Co-Citation Network Analysis: A Community Detection Approach

Murat Yusuf Taze

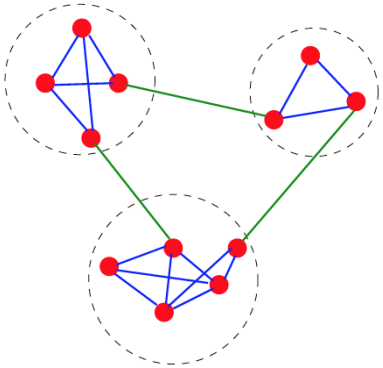
Outline

- Community Detection
- Applications
- Co-Citation Network
- Motivation, Problem Definition
- ACM Dataset
- Study Plan

Community Detection

- Community is a set of nodes densely connected to each other and sparsely connected with the rest of the network
- Community detection

  a.k.a. grouping, clustering, finding cohesive subgroups
  - Given: a social network
  - Output: community membership of (some) actors

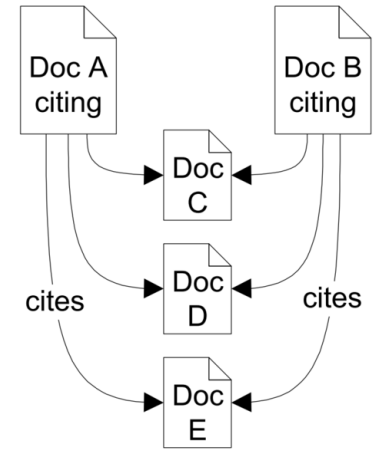


Applications

- Understanding the interactions between people
- Visualizing and navigating huge networks
- Website mirror server assignment
- Recommendation system
- Centrality Analysis/Influence Study
- Network hierarchy inference

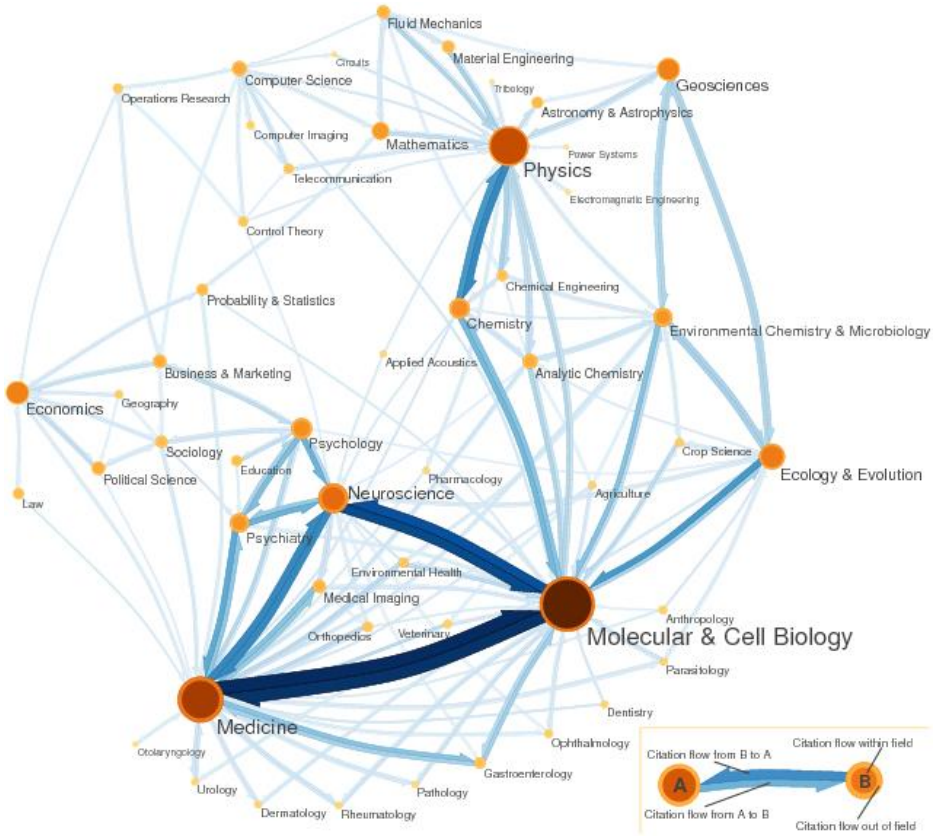
Co-Citation Network

- Co-citation is defined as the frequency with which two documents are cited together by other documents.
- If at least one other document cites two documents in common these documents are said to be co-cited.



Citation Network

- computed using citation data from the Thomson Reuters Journal Citation Reports (JCR).



Motivation

- Do scientists tend to cite friends and research colleges more frequently?
- Matthew effect in science?
  - Classics tend to receive even more reputation?
- Evolution of disciplines
- Problem: Can we apply community detection algorithms to find answers to these questions?

Dataset

- ACM 2010-05-15
- 629,814 papers and >632,752 citation relationships

Format:

- #* --- paperTitle
  #@ --- Authors
  #t ---- Year
  #c --- publication venue
  #index 00---- index id of this paper
  #% ---- the id of references of this paper (there are multiple lines, with each indicating a reference)
  #! --- Abstract

Study Plan

- Scanning Literature
- Implementation of Community Detection Algorithm
- Evaluation Results

References

- B. M. Althouse, J. D. West, T. C. Bergstrom, and C. T. Bergstrom (2009) Differences in impact factor across fields and over time
- S. Fortunato. Community detection in graphs. Physics Reports, 486:75{174, February 2010.
- S. E. Ahnert, Generalised power graph compression reveals dominant relationship patterns in complex networks, Nature Scientific Reports 4, 2013
- B. Gipp and J. Beel, “Citation Proximity Analysis (CPA) – A new approach for identifying related work based on Co-Citation Analysis,” in Proceedings of the 12th international conference on scientometrics and informetrics (issi’09), Rio de Janeiro (Brazil), 2009, pp. 571-575.