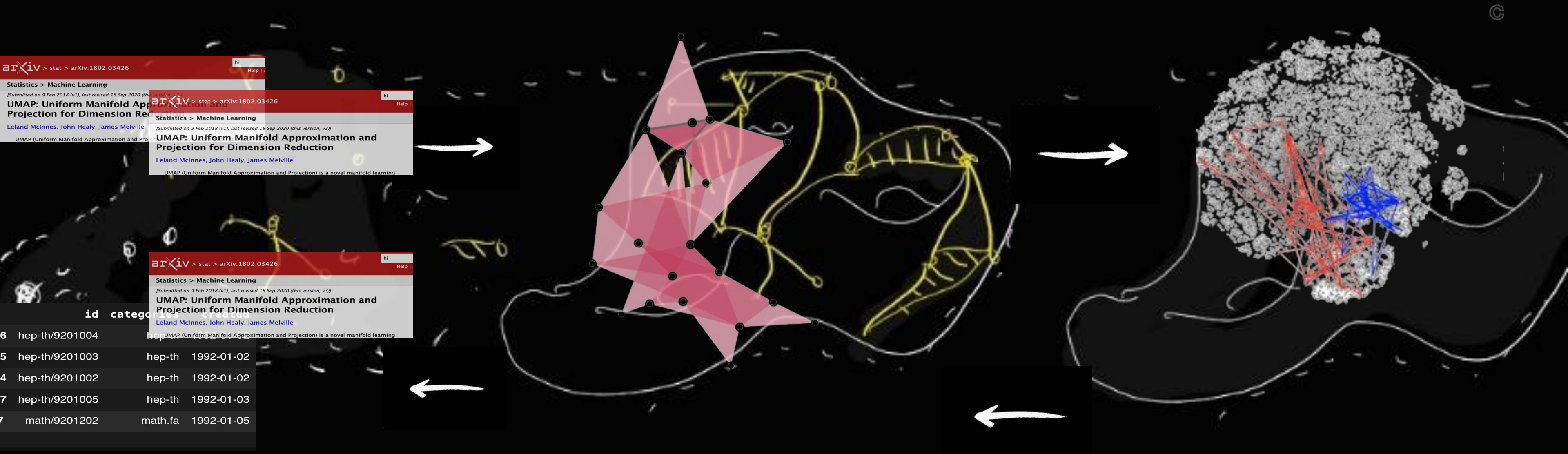
Welearn project, LPI

Singh, Tupikina et al. Charting mobility patterns in the scientific knowledge landscape EPJ 2023

General approach:

Data

Data Hypergraphs Manifolds



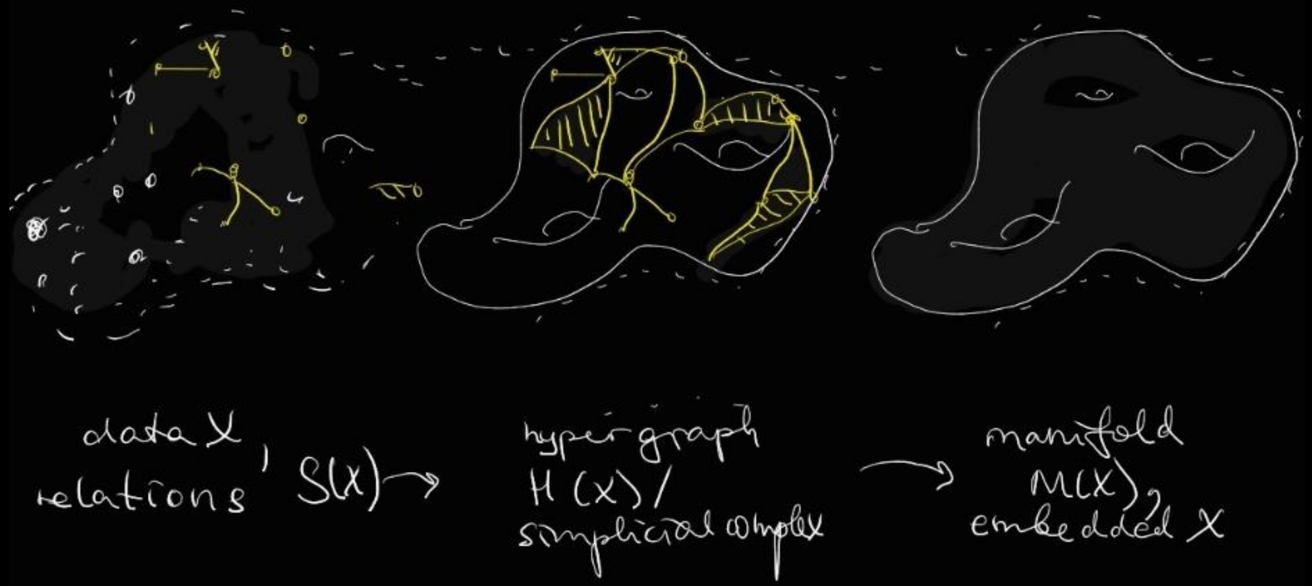
Topological data analysis
[Bernstein, 2020, Carlson 2009]

Manifold learning, hypergraph embedding
[UMAP, 2020, Singh, Tupikina 2009]

Main question:

How to use higher-order theory for data analysis, embeddings theory for high dimensional data representation?

Data Hypergraphs Manifolds



Main question:

How to use higher-order theory for data analysis, embeddings theory for high dimensional data representation?

Research | Open access | Published: 20 February 2024

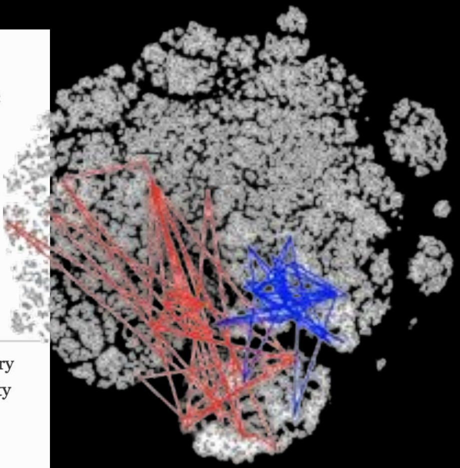
Charting mobility patterns in the scientific knowledge landscape

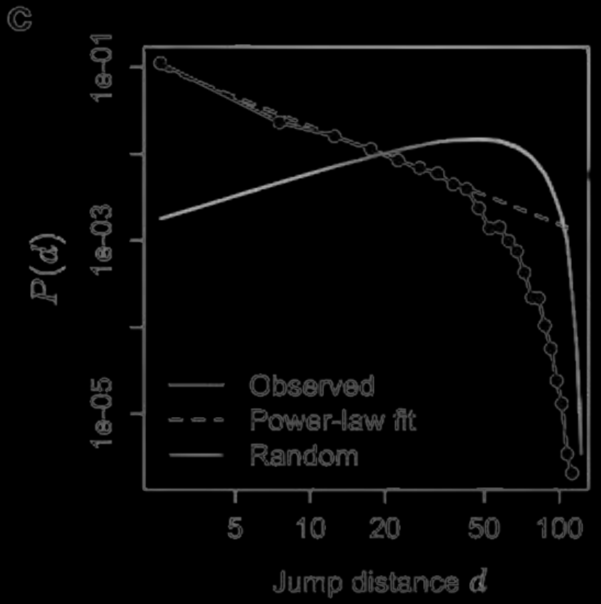
Chakresh Kumar Singh, Liubov Tupikina, Fabrice Lécuyer, Michele Starnini & Marc Santolini

EPJ Data Science **13**, Article number: 12 (2024) | Cite this article

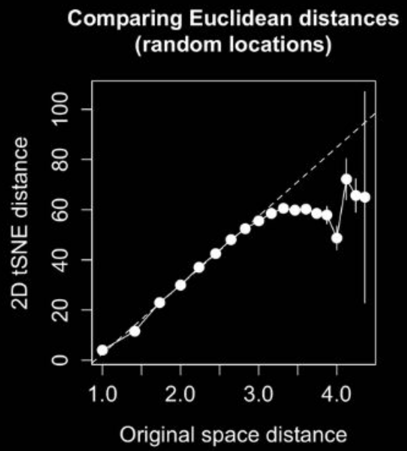
Abstract

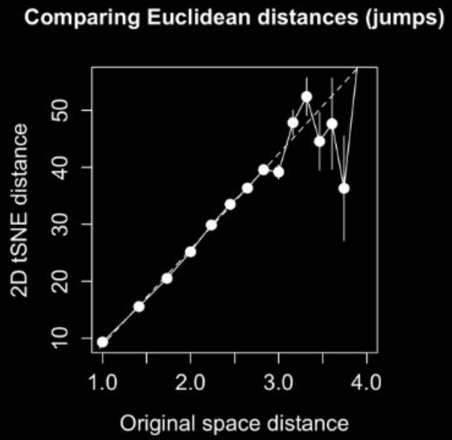
From small steps to great leaps, metaphors of spatial mobility abound to describe discovery processes. Here, we ground these ideas in formal terms by systematically studying mobility patterns in the scientific knowledge landscape. We use low-dimensional embedding techniques to create a knowledge space made up of 1.5 million articles from the fields of physics, computer science, and mathematics. By analyzing the publication histories of individual researchers, we discover patterns of scientific mobility that closely resemble physical mobility. In aggregate, the trajectories form mobility flows that can be described by a

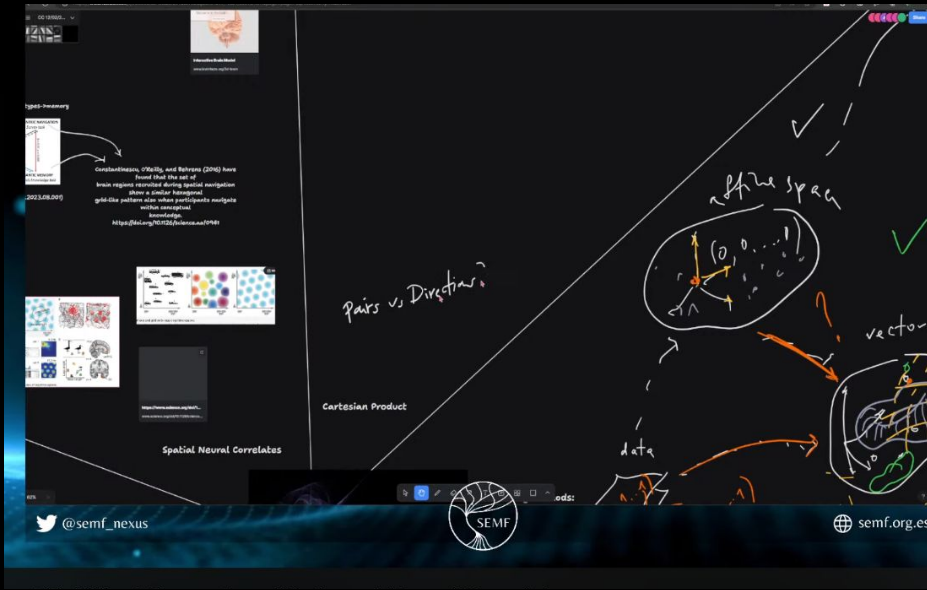




Reviewers questions

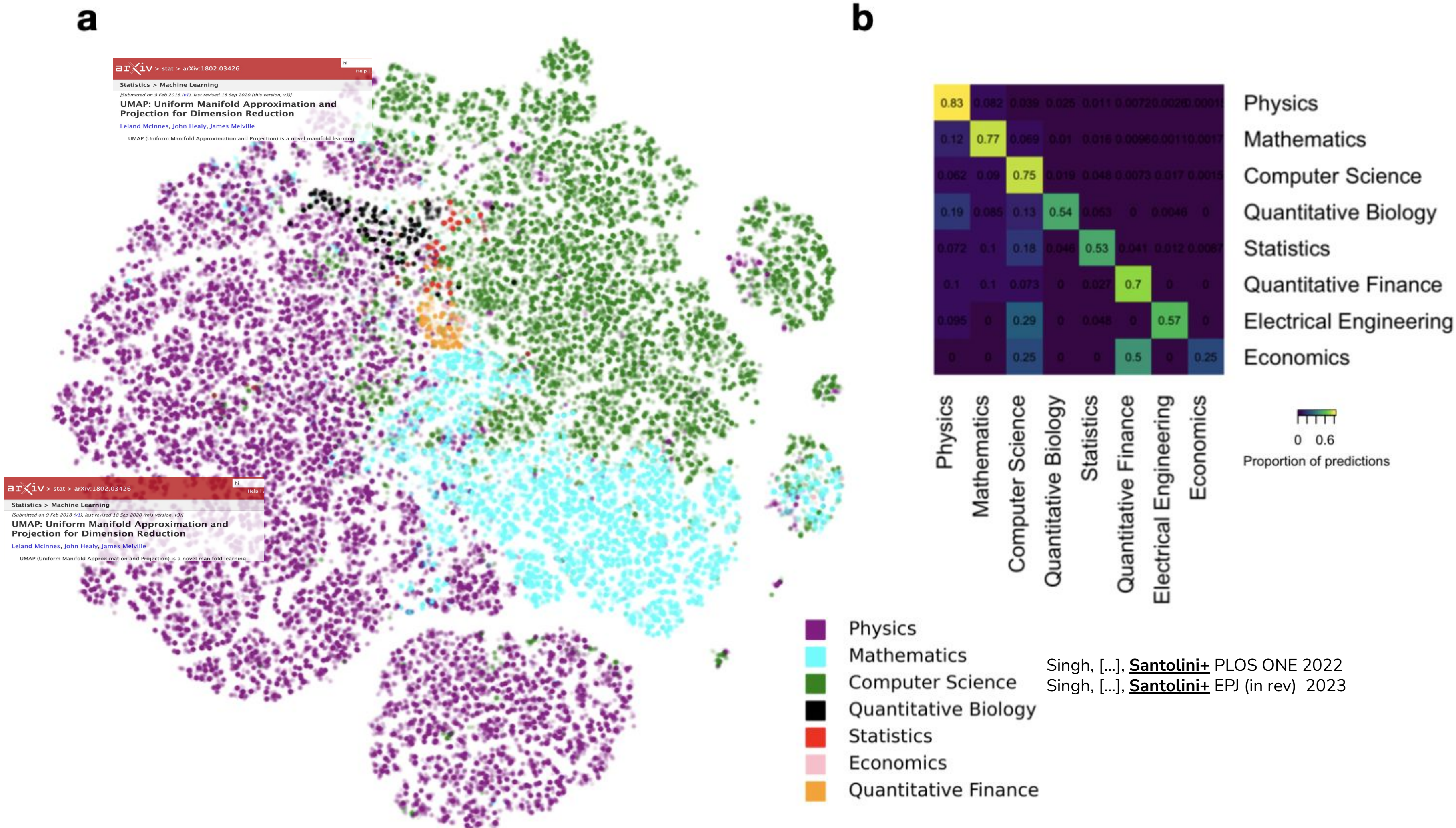






Singh, Tupikina et al.
[Charting mobility patterns in the scientific knowledge landscape 2023]

If you are interested to discuss this, send us a message:
[liubov.tupikina@cri-paris.org]



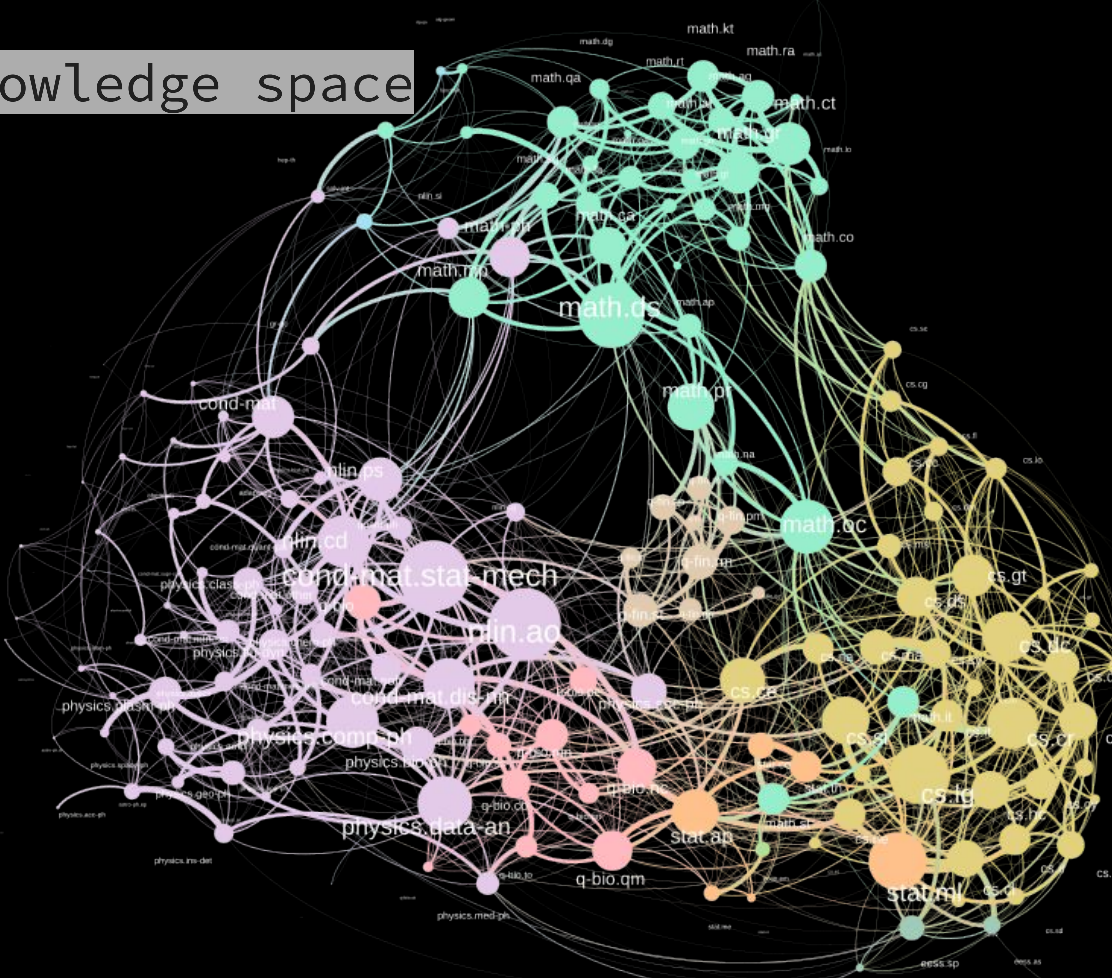
Singh, [...], **Santolini+** PLOS ONE 2022
Singh, [...], **Santolini+** EPJ (in rev) 2023

Binary reconstructing knowledge space

$$p_{ij} = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

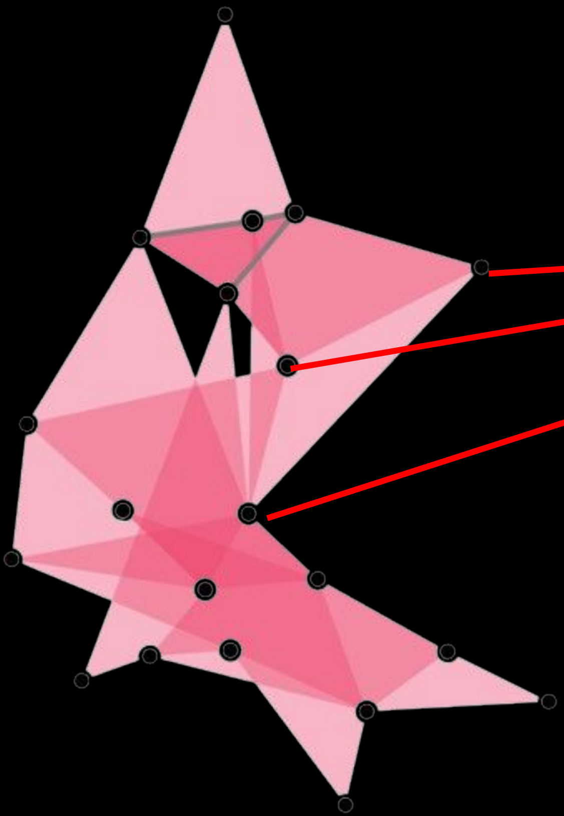
Edge Weight = $-log_{10}(p_{ij})$

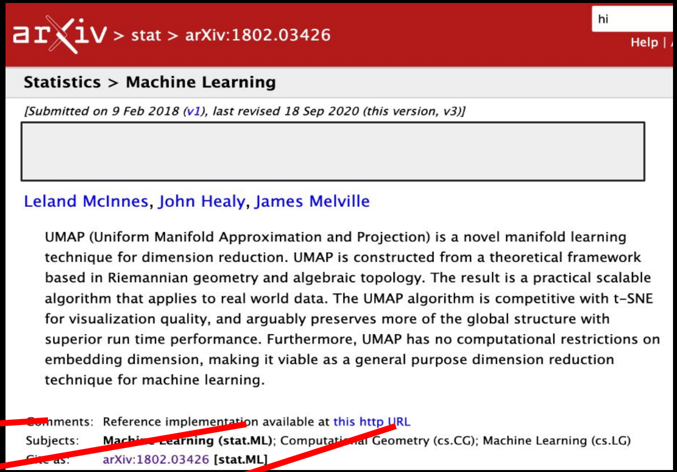
N - total articles, K - articles in field i, n - articles in field j, k - common articles bw i, j



Co-tags network, Singh et al. ‘Quantifying the rise and fall of scientific fields’ Plos One 2022

Higher-order reconstructions of knowledge space from arxiv data





arXiv > stat > arXiv:1802.03426

Help |

Statistics > Machine Learning

[Submitted on 9 Feb 2018 (v1), last revised 18 Sep 2020 (this version, v3)]

Leland McInnes, John Healy, James Melville

UMAP (Uniform Manifold Approximation and Projection) is a novel manifold learning technique for dimension reduction. UMAP is constructed from a theoretical framework based in Riemannian geometry and algebraic topology. The result is a practical scalable algorithm that applies to real world data. The UMAP algorithm is competitive with t-SNE for visualization quality, and arguably preserves more of the global structure with superior run time performance. Furthermore, UMAP has no computational restrictions on embedding dimension, making it viable as a general purpose dimension reduction technique for machine learning.

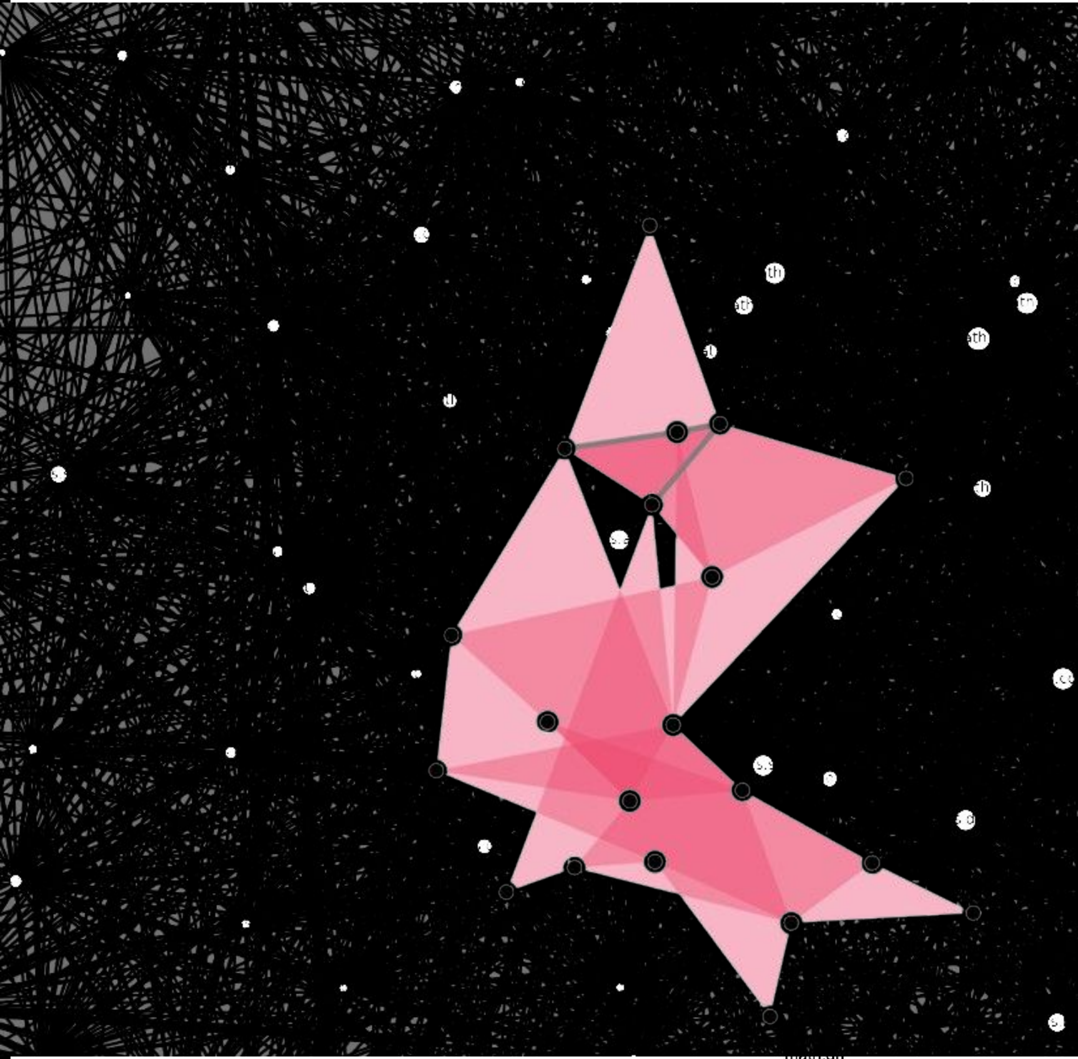
Comments: Reference implementation available at this http URL

Subjects: Machine Learning (stat.ML); Computational Geometry (cs.CG); Machine Learning (cs.LG)

Cite as: arXiv:1802.03426 [stat.ML]

Hypergraph of co-tags is generalisation of co-tags network.
'Quantifying the rise and fall of scientific fields' Plos One 2022

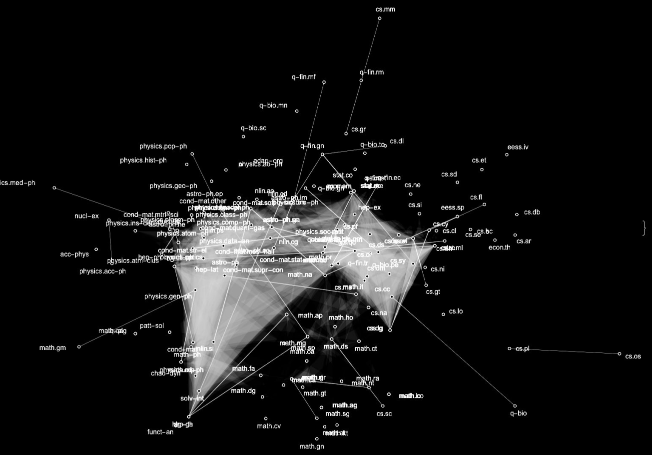
Nodes = fields
Hyperedges = papers

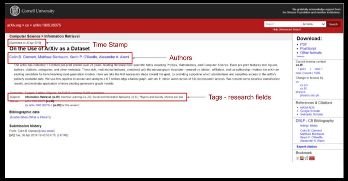


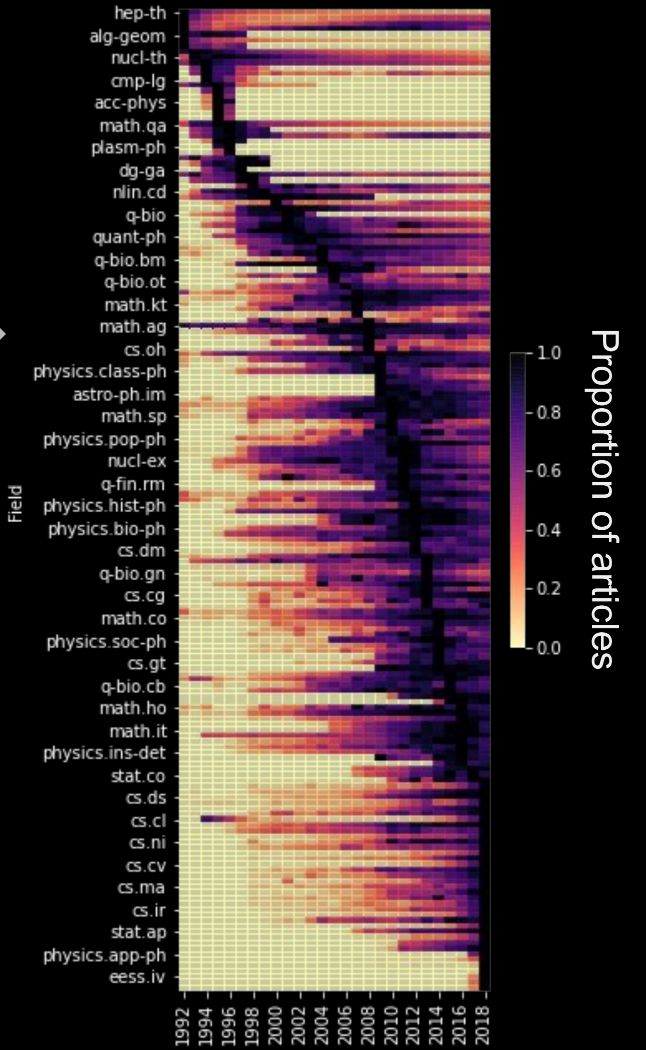
Hypergraph of co-tags is generalisation of co-tags network.
'Quantifying the rise and fall of scientific fields' Plos One 2022

Nodes = fields
Hyperedges = papers

Collective level: Rise and Fall of Research Fields



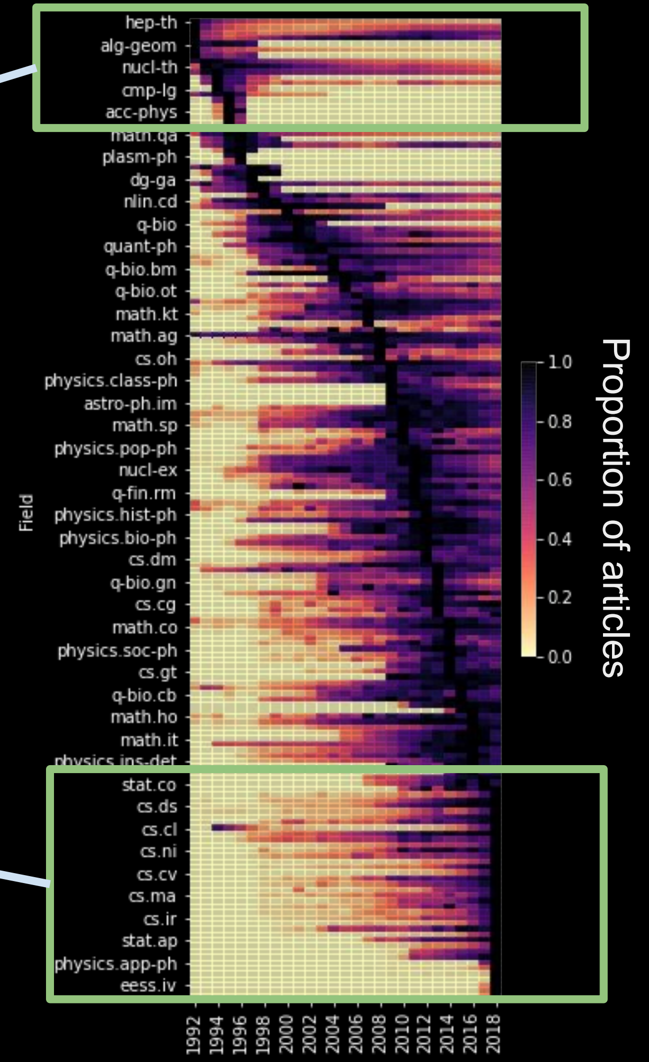




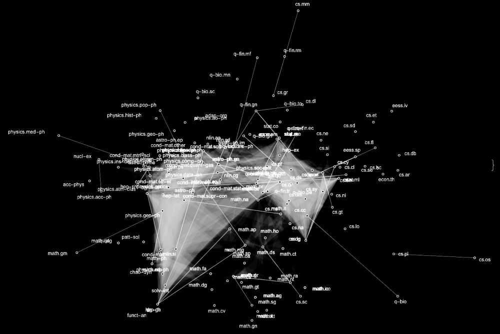
Singh (...) Tupikina, Santolini **Quantifying the rise and fall of scientific fields**", Plos One (2022)

High Energy Physics, Accelerator Physics, Nuclear Physics, Algebraic Geometry

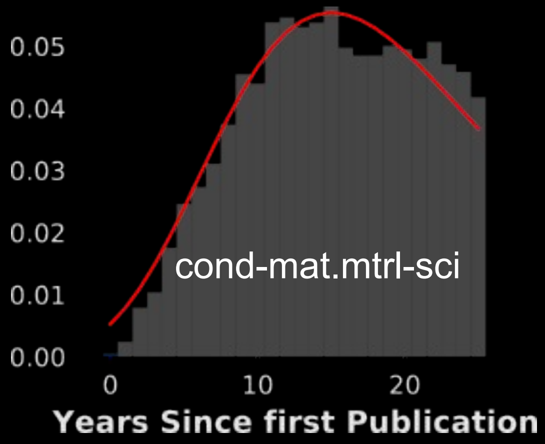
AI, Computation and Language, Vision and Pattern Recognition, Data Structures, Computation, Applications, Applied Physics



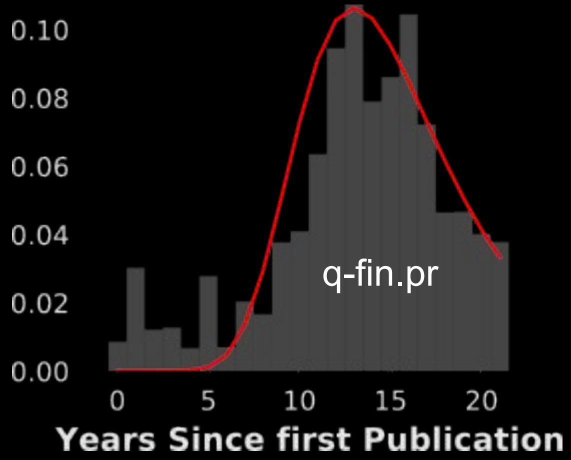
Rise and fall of scientific fields



$$G = \frac{1}{\beta} e^{\frac{-(x-\alpha)}{\beta}} e^{-e^{\frac{-(x-\alpha)}{\beta}}}$$



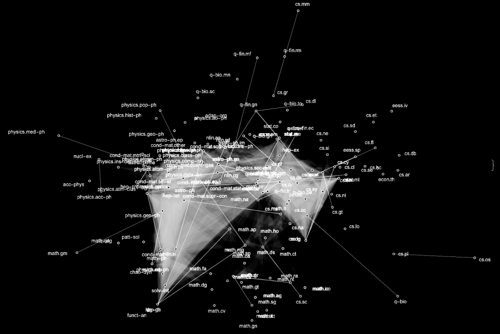
Material Sciences

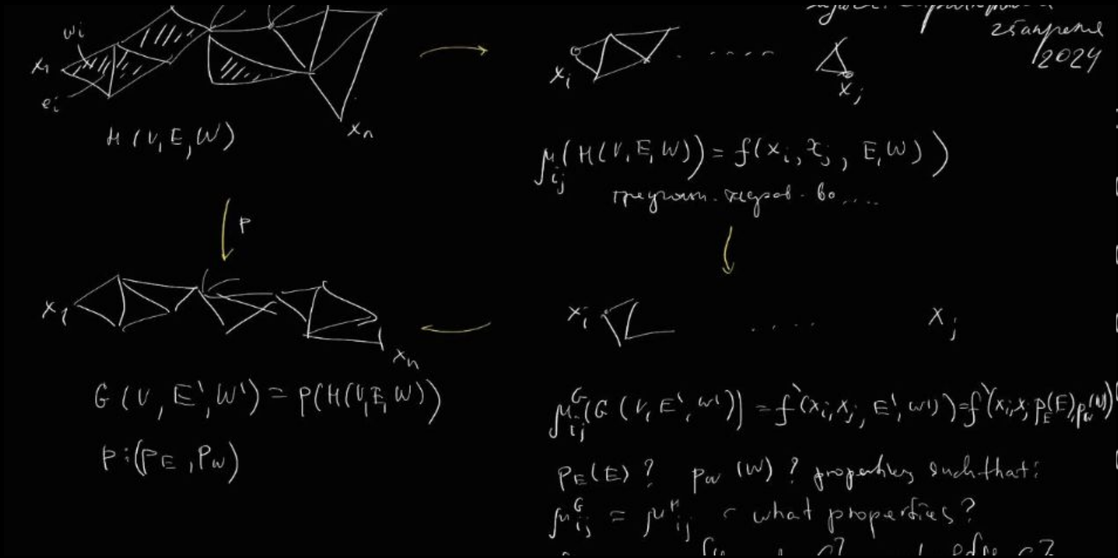


Quantitative Finance (Pricing)

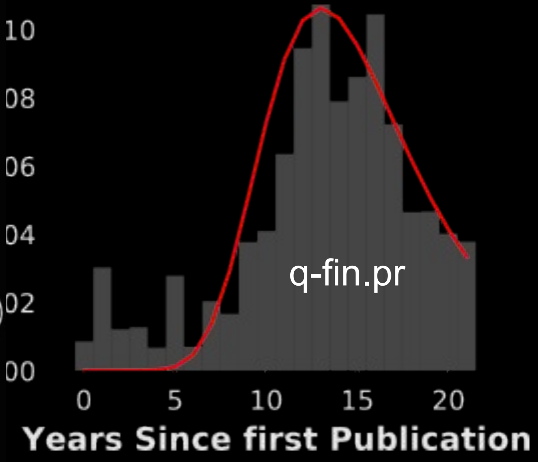
Problem of modeling the emerging weighted knowledge hypergraph:

Cognitive distance genrealisation for hypergraphs



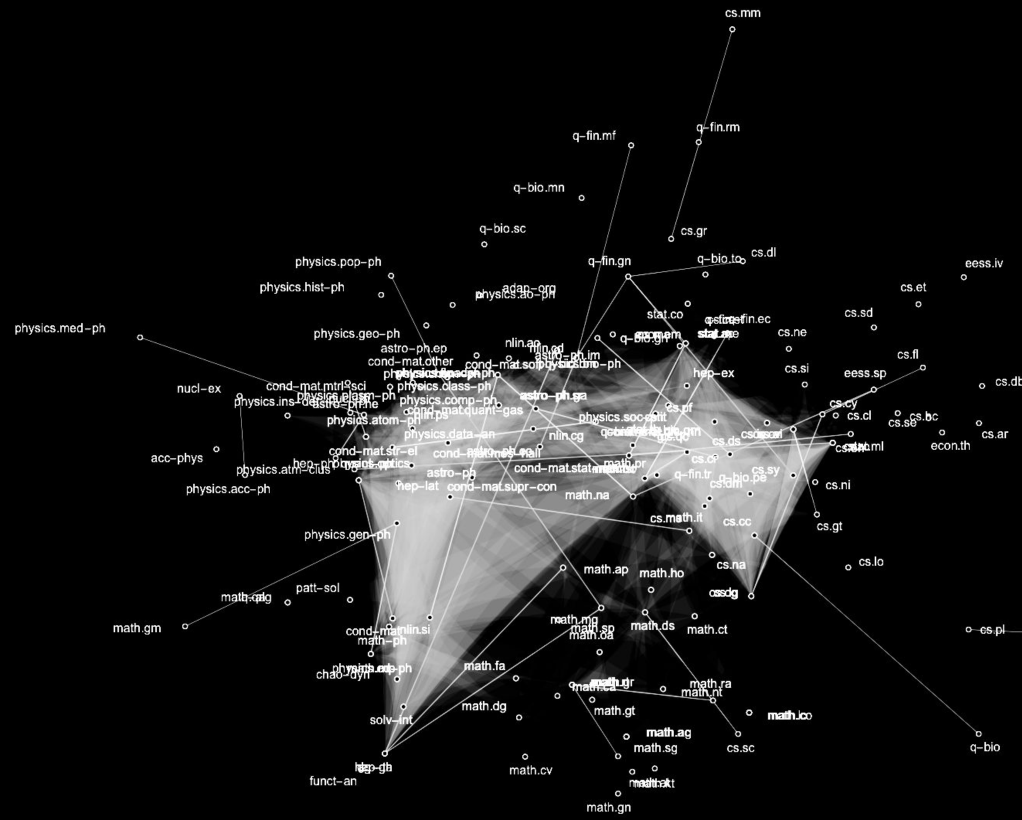


Material Sciences



Quantitative Finance (Pricing)

How to analyze knowledge hypergraphs?

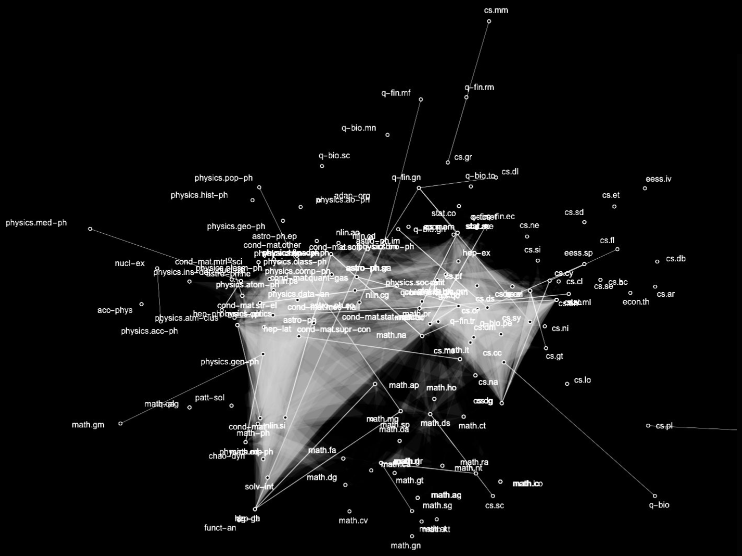


Dual representation
Possible vector centrality application area
Application with XGI development
Collaboration ongoing

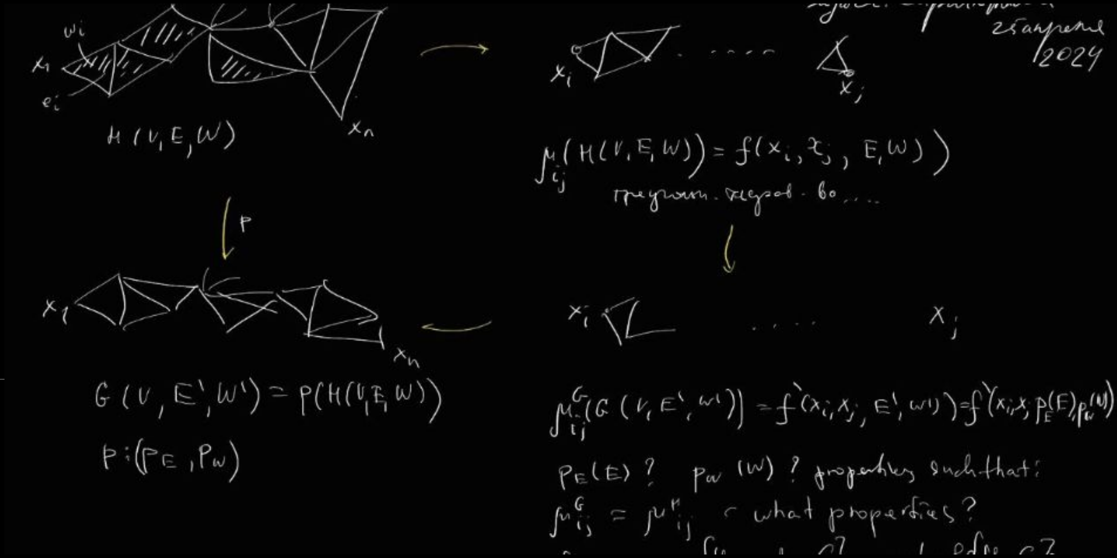
Co-tags network, Singh et al. ‘Quantifying the rise and fall of scientific fields’ Plos One 2022
Hypergraph motifs https://github.com/Liyubov/hypergraph_structures

Problem of modeling the emerging weighted knowledge hypergraph:

Cognitive distance genrealisation for hypergraphs



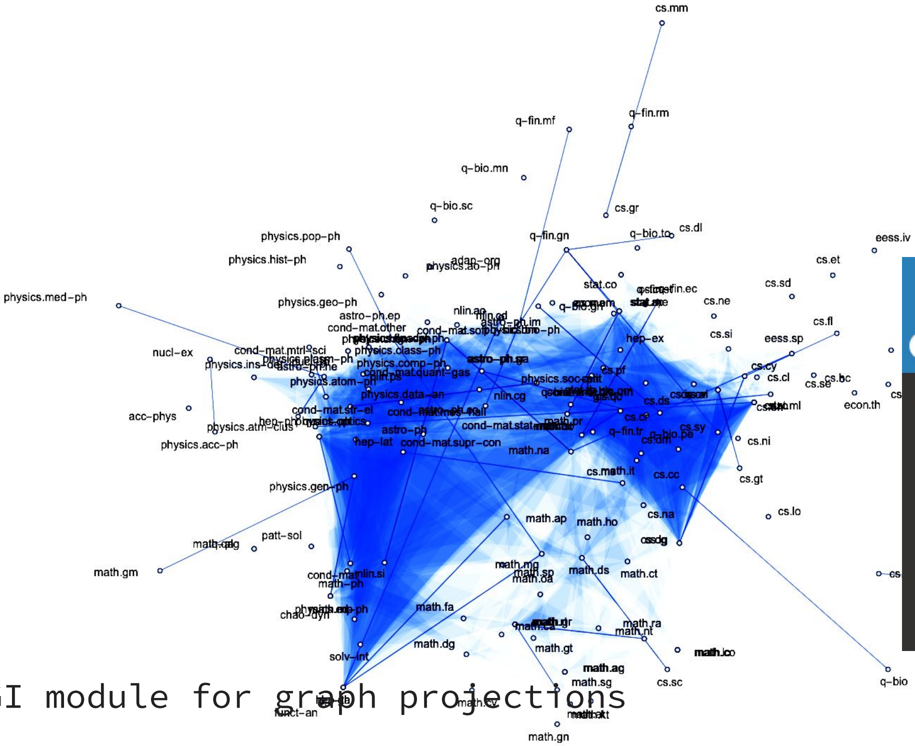
Hypergraphs



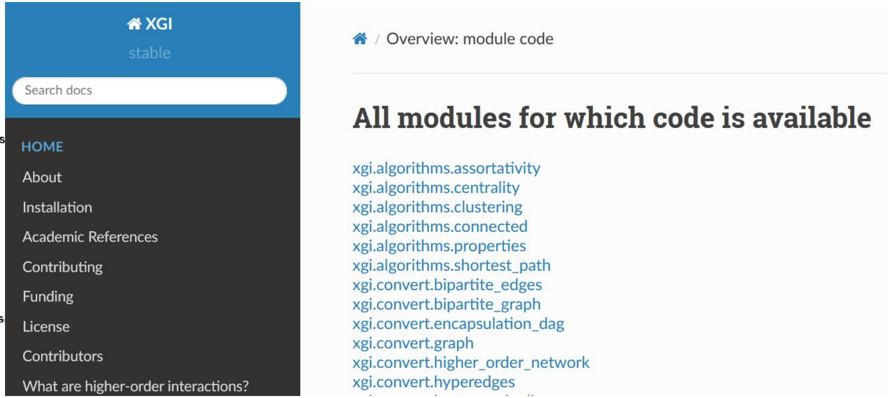
Measures for hypergraphs

Motivation to develop higher-order mathematics

Hypergraphs are often represented using lower-dimensional structures (networks), yet using the low dimensional representation removes the information from the structure.



XGI module for graph projections



Hypergraph application to knowledge space modeling

Analysis and inference of higher-order systems is still an open and challenging problem (higher arity, non-associativity):

Application of advanced mathematics on the higher order structures for data analysis:

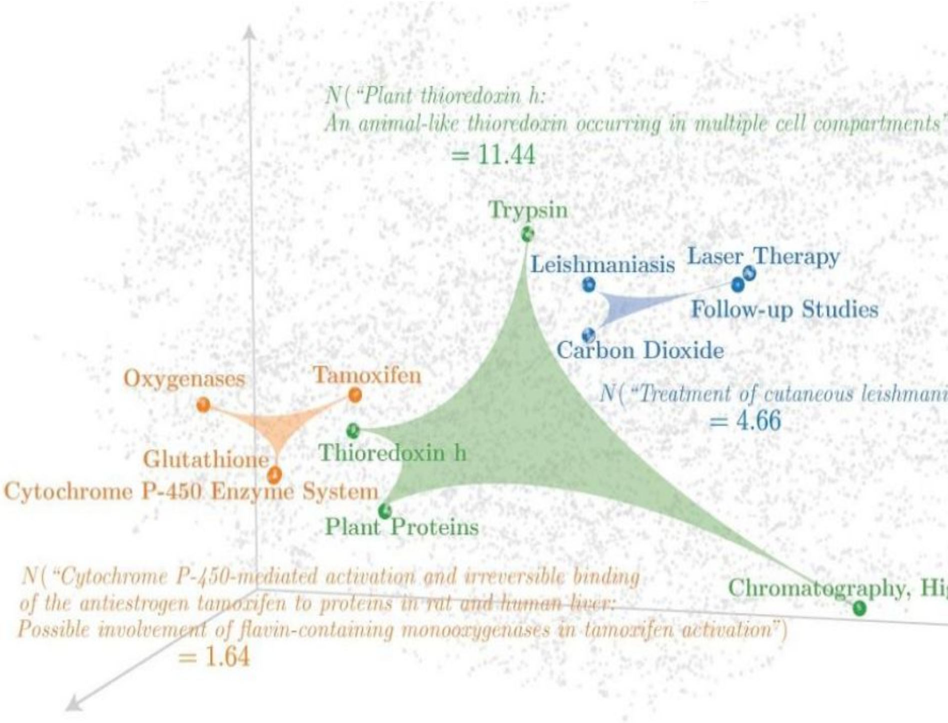
Hypergraph theory, n-arity algebras (Bretto 2013, Ghang 2020)

Computational issues to work with hypergraphs:
Combinatorial complexity, motifs counting

Evans et al. Knowledge hypergraphs 2020

The vast majority of data available on network systems contain only records of pairwise interactions, even when the underlying rules rely on higher-order patterns.

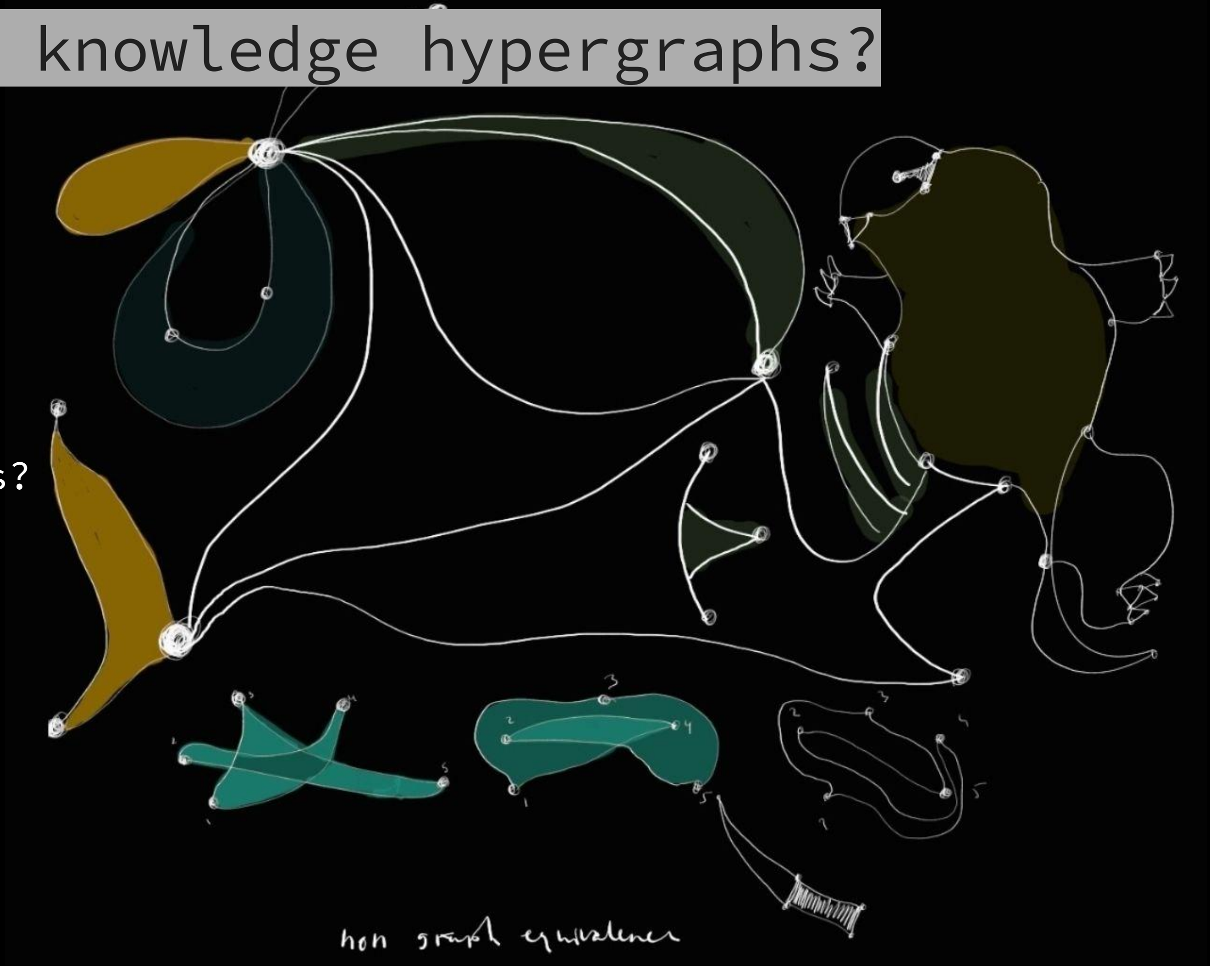
Battiston et al. Nat.Ph. (2020)



How to analyze knowledge hypergraphs?

a) how can we study higher order zoo?

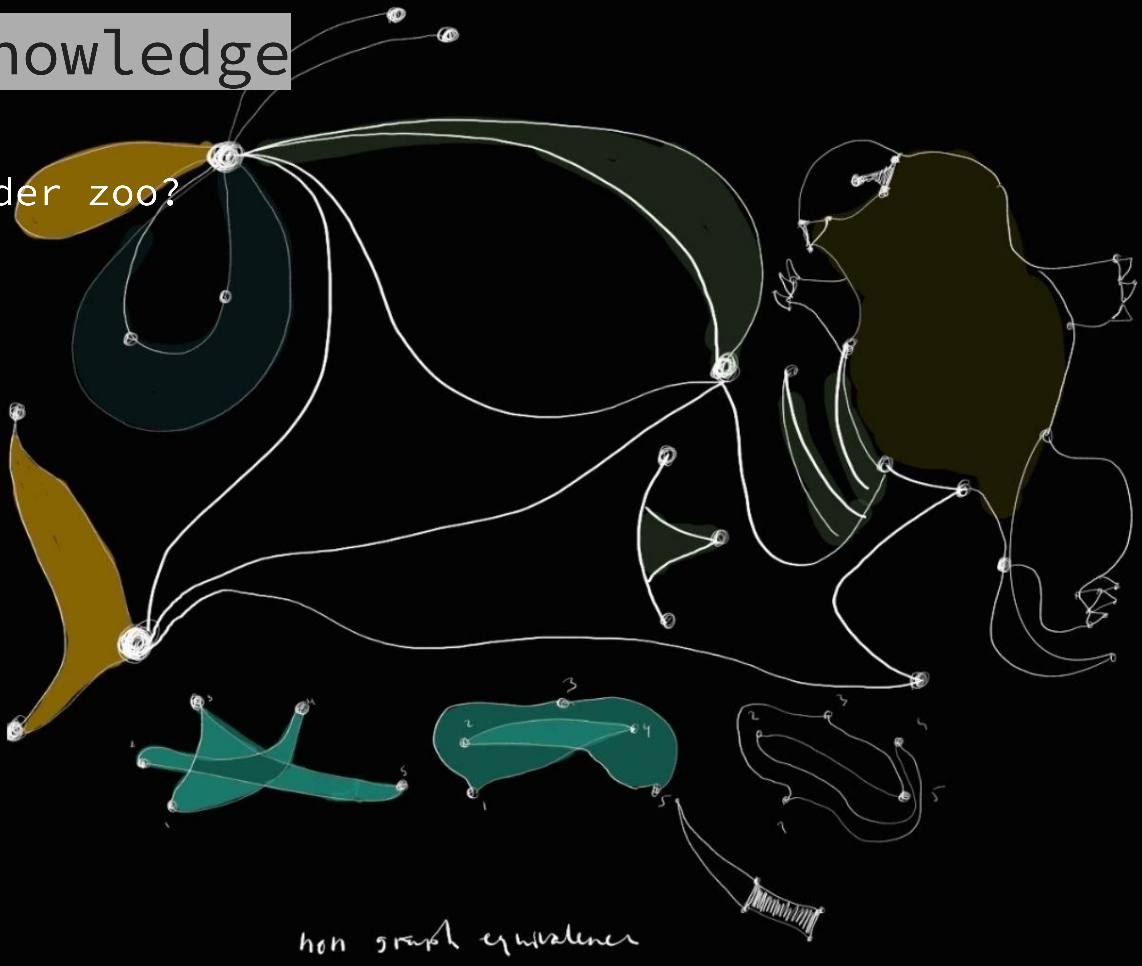
b) which lower-order measures are still applicable on higher-order structures?



OK, so now we arrived to a higher order zoo, but how can we study it?

How to analyze knowledge hypergraphs?

how can we study higher order zoo?



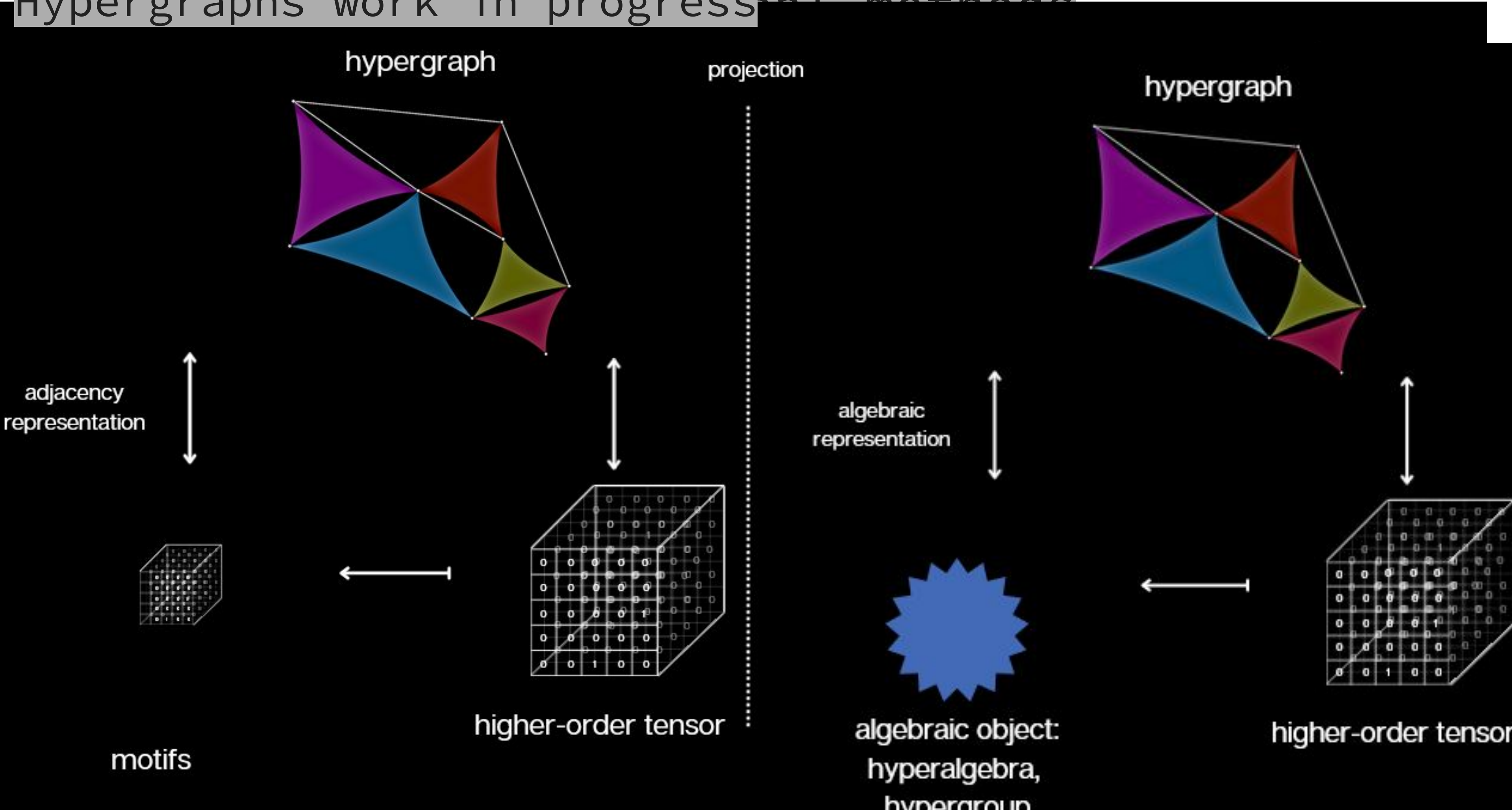
Apply generalized Laplacian …
Apply colored algebra

Apply all possible graph measures

Apply rewriting :)

XGI
stable

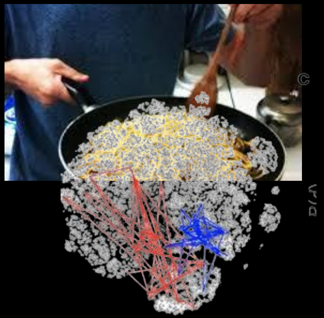
Hypergraphs work in progress[redacted]

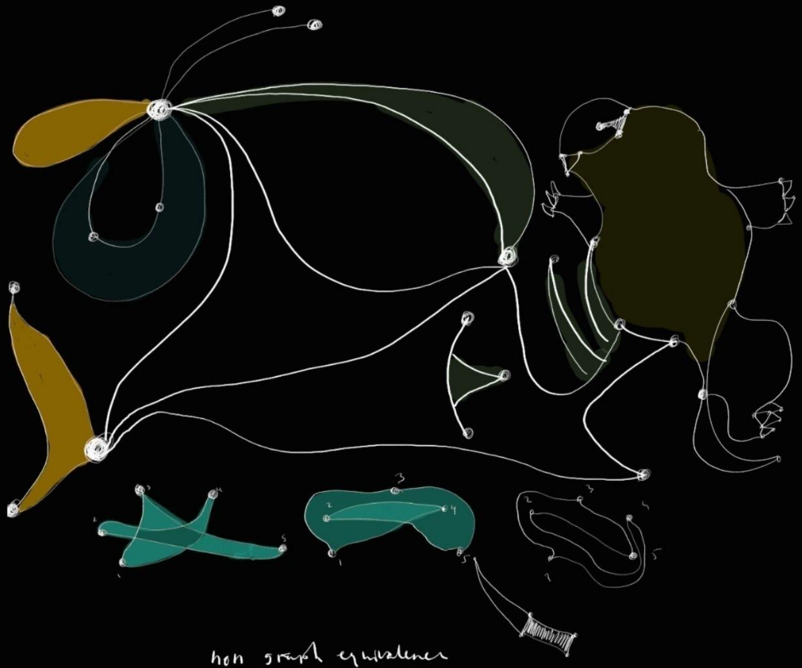


Some take-aways

Encode the hypergraphs to algebraic structures, introduce metrics with small G-H distance

Dyadic vs. higher-order motifs presence gives different results for higher-order





Questions?
liubov.tupikina@cri-paris.org

Singh, [...], **Santolini+** PLOS ONE 2022
Singh, [...], **Santolini+** EPJ (in rev) 2023