SOCIAL NETWORK ANALYSIS ON RESEARCH

YUSUF ERDEM NACAR & MEHMET EMRE AKBULUT
ADVISORS: SUZAN ÜSKÜDARLI & H. BIRKAN YILMAZ
BOĞAZIÇI UNIVERSITY, COMPUTER ENGINEERING DEPARTMENT

PROBLEM STATEMENT

To design and develop an analysis tool that allows the researchers to perform **social network analysis** on research networks that describe articles, authors and the relationships between them in order to uncover helpful and quantifiable information.

MOTIVATION

The core elements of scientific research include articles, researchers, and institutions. Since scientific research is the cumulative effort of researchers to increase the understanding of the world around us, the relationships between these elements are as important as the scientific results themselves.

Gaining insight into the relationship between the core elements of scientific research can be useful for a variety of purposes, such as;

- Guiding scientific effort toward better use of resources
- Inferring comparative results between fields of research
- Better representing the importance of certain research fields and/or research groups
- Discovering key researchers, and/or articles for networking

Our motivation in this project is to develop a way to provide the researchers with quantifiable information about the relationships between these elements so that it can be used for such purposes. This quantifiable information includes graph measures of individual elements of a graph as well as the graph measures of the whole graph.

STATE OF ART

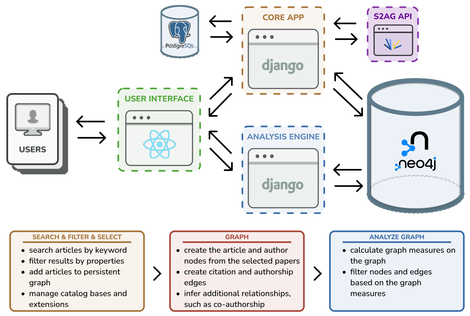
There are several online services that provide similar graph building, however, none of them aim to let the users analyze the graph they built. These services are mainly focused on providing the researchers with a more convenient way of finding articles that are related to an article that they are interested in. The most well-known tools of this nature are

- Connected Papers
- Research Rabbit
- Litmaps

Although these tools provide their users with a research graph of articles, they do not provide the measures of the graphs they built since their focus on building these graphs is to enhance the literature exploration of their users.

DESIGN

Our main approach in the design of the project is to divide the project into 3 main parts, **Search & Select**, **Graph**, **Analyze**, and additionally a user interface. To search for information about articles, authors, and citations, Semantic Scholar Academic Graph (S2AG) API is used.



The graphs created using this design are composed of 2 types of nodes and 3 types of edges, one of which is inferred from a specific configuration of certain node and edge types.

Fig.1 Authorship between an author and an article

Fig.2 Citation between two articles

Fig.3 Authors of an article

Fig.4 Co-authorship between two authors of an article, inferred from **Fig.3**

To investigate the relationship between the graphs constructed and other structured knowledge representations such as ontologies, additional node types, and relationship types as shown in Fig.5 and Fig.6 are designed to be added. These new node and edge types, or analogous entities and relationships are typically found in many ontologies and serve as the interface between the research graphs constructed and the ontology to be connected.

Fig.5 Article mentions concept

Fig.6 Concept belongs to a semantic type

RESULTS

Given the nature of any research field, the network of scientific knowledge and researchers is immensely complex. Therefore, to make sense of how aspects of these graphs relate to each other, one would not only need the quantifying information on the graph but also how this information changes as the graph itself evolves.

Here we present an example coauthorship graph with 4569 authors constructed using the 24 highest-ranked-by-S2AG papers retrieved by the keywords *Toxoplasma* and *Mitochondria*, the papers citing these papers, and the papers referenced by these papers. These many authors yield 29487 coauthorship relationships.

In Fig.7, a subgraph with author nodes size-scaled with the number of papers authored and colored by PageRank score is presented.

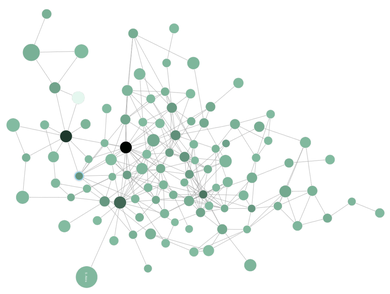


Fig.7 Coauthorship subgraph of *Toxoplasma* and *Mitochondria* research graph, constructed using the 24 highest-ranked-by-S2AG papers, where the author nodes size-scaled with the number of papers authored and colored by PageRank score

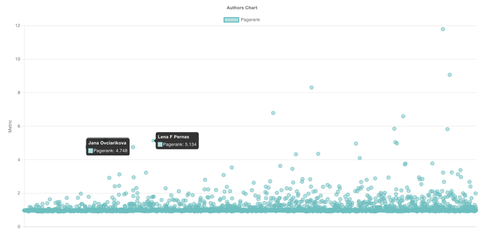


Fig.8 Pagerank scores of the authors where the x-axis is increasing with the number of papers published.

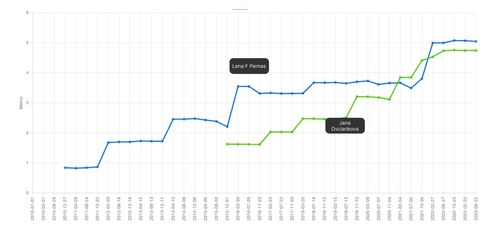


Fig.9 Pagerank scores of two authors with respect to time

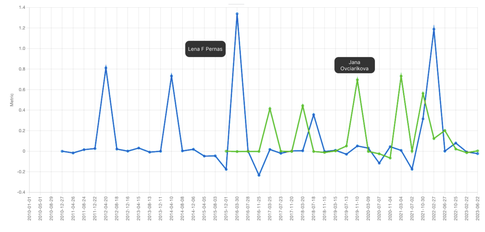


Fig.10 Change in Pagerank scores of two authors with respect to time.

Having a graph-structured representation of the research world allows the addition of explicit connections to other graph-structured knowledge representations. One prime example of such knowledge representations is ontologies. In this part of the project, the possible connections between a selected ontology, Unified Medical Language System (UMLS), have been investigated. The connections between the papers and the UMLS concepts are constructed by passing the abstracts of the papers through a named entity recognizer called MedCAT. Then, "MENTIONS" relationships between the articles and the UMLS concepts are constructed as proof of concept.

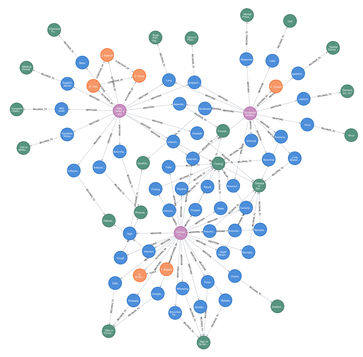


Fig.11 Example research graph constructed using 3 papers that include the word *pathology* in their abstracts, the ontological concepts, and the relationships between them.

This representation reveals a bigger picture in terms of how the aspects of research and the semantic concepts relate to each other.

CURRENT STATUS

We can summarize the capabilities of the project as follows:

- Listing papers related to a search query with the search engine
- Applying filters on these papers, listing the papers that these papers cite and the ones that cite these papers
- Analyzing graphs using various graph measures and how they change w.r.t time.
- Interfacing UMLS by using named entity recognition.

There are several key differences between our project and similar services. These are:

- Letting users freely choose the articles that they want to have in their graph and allowing users to extend their data with some options such as 'add citations of my papers', 'add references of my papers', etc.
- Focusing on network analysis and obtaining meaningful information using social network analysis

FUTURE WORK

There are a few key steps that we aim to develop about the application. These next steps in order of importance are:

- Integration graph analysis with the information obtained from named entity recognition and medical ontologies
- Exploring different metrics or graph topology measures such as community detection, etc.
- Redesigning and reimplementing the user interface for a better experience

REFERENCES

[1] S. Wasserman and K. Faust, "Social Network Analysis: Methods and Applications," 1994.

[2] J. M. Harris, J. L. Hirst, and M. J. Mossinghoff, "Combinatorics and graph theory," 2000.

[3] A. D. Wade, "The Semantic Scholar Academic Graph (S2AG)," Companion Proceedings of the Web Conference 2022, 2022.

[4] O. Bodenreider, "The Unified Medical Language System (UMLS): integrating biomedical terminology," Nucleic acids research, vol. 32 Database issue, pp. D267-70, 2004.

[5] Z. Kraljevic et al., "Multi-domain Clinical Natural Language Processing with MedCAT: the Medical Concept Annotation Toolkit," Artificial intelligence in medicine, vol. 117, p. 102083, 2020.