

Univerza v Ljubljani  
Fakulteta *za računalništvo  
in informatiko*



Luka Žontar

**INTERIM PUBLIC SUMMARY AND  
ESTIMATED PLAN ON MASTER'S  
THESIS TITLED:  
RECOMMENDING COLLABORATION  
LINKS IN CO-AUTHORSHIP  
NETWORKS**

Mentor: Assoc. Prof. dr. Tomaž Curk

Ljubljana  
19. 5. 2024



# Main goals of the thesis

- Investigate the nature and influence of new collaborations between EUTOPIA institutions and departments.
- Develop a recommender system for interdisciplinary and cross-institutional academic partnerships using graph neural networks to foster new collaborations within the EUTOPIA organization.
- Ensure recommendations adapt to evolving research trends.
- Provide explainable and real-time collaboration suggestions through an interactive dashboard.



# Work done & main results

- Established integration for daily updates and historic data pull on 4 distinct data sources: Crossref, Unpaywall, ORCID and Elsevier (Elsevier is still work-in-progress).
  - For each data source we prepared a script for historic data pull and an ingestion pipeline to update datasets with recent data.
- Combined all the datasets into a bipartite network of articles and authors connected by authorship.
- Working on preliminary analysis on enhanced datasets
  - some results shown on the next slide.

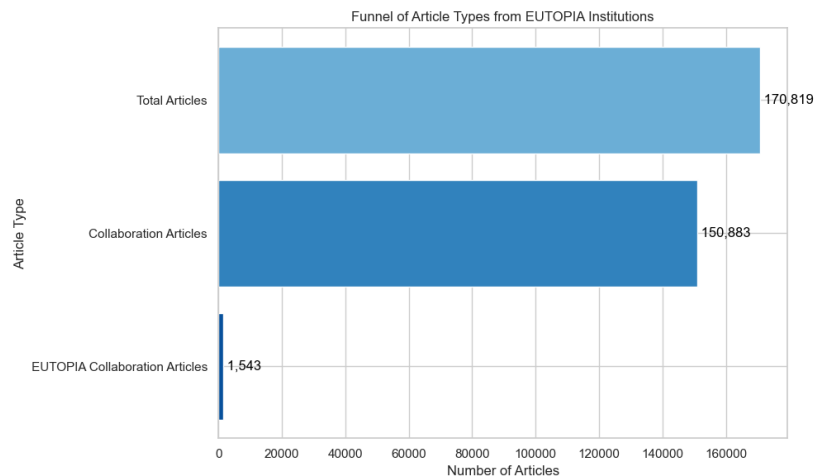


# Work done & main results

## Preliminary analysis

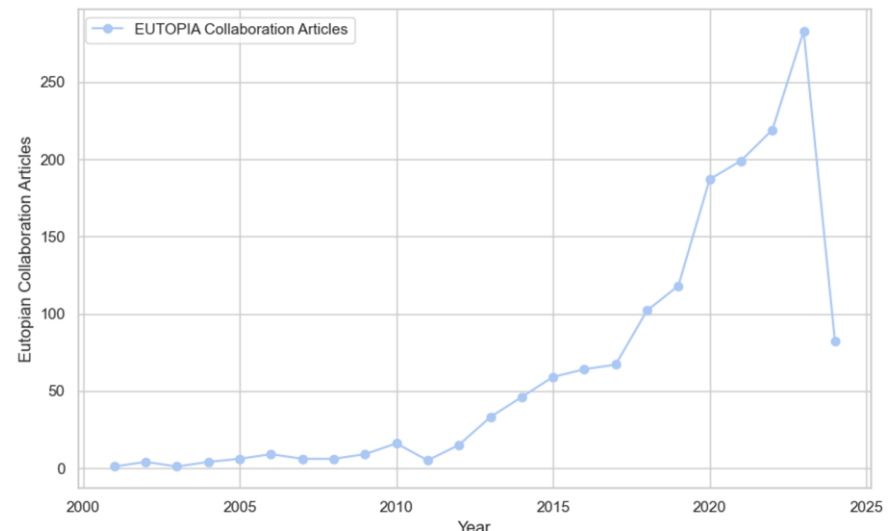
### General Data Statistics

- 3M articles from Crossref, where ~60% has affiliation data.
- 320K articles from EUTOPIA institutions, only 1500 articles made in collaboration between different EUTOPIA institutions.



### Collaboration Trends

- Both EUTOPIAN and other collaboration articles are trending positively.



### (WIP) Testing hypotheses:

1. Following first collaboration event, collaboration between institutions increase exponentially.
2. A new collaboration significantly impacts the research topics of both institutions.
3. Collaborations are most likely to be motivated by expertise search.



# Estimated plan

- (1. 6. 2024) Data from Unpaywall, Crossref, ORCID and Elsevier combined and cleaned for further analysis along with potential other data sources showing outside trends which are relevant for the collaboration recommendation (e.g. Google Trends).
- (30. 6. 2024) A prototype of the data exploration dashboard developed (not including any recommendation only relevant for interactive data exploration & hypothesis validation).
- (31. 7. 2024) An optimized version of the recommendation model trained and delivered via REST API. This part includes extensive experimentation on the available dataset and comparing results to the baseline.
- (8. 8. 2024) Recommendation system included in the data exploration dashboard and finalized version of the dashboard delivered.
- (1. 9. 2024) First version of the thesis delivered to review to mentor.