

Extraction and Analysis of Citation Data from Student Output in Order to Improve Library Instruction

Discerning Disciplinary Deviance

Institutions at Mid Sweden University (Miun) have voiced concerns that students might be deviating too much from their field of study when writing theses. They have also desired insights regarding the sources which influence students.

We have explored a method which allows extraction of references for analysis from student papers stored on our institutional repository. Our corresponding researchers' citation data has been used to evaluate student citation patterns.

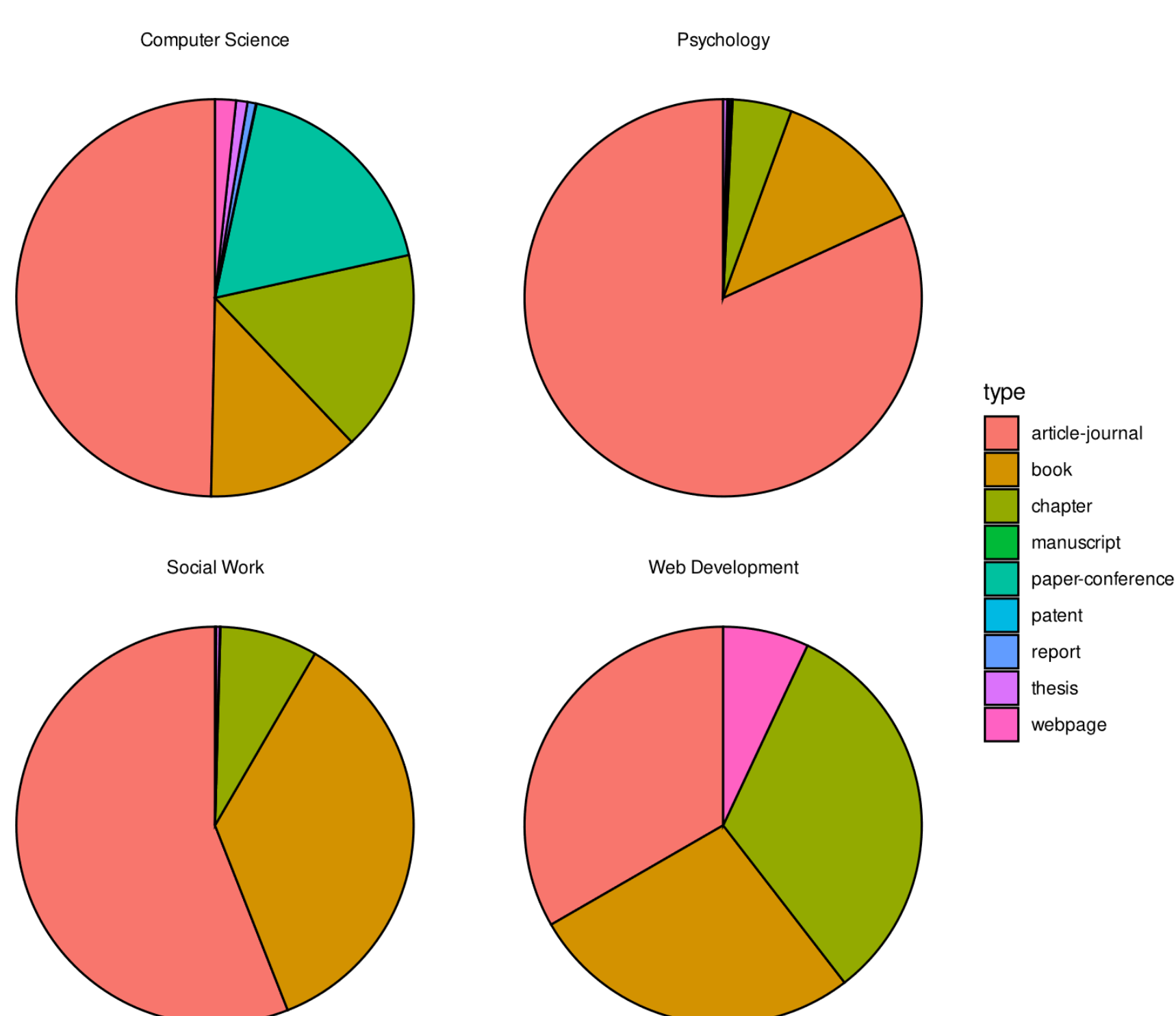


Fig. 1. Student use of sources. 4 different programs at Miun.

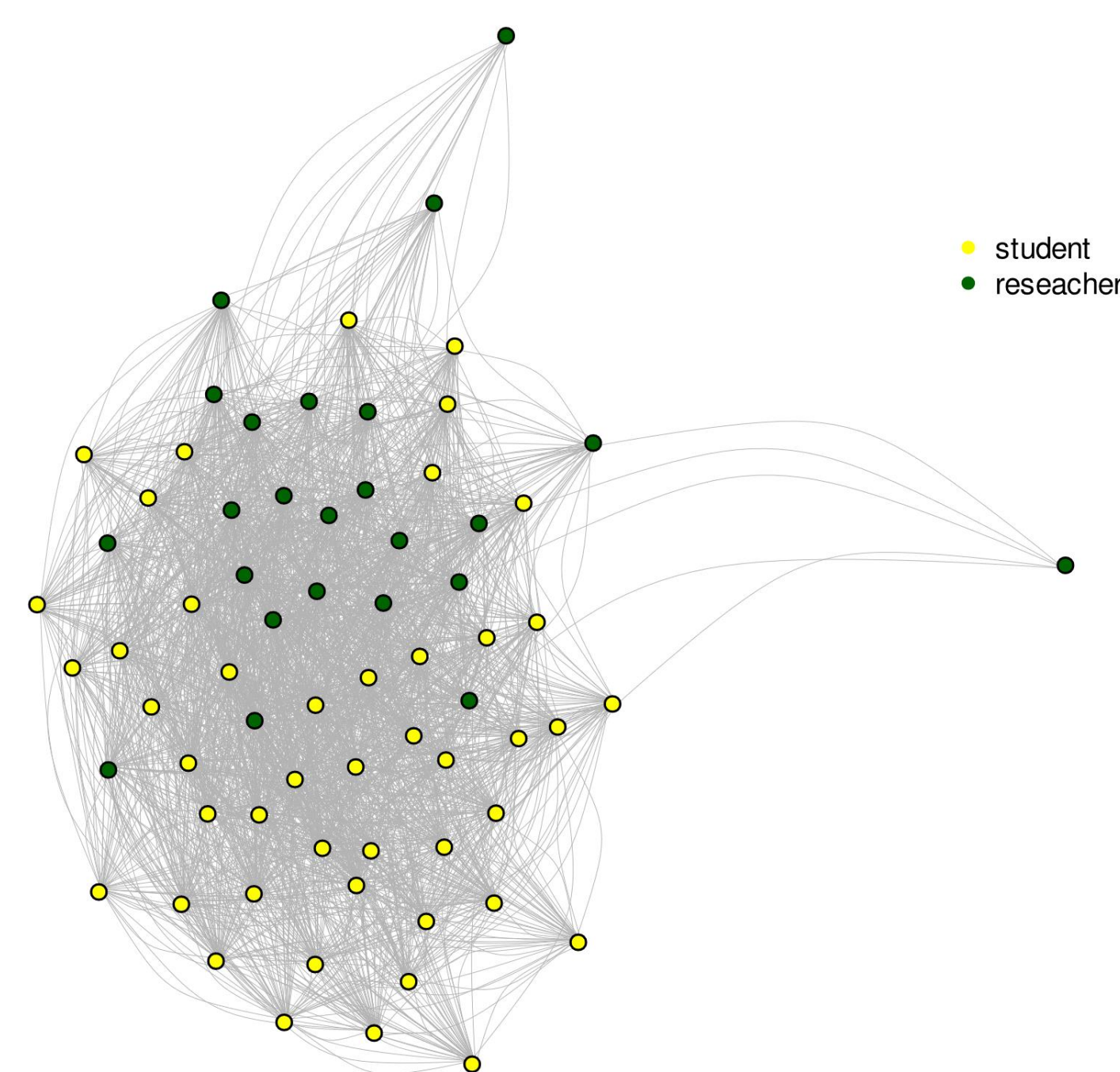


Fig. 2. Bibliographic coupling (cited author): Tourism Studies at Miun.

Extraction of Sources Used by Students

The R environment provides tools to download papers, extract and parse references by calling *anystyle* (Keli, 2023), compile datasets, analyze, and visualize the results using the *igraph* R package (Csardi & Nepusz, 2006).

We have also used the OpenAlex API, via *openalexr* (Aria & Lee, 2023), to harmonize journal names and collect researcher citation data.

Quality of student bibliographies affect data extraction success rate considerably.

Analysis of Sources in Relation to Researcher Output

Data can be used to e.g.:

- gain a better understanding of which publication types students use,
- obtain insights regarding how well-aligned student use of sources is with researcher use via bibliographic coupling,
- visualize which journals students cite by co-citation analysis and compare with researcher citing frequencies.

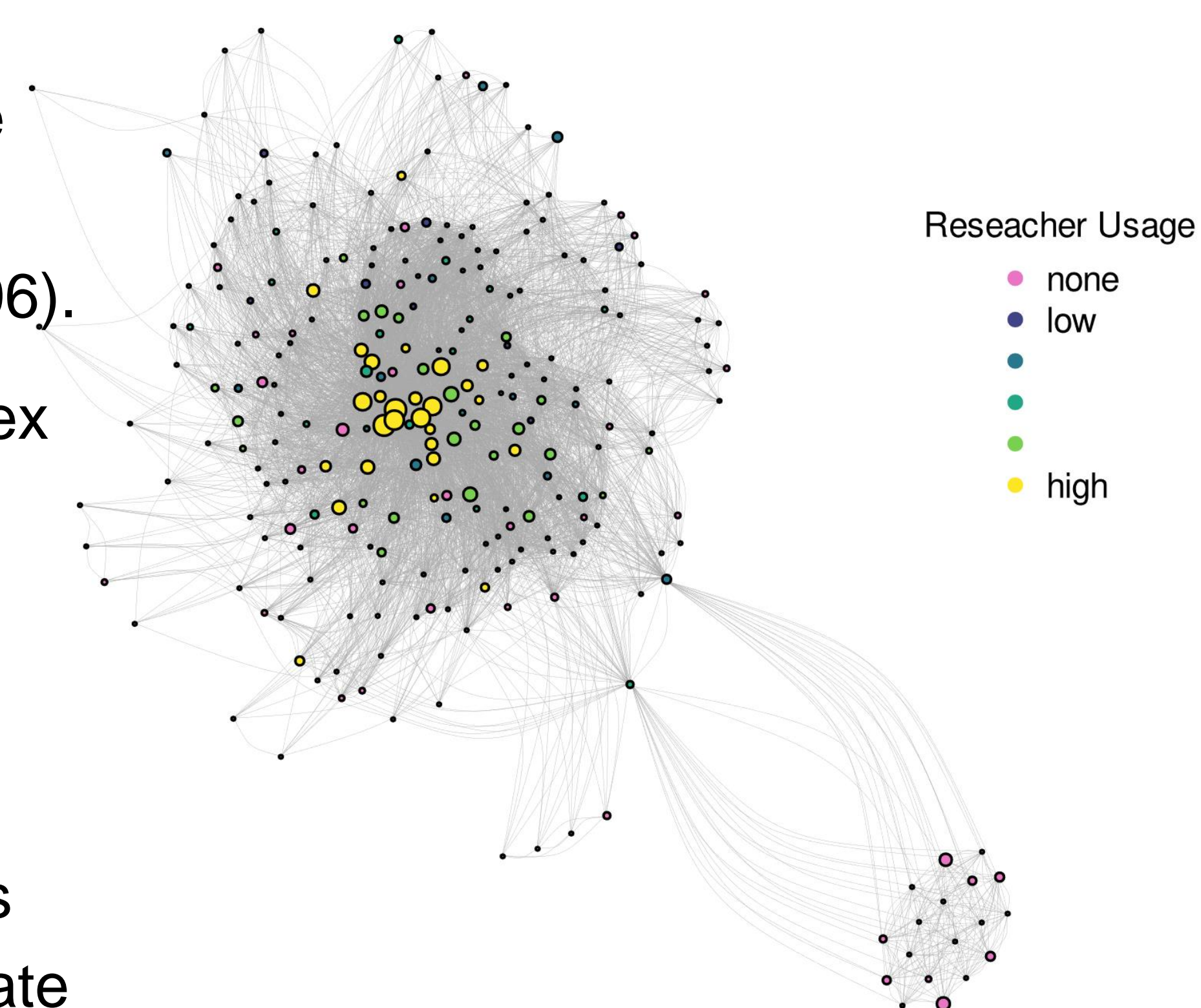


Fig. 3. Co-citation analysis: Tourism Studies at Miun. Journals cited by students and researcher use. Node color indicates researcher usage, node size indicates student usage.

Implications for Librarians and University Teachers

- Library instruction can be revised using findings.
- Findings can be employed in course development and assessment making the Library an active partner.
- Inter-disciplinary influences on student output can be visualized.
- Differences between student citation patterns and researcher patterns can be detected and studied.

References

- Aria, M., & Lee, T. (2023). *openalexR: Getting Bibliographic Records from "OpenAlex" Database Using "DSL" API* [R package]. <https://github.com/ropensci/openalexR>.
- Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, 1695. <https://igraph.org>
- Keil, S. (2023). *anystyle: Finds and parses bibliographic references* (1.1.0) [Ruby]. <https://github.com/inukshuk/anystyle-cli>