Mapping Change in Large Networks (25-1-R-13)

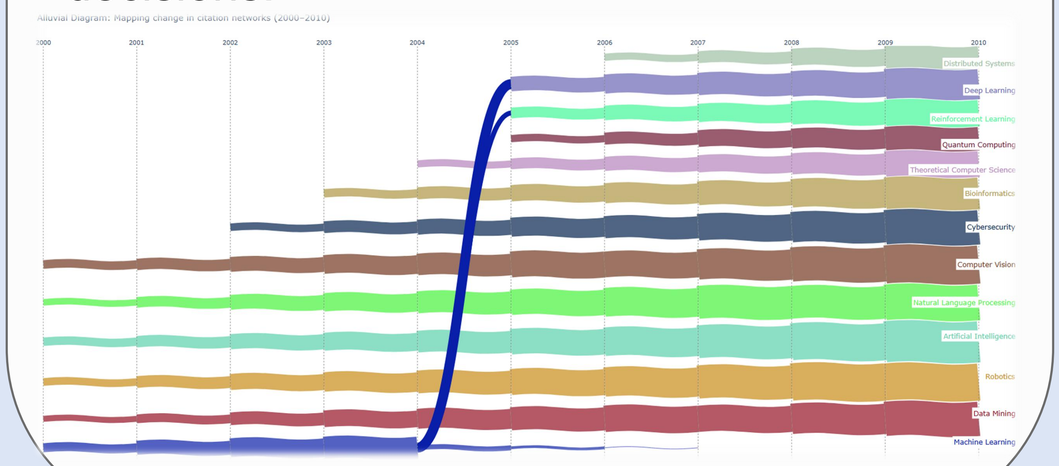
Students: Yossi Shem Tov, Shimron Ifrah

Supervisors: Dr. Renata Avros, Prof. Zeev Volkovich

BRAUDE College of Engineering, Karmiel

1. Motivation

Mapping the evolution of scientific domains is essential to understand emerging trends and shifts in research focus.
Traditional methods of analyzing citation networks are limited in scalability and clarity. This project aims to provide a scalable and interpretable system to track structural change in academic citation networks over time, helping researchers make informed decisions.



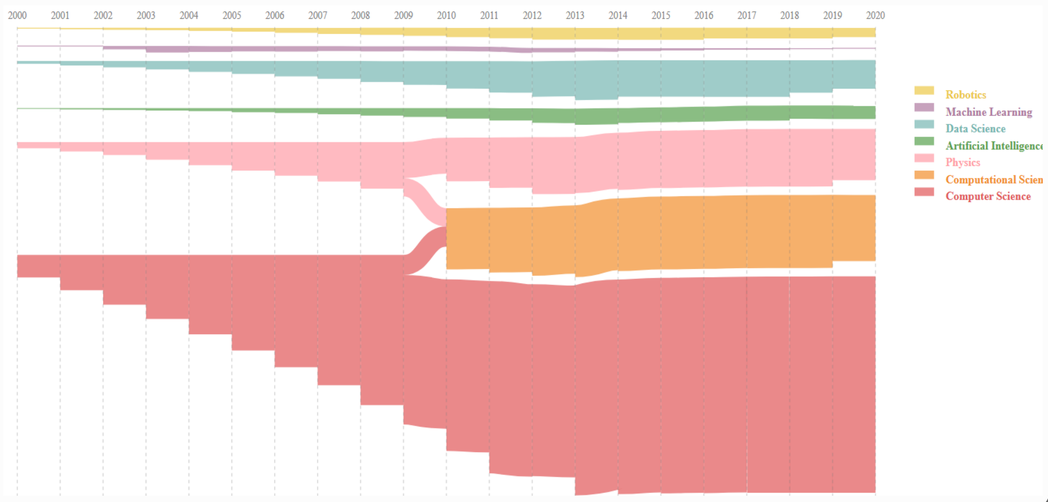
2. Proposed Solution

We developed a modular pipeline composed of:

- **Data Filtering**: Using metadata from Open Academic Graph (OAG v2.1), we filtered papers by years and Fields of Study (FOS)
- **Semantic Classification**: Applied Sentence-BERT embeddings for improved label consistency
- **Graph Embedding**: Used GraphSAGE++ to create embeddings from citation networks
- **Clustering**: MiniBatch KMeans for efficient clustering across time slices
- **Visualization**: Sankey-based Alluvial Diagrams illustrate the temporal flow of research clusters
- **User Interface**: A Tkinter-based GUI allows parameter configuration and dynamic result generation

3. Performance and Achievements

- The system successfully analyzed 20 years of academic citation data.
- Captured cluster dynamics such as emergence, merging, and decay (e.g., a "Computational Science" cluster formed from "Physics" and "Computer Science").
- Visual output clearly demonstrated topic evolution over time.
- Achieved interpretable, repeatable, and user-configurable analysis through the GUI.
- Alluvial diagrams provided visual insights into the academic landscape’s structural changes.



4. Challenges and Solutions

- **Data Scale**: Handled large datasets efficiently using JSONL streaming and pre-filtering by year and domain.
- **Clustering Stability**: Replaced HDBSCAN with MiniBatch KMeans for faster convergence.
- **Semantic Noise**: Sentence-BERT improved label relevance and clustering quality.
- **Flow Visualization Sensitivity**: Introduced Citing Percent and Cited Ratio thresholds to tune visual flow accuracy.
- **Interface Design**: Created a modular GUI supporting validation and reuse.

5. Conclusion

The project achieved its objectives by providing an efficient and interpretable system for analyzing topic evolution in citation networks. The results aligned with expectations, showing the potential for this tool in scientific trend analysis. The modular architecture and interactive interface offer a robust foundation for future research and development.