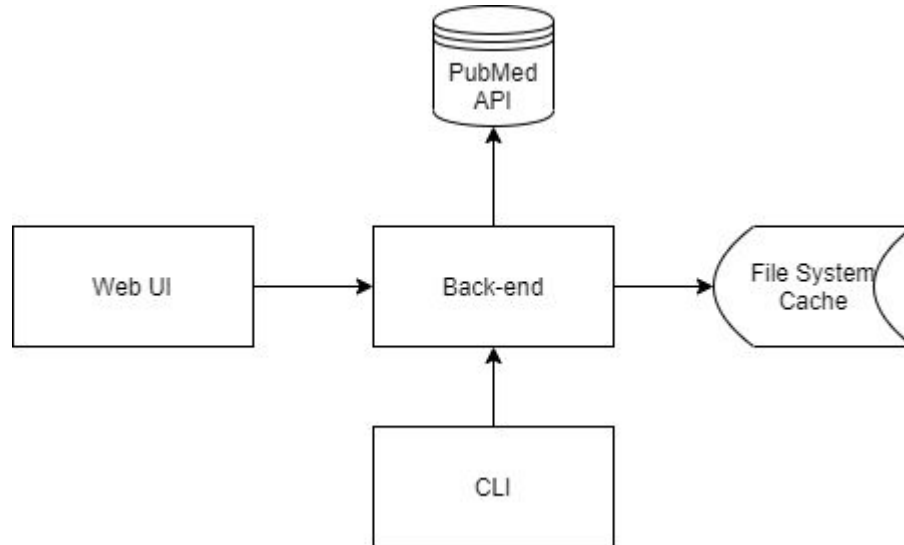


Project 02 - Author Maps

Group 5 (I. Yalchyk, M. Bareth, P. Mishra, D. Rathod)

Project Overview

- Create a pipeline that generates a co-author network which details how strongly connected the queries author is to his/her co-authors.



Other approaches (only Papers)

- Automation to solve name disambiguity
https://philippmayr.github.io/papers/Co-Authorship_JCDL2016.pdf
- Also include shared topics and publishers
http://www.ccs.neu.edu/home/yzsun/papers/asonam11_pat_hpredict.pdf
- Weighting of publications by metrics like:
 - Influence of shared papers (<https://doi.org/10.1016/j.physa.2014.10.006>)

Team roles

Dhruv: backend (data layer, including requests to PubMed, caching)

Marlo: backend (business logic, including graphs creation)

Pragya: Web UI, QA

Ilya: CLI, team coordination

Development processes

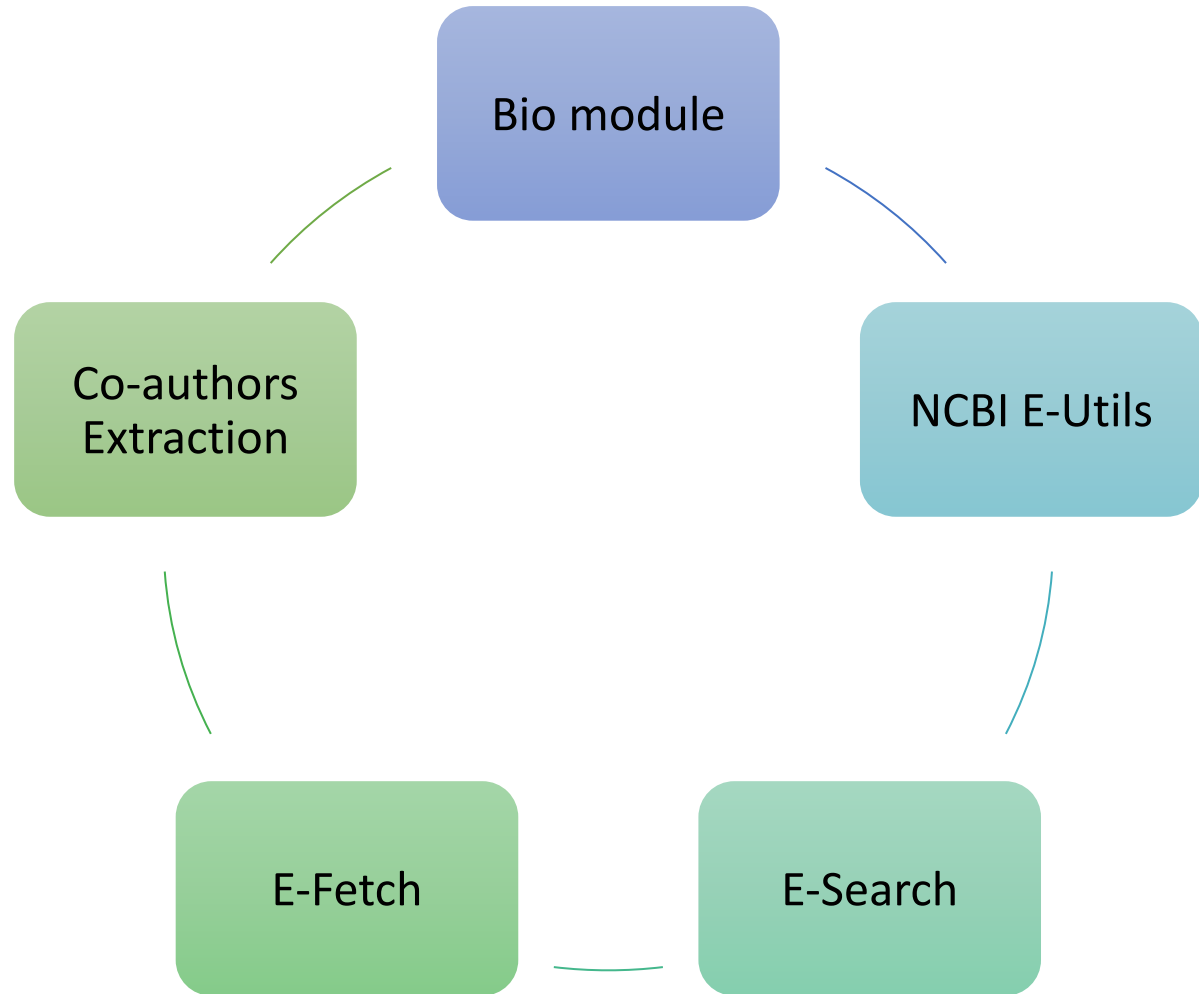
1. GitLab issues
 - a. bug
 - b. improvement
 - c. question
2. Each branch corresponds to a certain issue
3. Merge requests
4. Code review

CLI

6 CLI commands:

- `get_publications AUTHOR_NAME`
- `get_coauthors AUTHOR_NAME`
- `get_common_publications AUTHOR_NAME_1 AUTHOR_NAME_2`
- `get_common_publications_for_all_coauthors AUTHOR_NAME`
- `get_common_publications_counts_for_all_coauthors AUTHOR_NAME`
- `generate_graph AUTHOR_NAME OUTPUT_FILE [FORMAT]`

API



Important Methods:

- `def get_list_of_publications (self) -> List[str]:`
- `def get_list_of_coauthors_from_list_of_publications (self) -> List[str]:`

What if the author gets a new publication after caching the data once?

- Package checks for each cached record and each retrieved record if single record is missing it downloads it.

Difficulties

- No batch download available. Need to rely on slow responding NCBI server for large amount of data.

Graph preparation

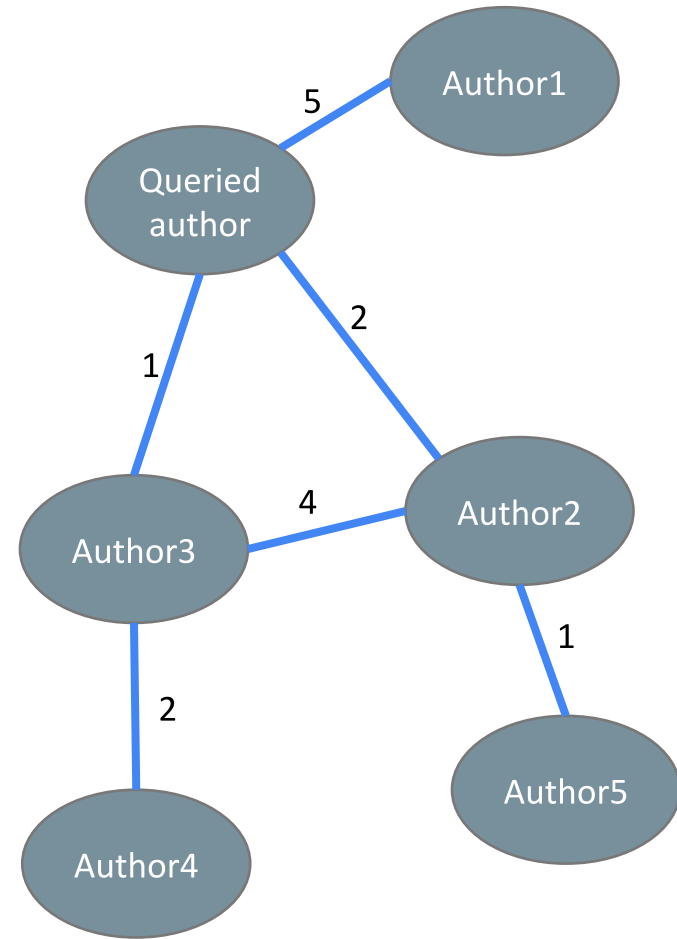
- Input is the first rank of authors as list
- Queries those authors and finds unions
 - only requires one function from dhruvs data layer
 - Creates a dict
 - key: author combination (tuple)
 - value: list of shared publications
- Function to transform values from list into integers

Graph class/creation

- Stores input and Graph object
- Export as string and file(cache or given path)

Problems:

- Only inhabits few visual cues for the user to orientate
- Large networks may be overcrowded

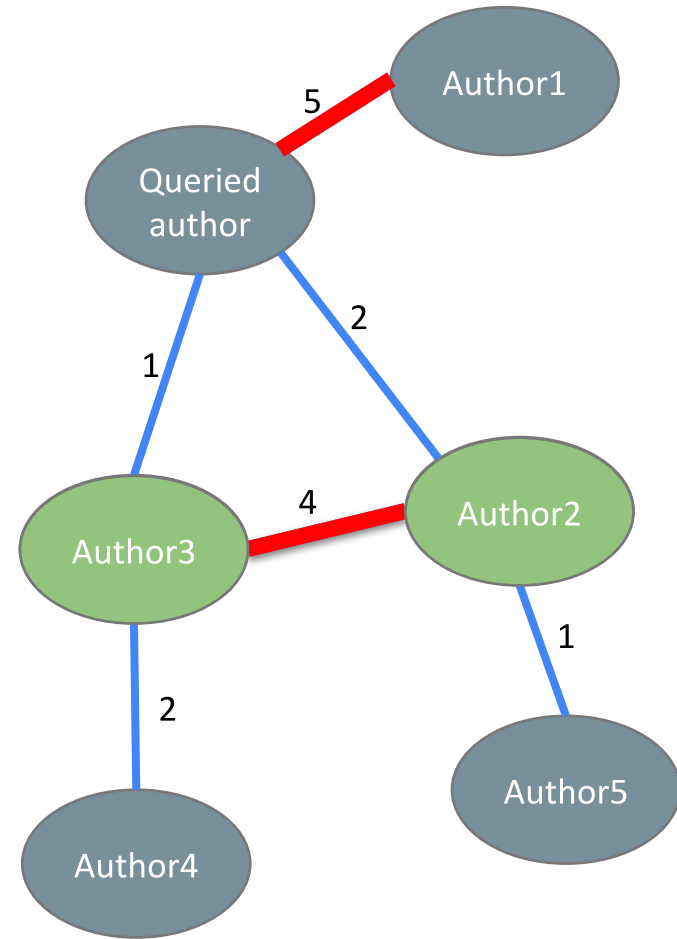


Graph creation

Adjust line thickness according to number of shared publications

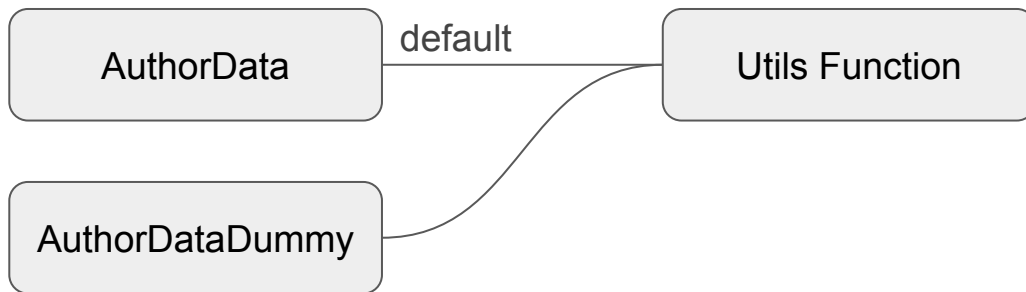
Color the top 10% in red to increase visibility

Allow selected nodes to be highlighted



Difficulties

- Data layer was not immediately available
- Solution: dependency injection for utility functions
 - create a dummy class in the unittest
 - overwrite dependency and uses testdata
 - pass tested function the dummy class as a parameter



Discussion - Graph

Varying number of shared publications and edges

Still sometimes overcrowded

Thank you for your attention! :)