DSSV 2023 Data Science, Statistics & Visualisation
ECDA 2023 European Conference on Data Analysis

Analyzing a citation network with a billion arcs

T. Koch[1,2], H. Peng[3], and Y. Chen[3,4,5]

[1]Zuse Institute Berlin, Applied Algorithmic Intelligence Methods Department, Takustr. 7, 14195 Berlin, Germany
[2]TU Berlin, Chair of Software and Algorithms for Discrete Optimization, Str. des 17. Juni 135, 10623 Berlin, Germany
[3]Department of Mathematics, National University of Singapore, Block S17, Level 4, 2 Science Drive 2, 117543 Singapore
[4]Risk Management Institute, National University of Singapore, 21 Heng Mui Keng Terrace, 04-03, 119613 Singapore
[5]Asian Institute of Digital Finance, National University of Singapore, 3 Research Link, 117602 Singapore

What would the graph of all publications and their references look like? What insights could we gain from analyzing it? We analyzed data from the World of Science from 1981 to 2015 and will present some statistics and discuss their relevance. An important question to consider is whether any observed results could be attributed to bias or randomness in the collected data. We introduced the Article's Scientific Prestige (ASP) metric [1], which uses eigenvector centrality to measure the scientific impact of individual articles. ASP accounts for both direct and indirect citations and provides a steady-state evaluation across different disciplines. Our findings indicate that ASP and #Cit, a metric based solely on the number of citations, do not align for most articles, with a growing discrepancy among less-cited articles. While both metrics are reliable for evaluating the prestige of articles such as Nobel prize-winning articles, ASP tends to provide more persuasive rankings than #Cit when the articles are not highly cited. Finally, we suggest some additional ideas for analyzing citation data.

[1] Y. Chen, T. Koch, N. Zakiyeva, K. Liu, Z. Xu, C. houh Chen, J. Nakano, and K. Honda, "Article's scientific prestige: Measuring the impact of individual articles in the web of science," *Journal of Informetrics*, vol. 17, no. 1, p. 101379, 2023.