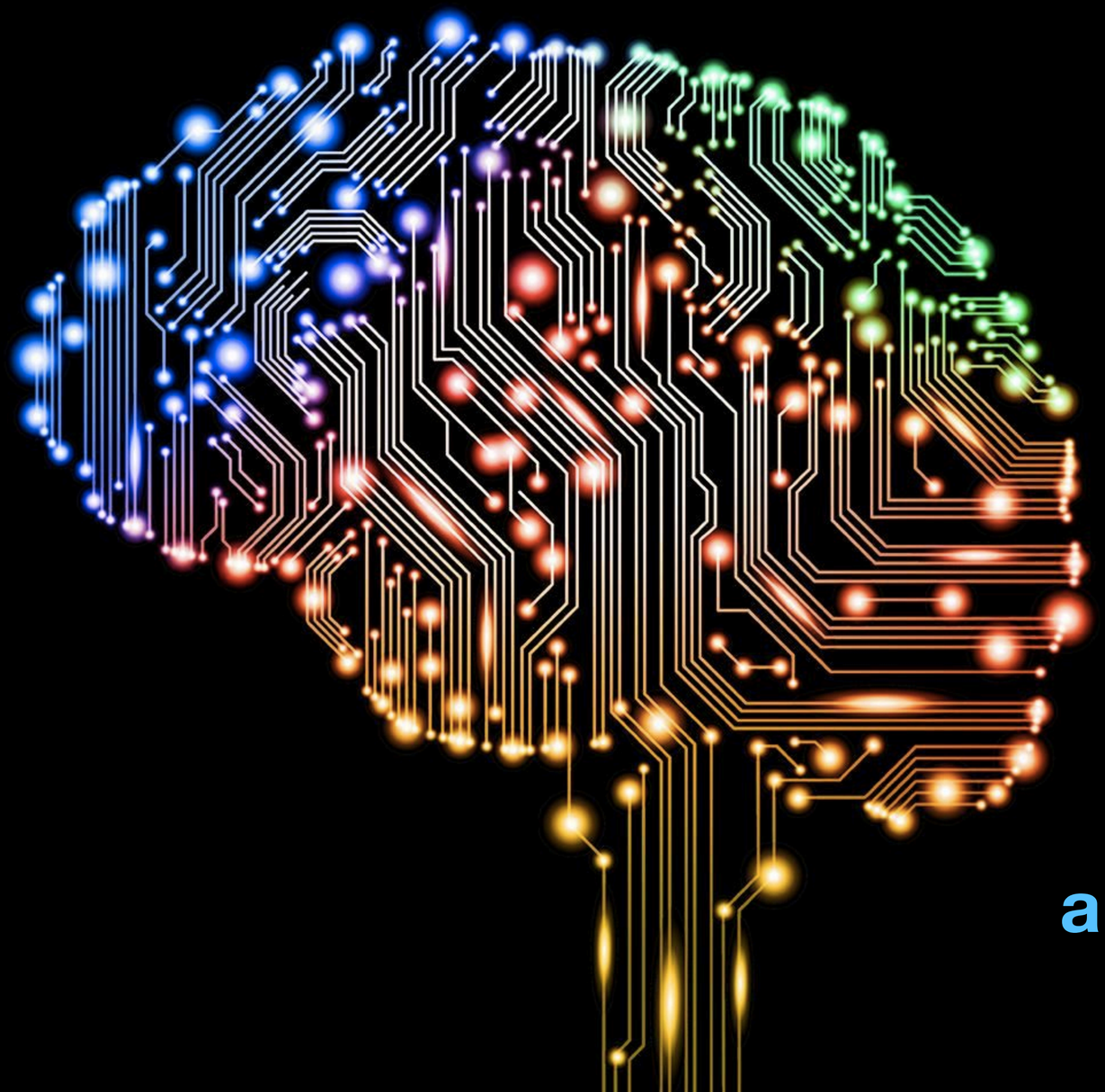
astrophysics data system



Machine / Deep learning and Natural Language Processing

Sergi Blanco-Cuaresma, and the ADS team

astrophysics data system

Senior Review 2020

5 years plan

Astrophysics Archives P
Proposal s
The NASA Astrophysi

Accelerating Dis
Enhanced Infor

Alberto Accomazzi, Michael J. Kurtz, Ed
Center for Astrophysics
3 Febru

Executive Summary

The NASA Astrophysics Data System (ADS
digital library 27 years ago, and has re
network for astrophysics research for m
space, despite massive changes in the
disseminate their results. These changes
experimental facility into a stable, robus
policies reflect the needs and priorities of th

The last five years have seen substantial
new management structure, has develo
migrated its community of 50,000 users
state-of-the-art search engine, a modern
access to the ADS data collections and
developed following an open source mode

The ADS's mission is aligned with NA
knowledge by enabling open science
research. The ADS has a unique role wit
focuses on the scientific literature to he
explore their connections. As interdisci
become organically connected and discov
readership. By further connecting the lite
ADS increases discoverability of both and

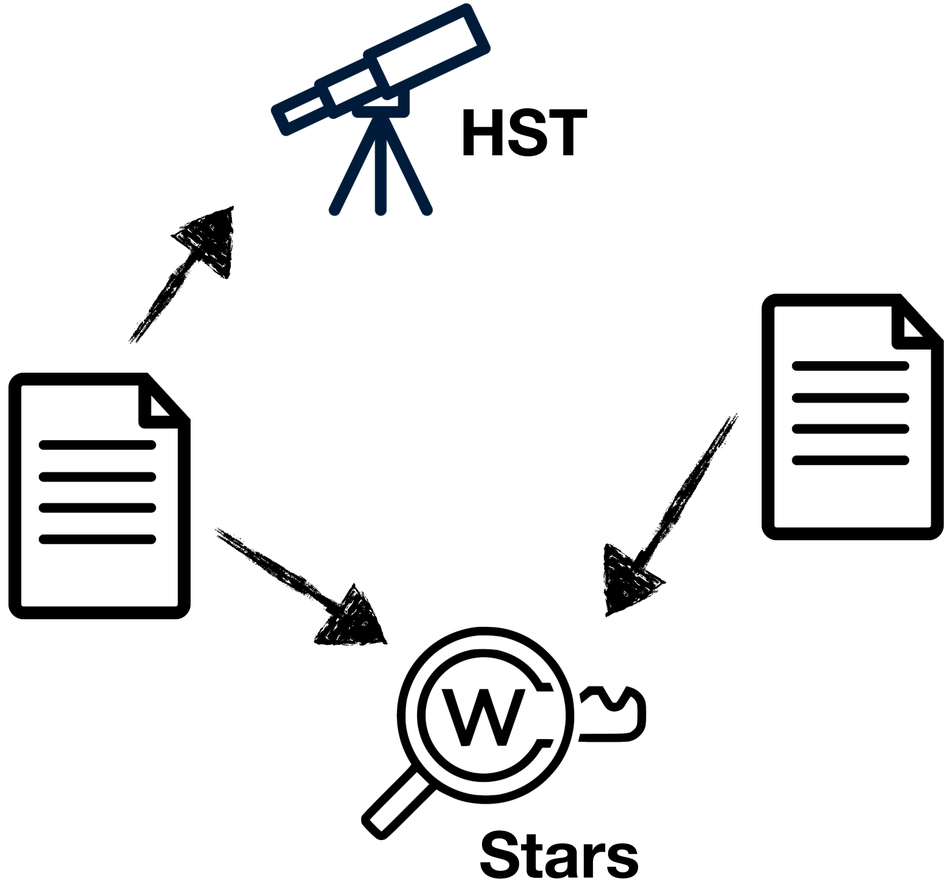
The NASA Astrophysics Data System

1. Improve Discovery. Combining improved metadata such as derived from named entity and concept extraction (He et al. 2019; Zhao et al. 2019, see section 3.1.4) with new graph-based clustering (Traag et al. 2019) promises new powerful avenues for human-machine interactions. Likewise, combining the full text of articles with the new context-sensitive vector space literature models (Devlin et al. 2018; Peters et al. 2018) will enable the use of language models to improve the efficiency and accuracy of scholarly information retrieval. These recent advances in the field of NLP and AI provide an opportunity for the ADS to significantly improve its ability to better disambiguate the content it indexes and better understand the intent behind a user query, which translates into more relevant results (see section 3.2.4). As the number of articles in the literature increases (and the relevant signal gets buried in a constantly increasing noise), **we intend to use state-of-the-art NLP techniques to improve our search, recommendation and notification systems.**

Senior Review 2020

5 years plan

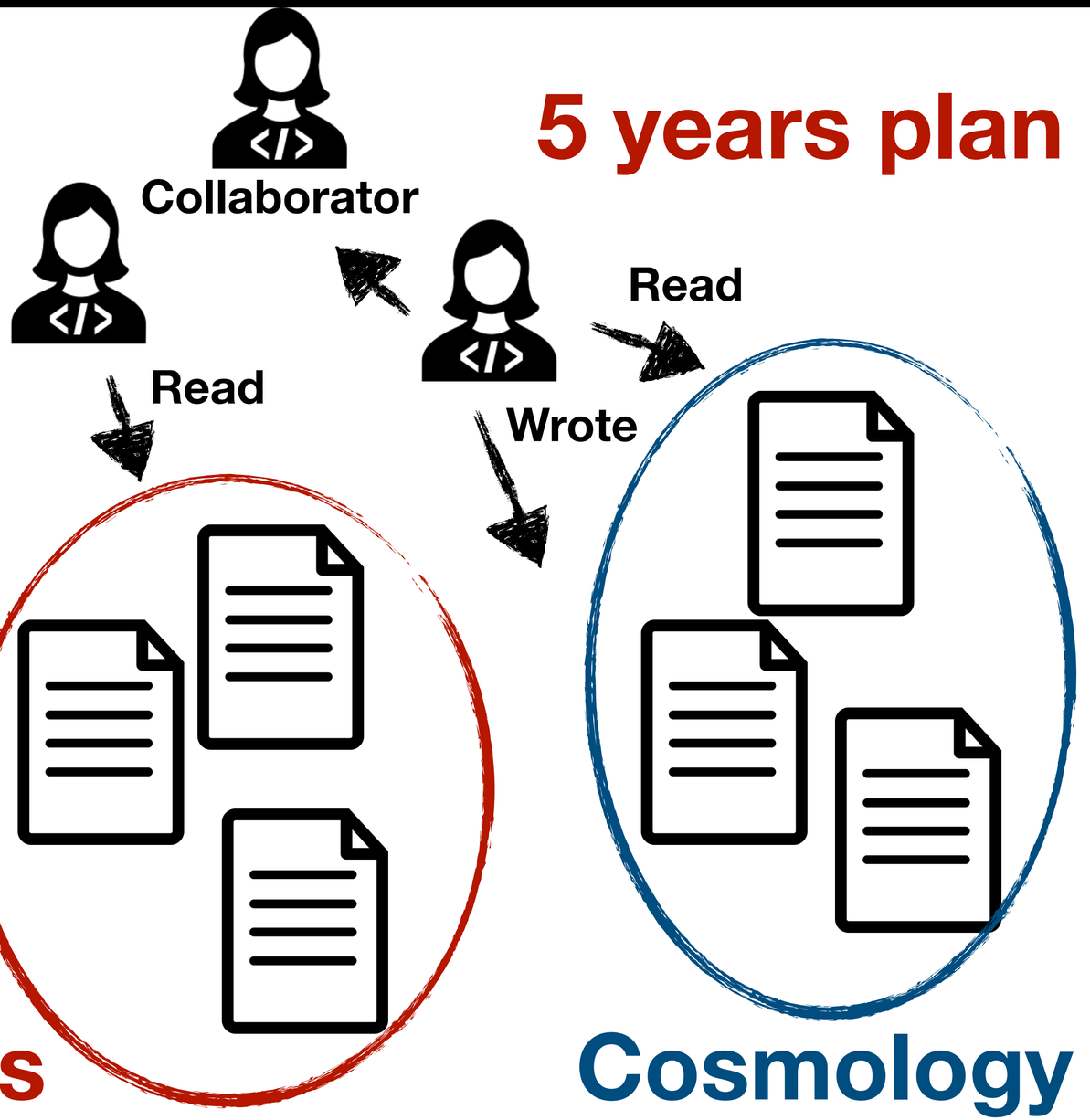
- Link papers to facilities and Unified Astronomy Thesaurus (UAT) terms



Senior Review 2020

- Link papers to facilities and Unified Astronomy Thesaurus (UAT) terms
- Implement an advanced recommendation system

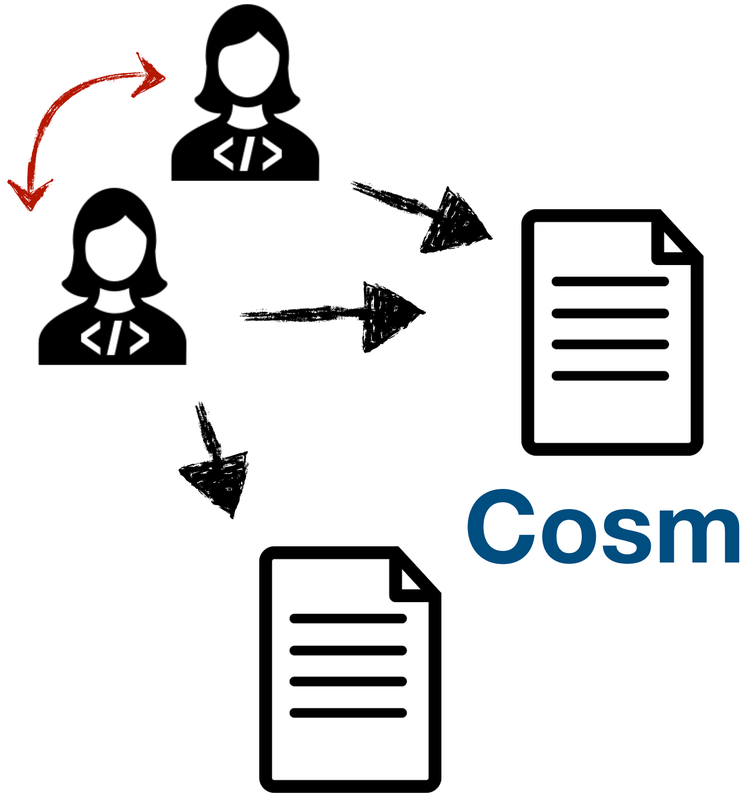
5 years plan



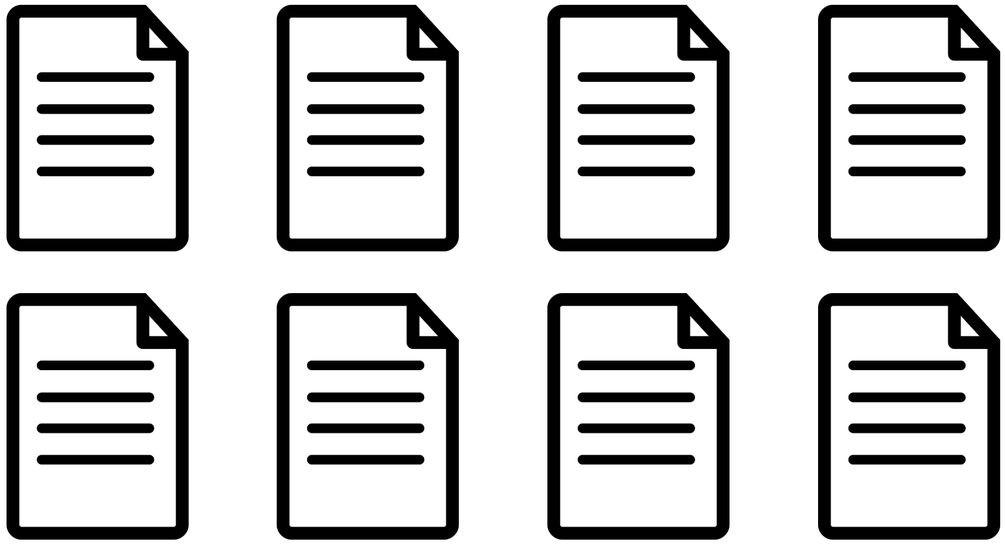
Senior Review 2020

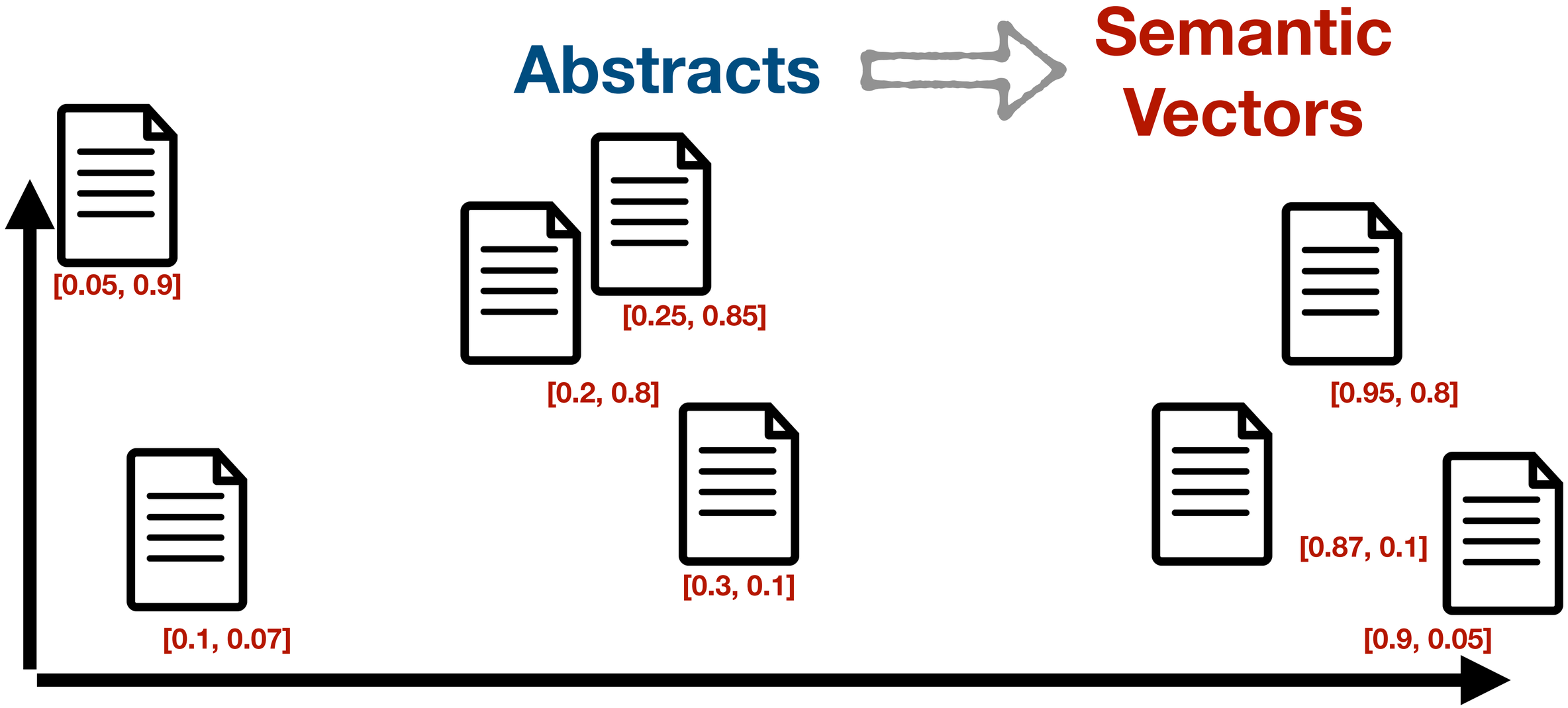
5 years plan

- Link papers to facilities and Unified Astronomy Thesaurus (UAT) terms
- Implement an advanced recommendation system
- Disambiguate authors (e.g., Is this your paper?)

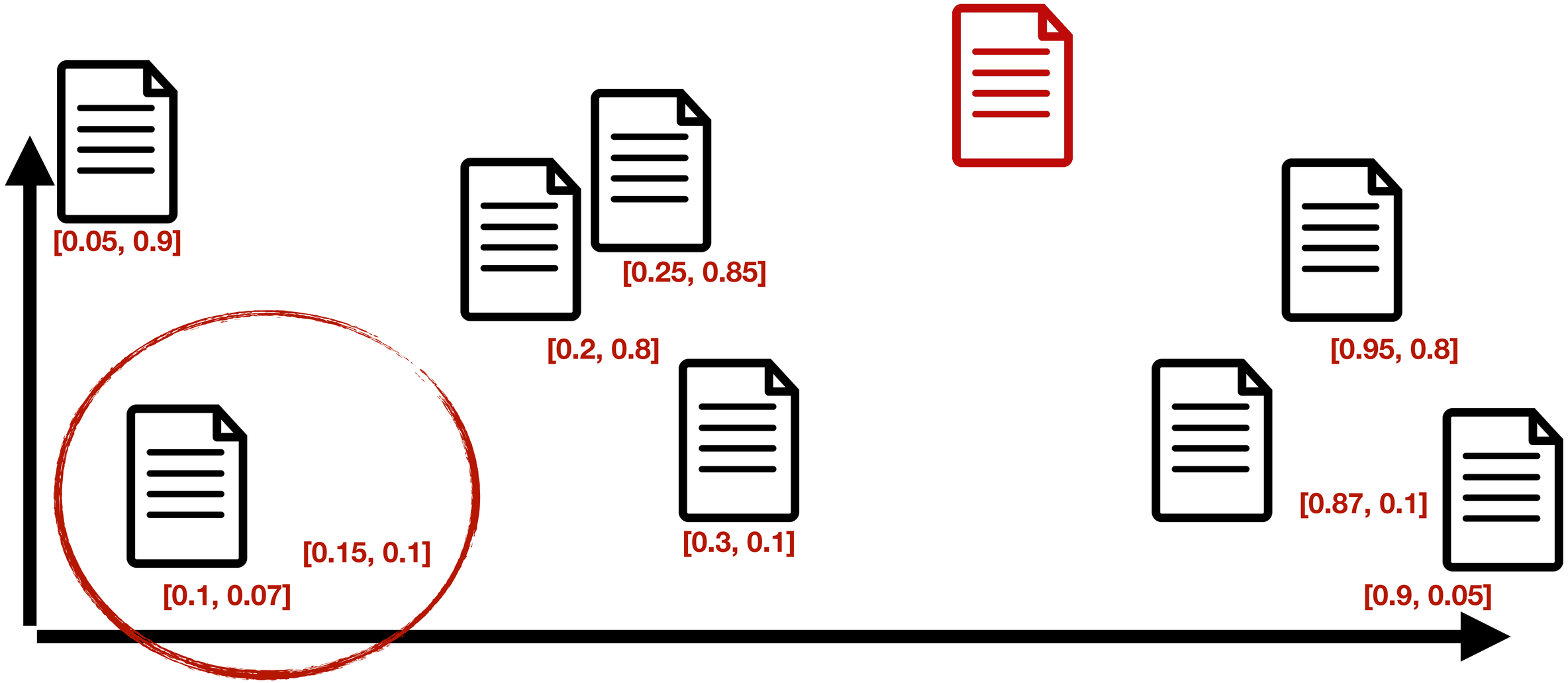


Abstracts

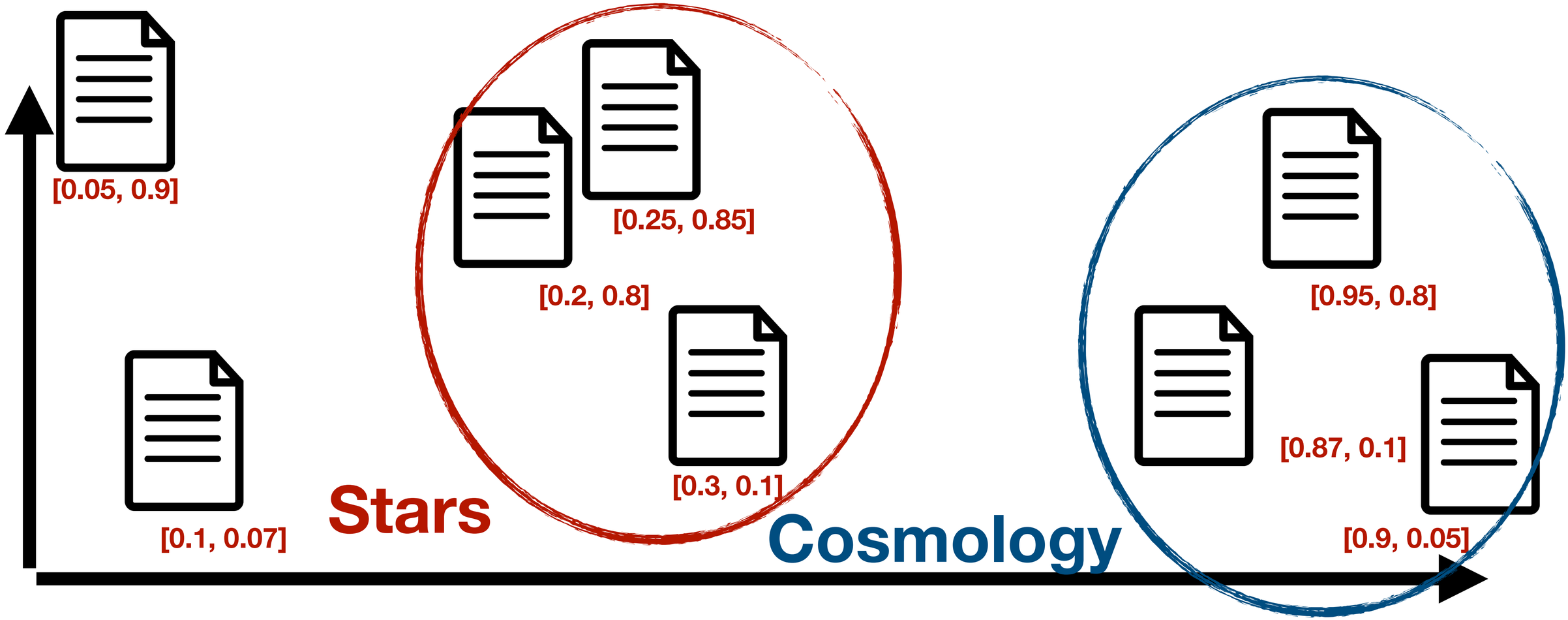



Abstracts
Semantic Vectors
[0.05, 0.9]
[0.25, 0.85]
[0.2, 0.8]
[0.95, 0.8]
[0.3, 0.1]
[0.87, 0.1]
[0.1, 0.07]
[0.9, 0.05]

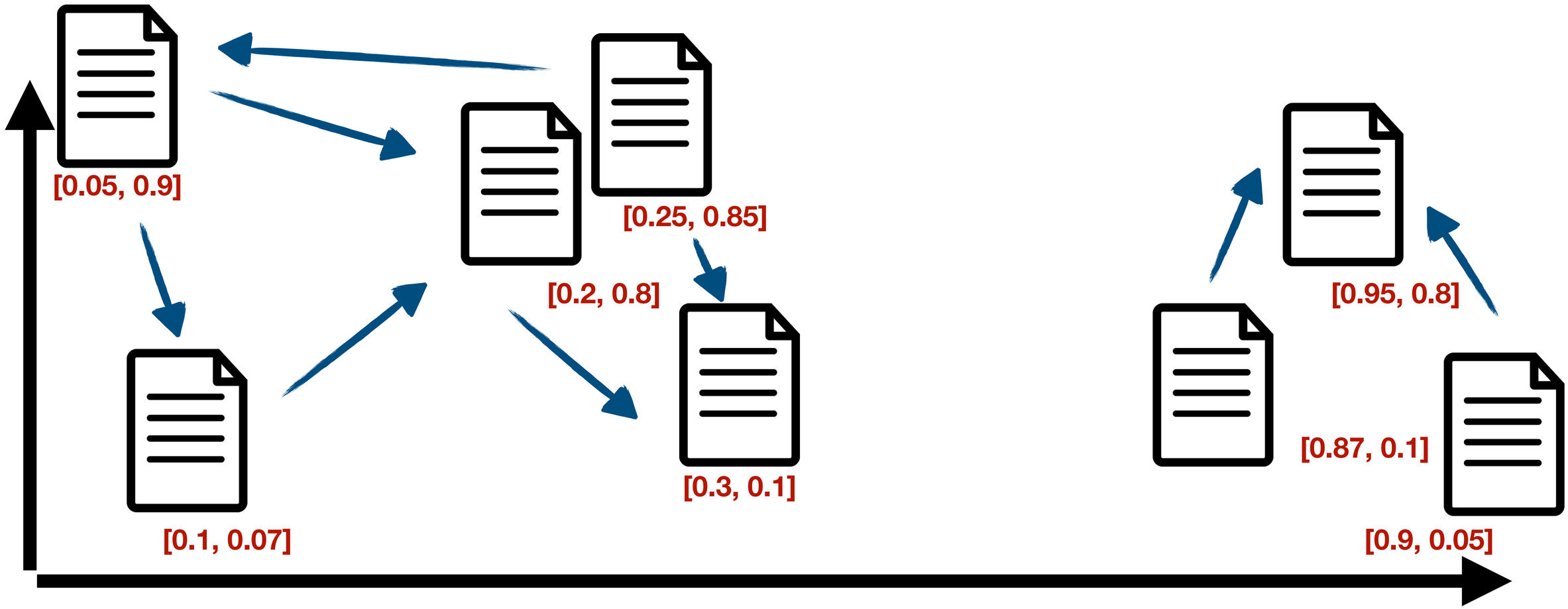
Approximate nearest neighbors



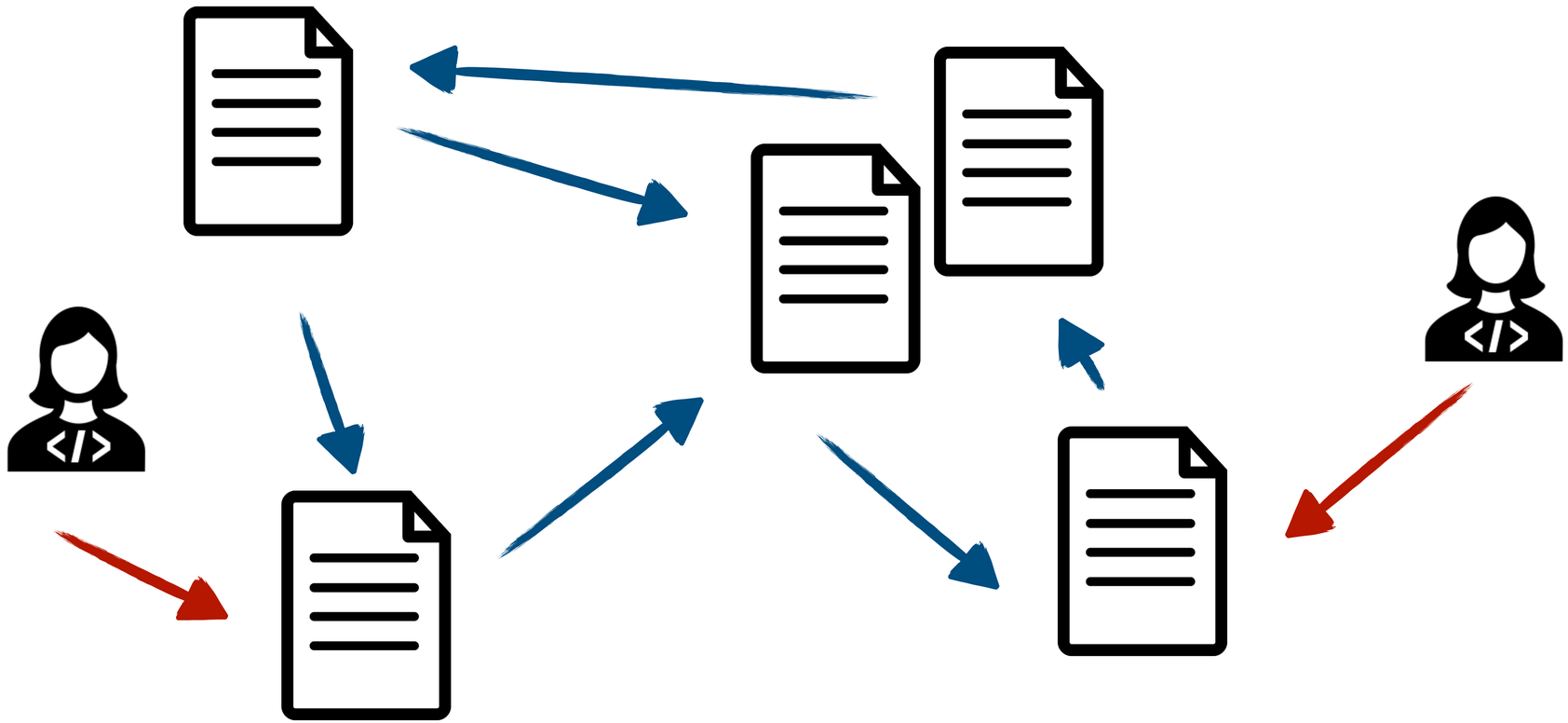
Clustering


[0.05, 0.9]
[0.25, 0.85]
[0.2, 0.8]
[0.3, 0.1]
[0.95, 0.8]
[0.87, 0.1]
[0.1, 0.07]
[0.9, 0.05]
Stars
Cosmology

Semantic content → Known relationships

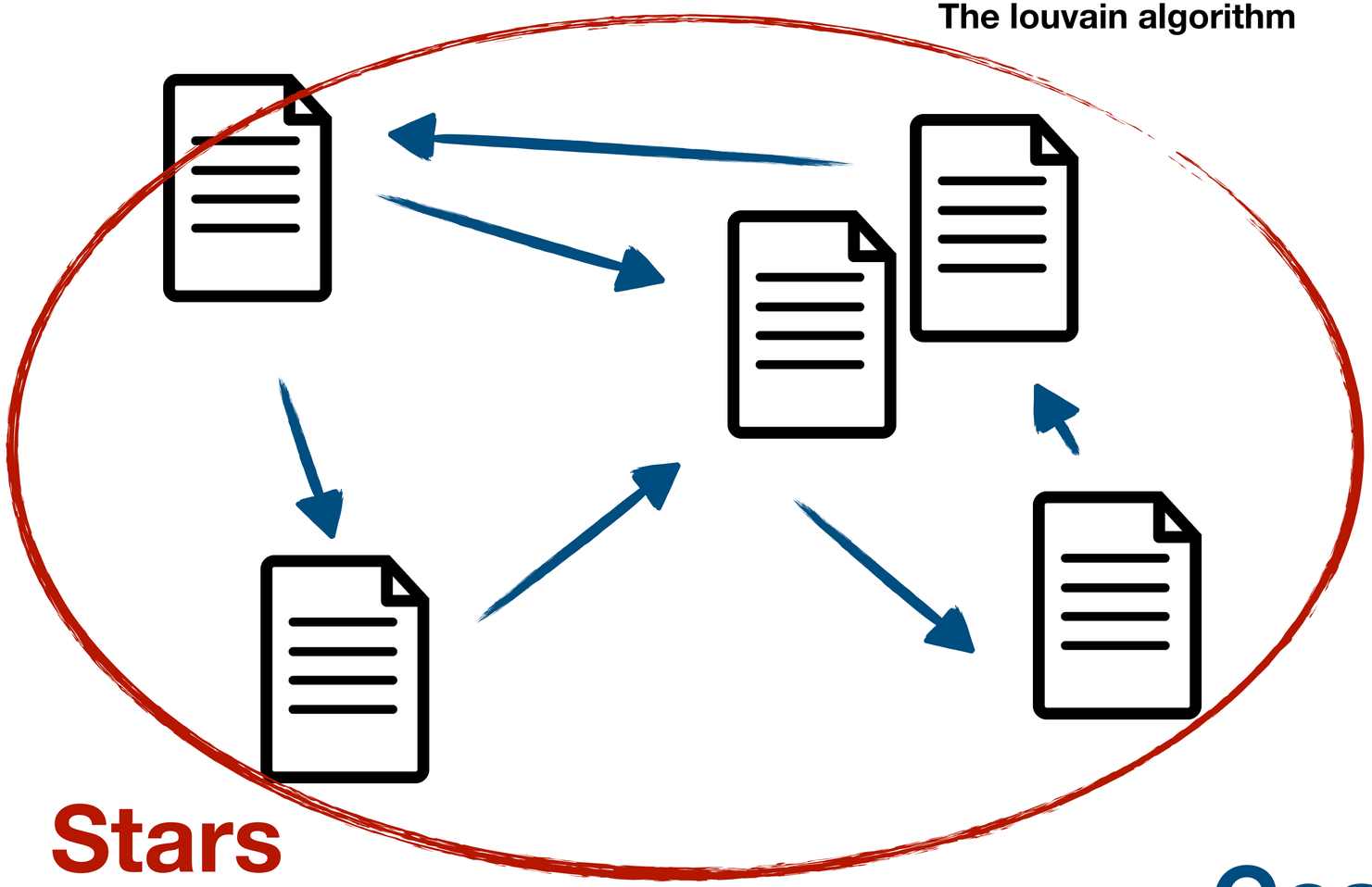


Known relationships → Graph

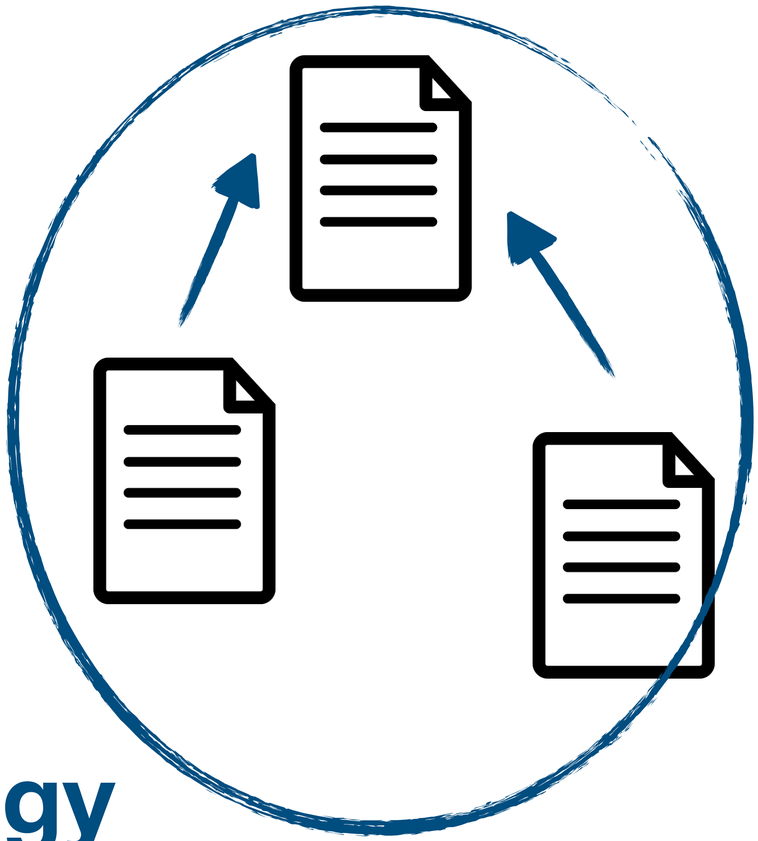


Community detection

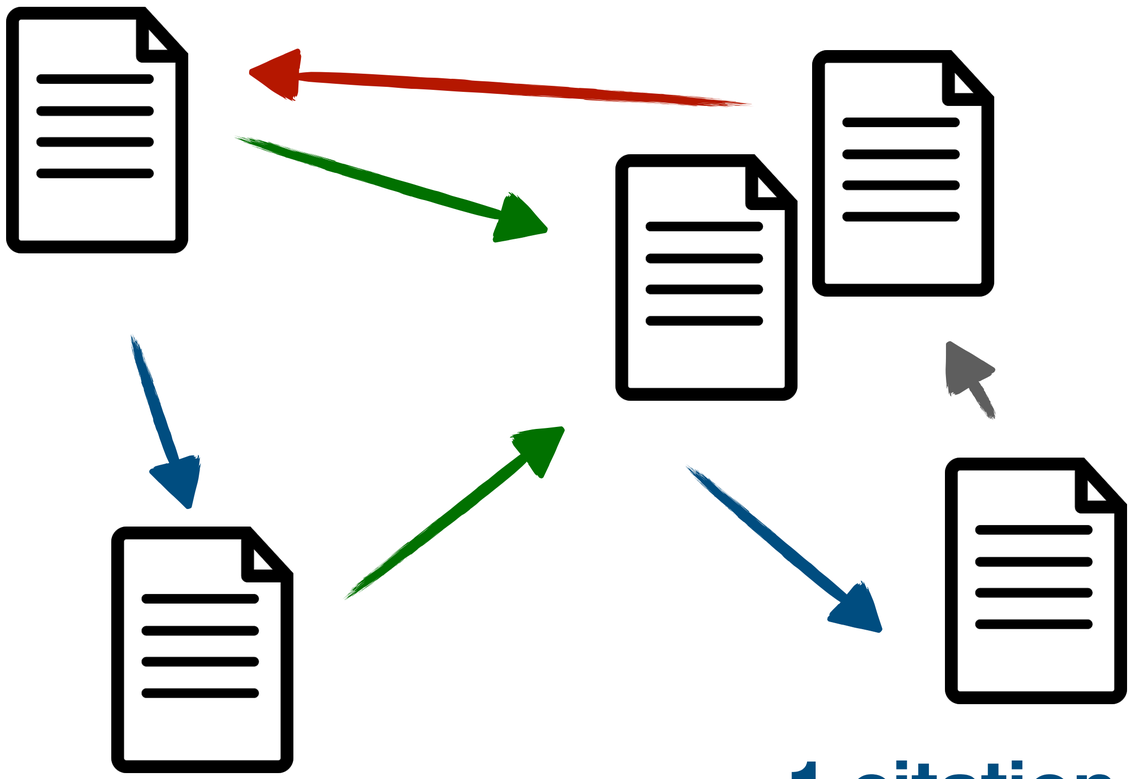
The louvain algorithm



Stars



Cosmology



1 citation with 1 citations

1 citation with 2 citations

PageRank

	PageRank	# Citations
1992ASPC...25..432K	3.428	125
2000A&AS..143...41K	3.268	87
1998PASP..110..934K	3.263	408
1993ASPC...52..132K	1.401	19
1985AJ.....90.1665K	1.244	28

PageRank

Summary

Improve discovery

- Link papers to facilities and Unified Astronomy Thesaurus (UAT) terms
- Implement an advanced recommendation system
- Disambiguate authors (e.g., Is this your paper?)

5 years plan…