COLLECTING AND MAPPING FACULTY COLLABORATION DATA

Mingzhu Zhu, Dr. Brook Wu, Dr. Nancy Steffen-Fluhr, Dr. S. Roxanne Hiltz, Dr. Katia Passerini, Dr. Anatoliy Gruzd, Regina Collins

Data collection from Scopus
Scopus author search
First name, last name
Affiliation filtering
Common names
Pub list fetcher
Extract title, year, source title...
Pub detail page fetcher
Extract bibliographic data
 mapping authors with their affiliation information
Duplicate removal
Author affiliation mapping
Author list
Pub list
Pub detail
Fig 1. Overview of co-authorship database development from Scopus
A list of NJIT faculty names was used to carry out Scopus author search. Fortunately, no NJIT faculty members have common names. We use affiliation information to filter out those authors who come from other organizations but have the same name with NJIT faculty members.
Co-authorship database was developed by extracting the bibliographic data for each faculty member.

Data collection from Google scholar
A list of NJIT faculty names was used to carry out Google scholar’s author search, and 63,937 raw search hits were collected. For the co-authorship database, only the publications with at least two NJIT authors qualify for inclusion.

Co-authorship Database development using Google scholar
An overview of co-authorship database development using Google scholar is illustrated in Fig-3.

Merge three Co-authorship Databases

Co-authored publication extraction:
If a paper is co-authored by two NJIT faculty members, it should appear in two different publication lists. For instance, suppose there are two authors A and B, two publication lists named Pa and Pb are denoted:
Publication list for author A: Pa={p(a)1, p(a)2, p(a)3, …… p(a)n}, and
Publication list for author B: Pb={p(b)1, p(b)2, p(b)3, …… p(b)m}.
If a publication p is coauthored by A and B, then p ∈ Pa and p ∈ Pb.
2,043 coauthored publication candidates were identified. After removing the duplicates, only 1,914 pubs were left.

Co-authorship network visualization

Fig 4. Co-authorship network created using keywords “Art”