

Mapping the Field of Educational Data Sciences (Master Thesis)

Kamila Misiejuk

kamilamisiejuk@yahoo.de

Department of Library and Information Science, Humboldt University of Berlin
Supervisors: Prof. Barbara Wasson, Prof. Vivien Petras

This research takes a scientometric approach to provide an overview of the EDS field by drawing on three research communities: educational data mining, learning analytics, and learning at scale. This particular selection was motivated by the three fields being mature enough to organize their own conference: the International Conference on Educational Data Mining (EDM), the International Learning Analytics and Knowledge (LAK), and the ACM Conference on Learning at Scale (L@S).

The research comprises desk research and data analysis. As depicted in the Figure 1, the first stage of the data analysis is data collection, that is, mining all available papers from EDM, LAK, and L@S conference proceedings. The second step is the data preparation of four datasets: citation dataset, keyword dataset, author dataset, and co-authorship dataset. Four analyses and visualizations using R programming language are performed in three steps: citation analysis, keyword analysis, author analysis, and co-authorship network analysis. In the final step, the results from all four analyses are compared, across conferences. Figure 2 presents the datasets used in each analysis.

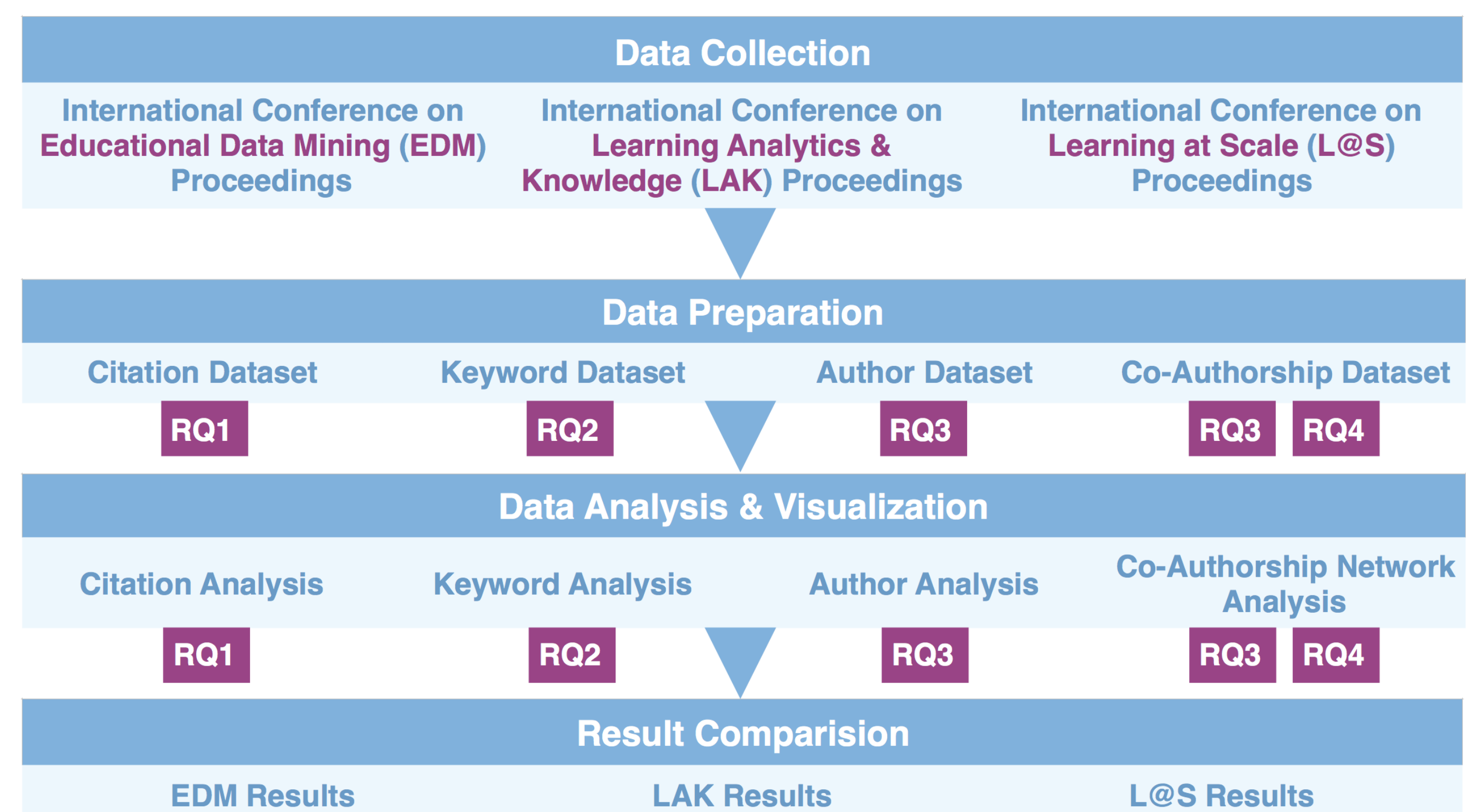


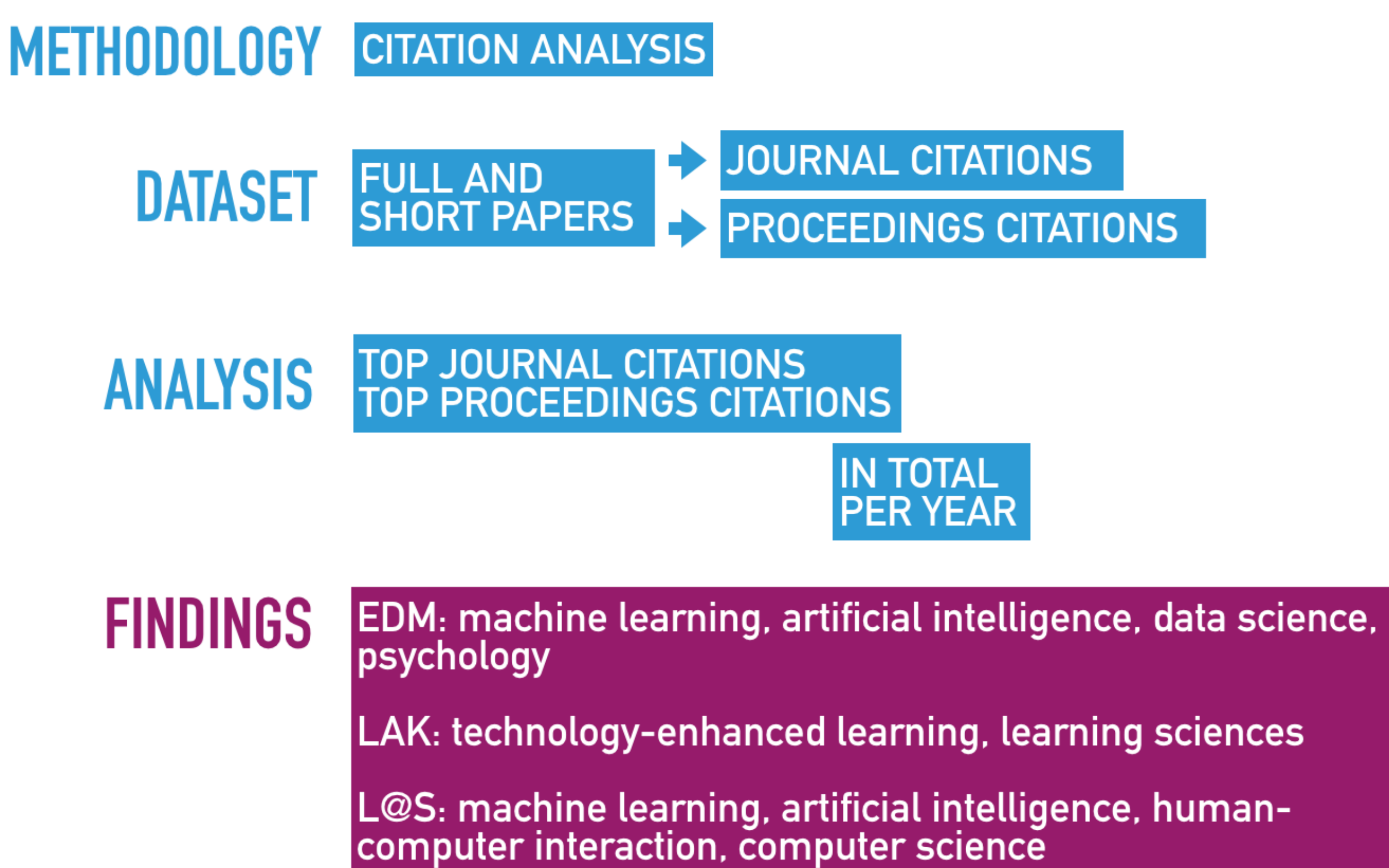
Figure 1: Data analysis flow with research questions

	EDM Proceedings (2008-2016)	LAK Proceedings (2011-2016)	L@S Proceedings (2014-2016)
Total Number of Papers	776 papers	365 papers	216 papers
Citation Dataset	404 papers ¹ 52% ² 5,348 citations	246 papers ¹ 67% ² 3,765 citations	55 papers ¹ 25% ² 892 citations
Keyword Dataset	471 papers ³ 60% ⁴ 2,052 keywords	354 papers ³ 97% ⁴ 1,708 keywords	190 papers ³ 88% ⁴ 862 keywords
Author Dataset	2,498 authors	1,299 authors	777 authors
Co-Authorship Dataset	696 papers ⁵ 89% ⁶ 1,244 authors	334 papers ⁵ 92% ⁶ 747 authors	199 papers ⁵ 92% ⁶ 555 authors

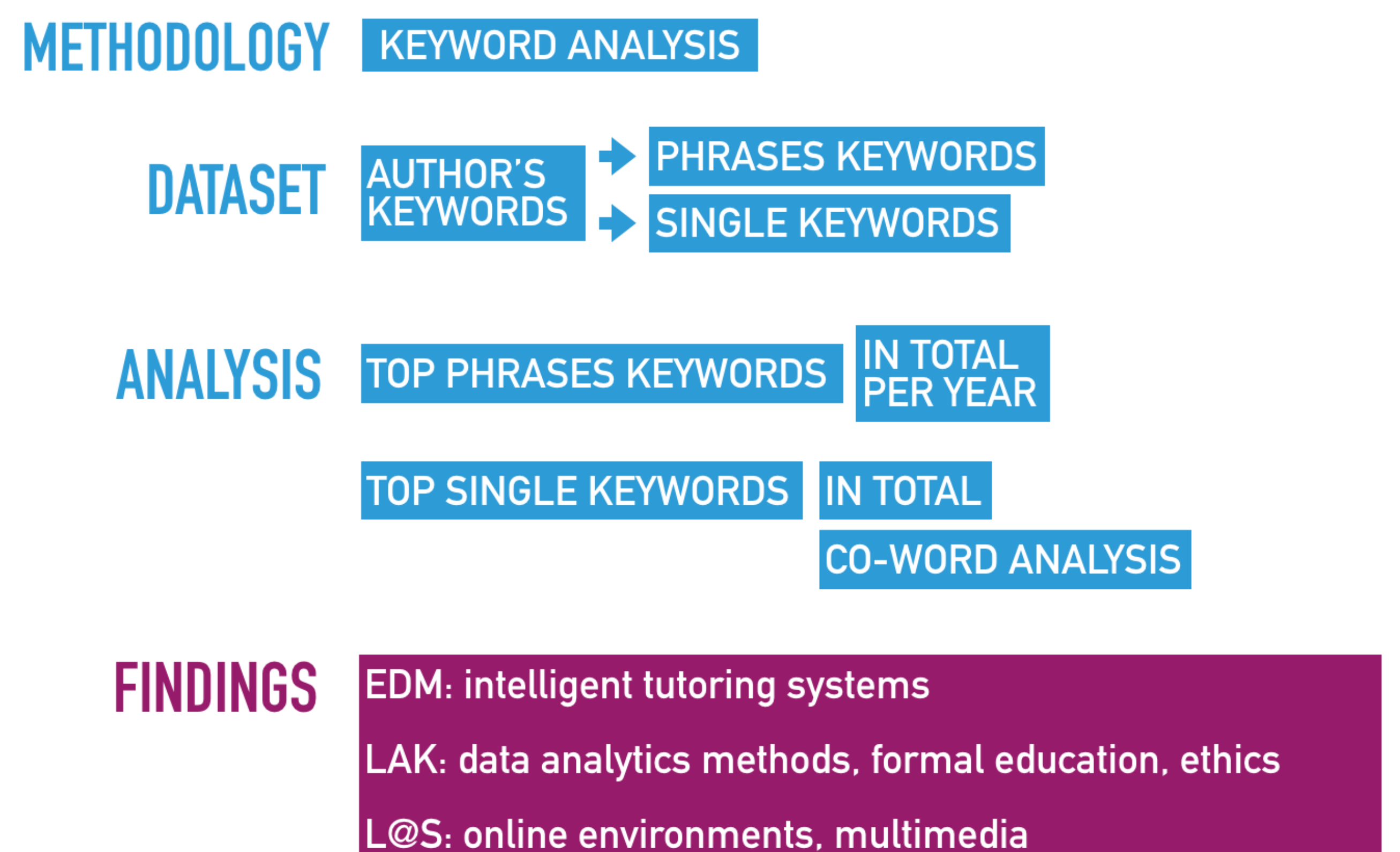
¹ Frequency of short and full papers
² Percentage of the short and full papers
³ Frequency of papers with keywords
⁴ Percentage of papers with keywords
⁵ Frequency of co-authored papers
⁶ Percentage of co-authored papers

Figure 2: Datasets

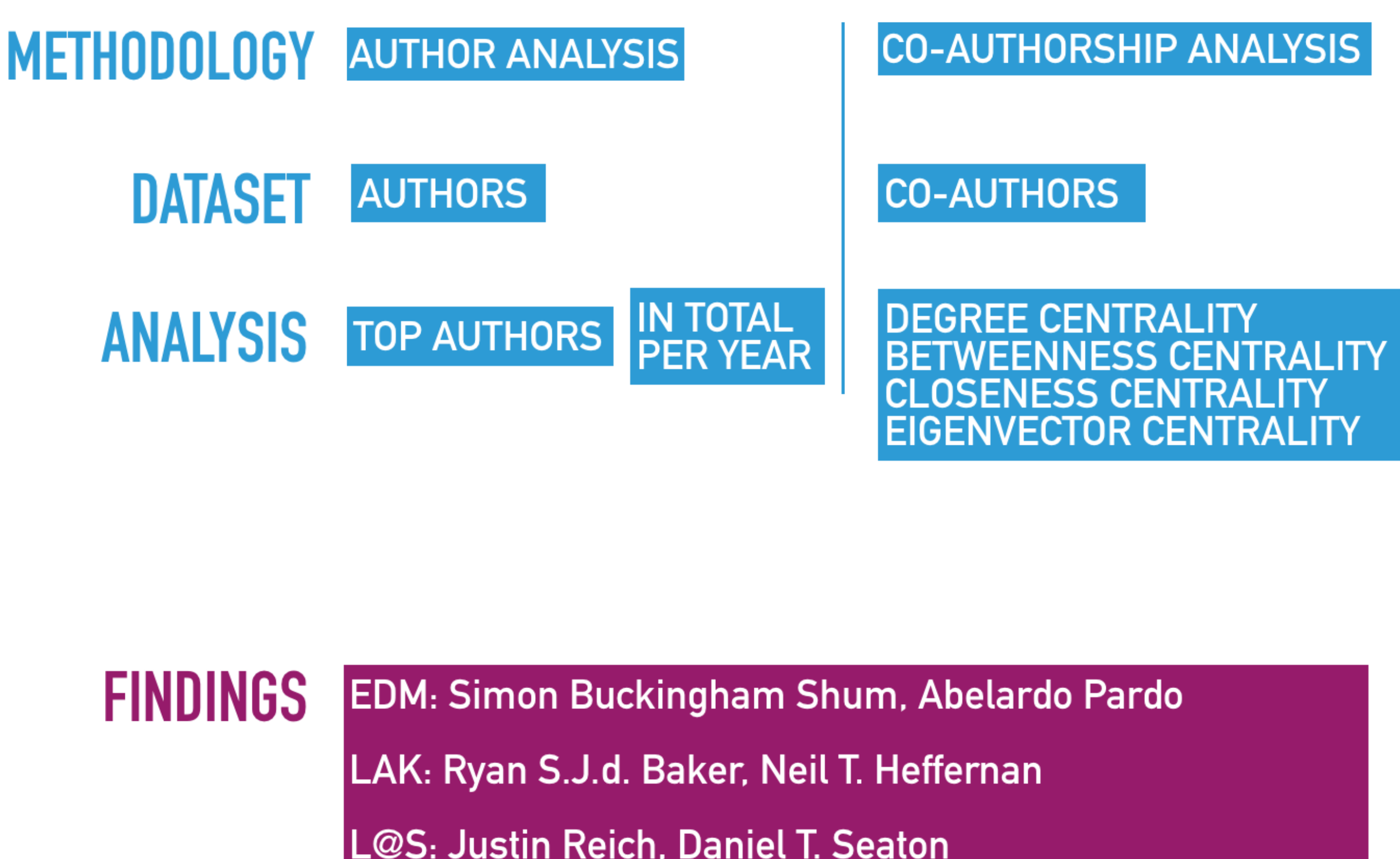
RQ1: WHICH DISCIPLINES INFLUENCE EDM, LAK, AND L@S?



RQ2: WHAT ARE THE MOST RESEARCHED TOPICS WITHIN EDM, LAK, AND L@S?



RQ3: WHO ARE THE MOST IMPORTANT AUTHORS WITHIN EDM, LAK, AND L@S?



RQ4: WHAT COLLABORATION NETWORKS EXIST WITHIN EDM, LAK, AND L@S?

