

RANDOM WALK IN THE ACADEMIC PUBLICATION

A data-driven approach

Ahmad Reza Ehyaei



Contents

Introduction	2
Affiliations	2
Authors	5
Citation Overview	5
Number of Publication	8

Introduction

Open Academic Graph (OAG) is a large knowledge graph unifying Microsoft Academic Graph (MAG) and AMiner. Two projects' major purpose is to develop a heterogeneous graph comprising scientific publication records and citation linkages between those articles, as well as authors, institutions, journals, conferences, and fields of study. After unifying the two datasets, the result can be found [here](#).

The dataset contains many tables, as:

- authors: includes some information about the authors, such as their names, affiliations, positions, and a summary of their publications.
- affiliations: contains the name of affiliation, type, and location by latitude and longitude.
- venues: shows information about where work was published.
- papers: this table has many fields and contains publication metadata such as title, publisher, issn, and...
- links: contains id tables for joining each table from MAG to AMiner.

If you have an Azure account, Microsoft has automated uploads of new versions of MAG to Azure Storage.

In this project, we can try to find new insights from the academic publication world!

Affiliations

The affiliation table contains information about where the papers were submitted. This table contains 25708 rows. In the below table, the fields are shown.

Affiliation Table Schema

Field Name	Field Type	Description	Example
id	string/long	affiliation id	136199984
DisplayName	string	affiliation name	Harvard University
NormalizedName	string	normalized affiliation name	harvard university
type	string	affiliation type	university
url	string	official page	http://www.harvard.edu/
WikiPage	string	Wikipedia page	http://en.wikipedia.org/wiki/Harvard_University
Latitude	string	latitude	42.37444
Longitude	string	longitude	-71.11694

Description is taken from AMiner

Top 10 countries with the most academic organizations

country	Frequency
USA	8549
UK	1836
India	1767
Japan	1057
China	831
Canada	817
Germany	678
Australia	516
France	515
South Korea	338

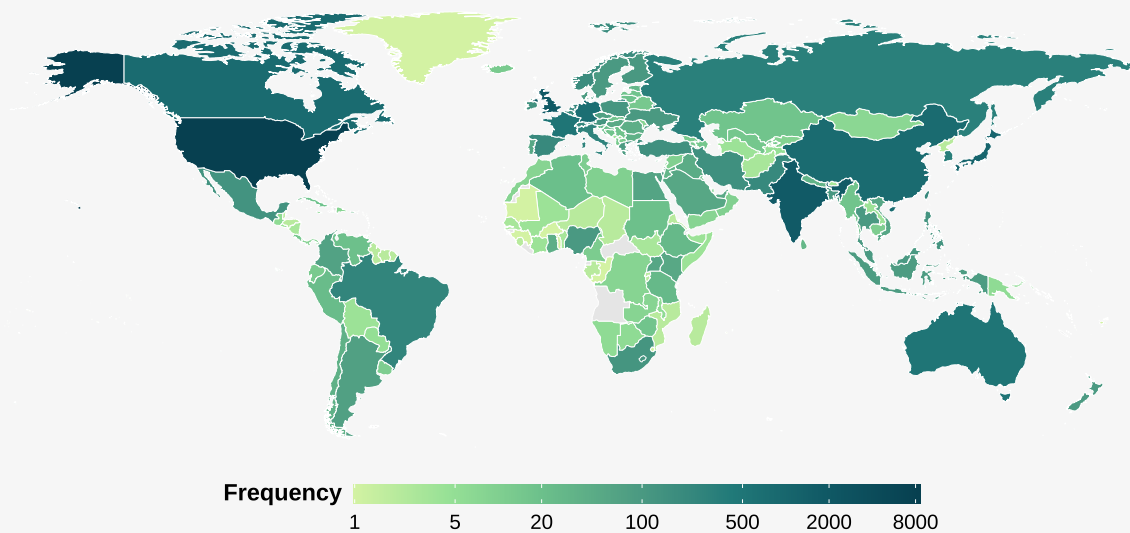


Figure 1: The distribution of academic organizations on the world map.

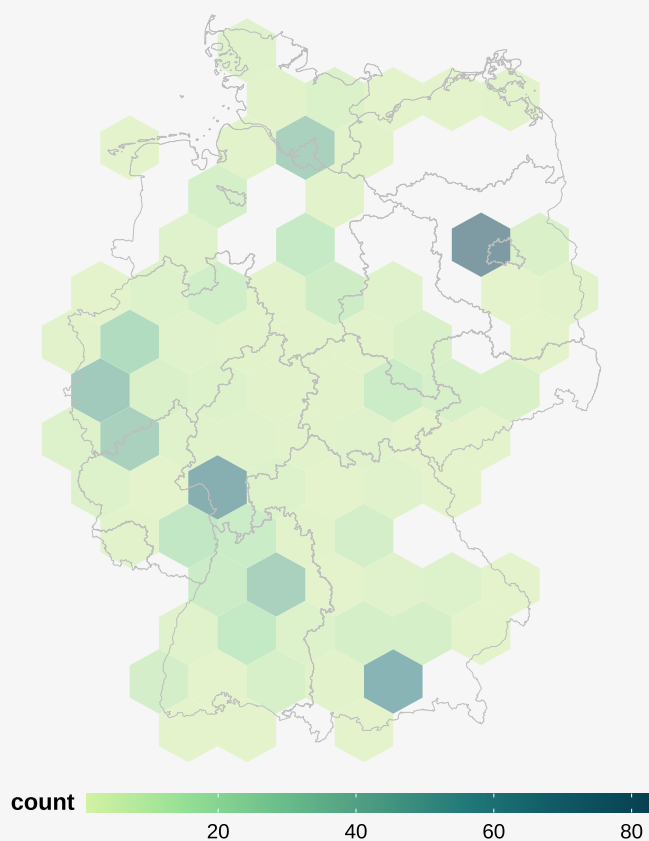


Figure 2: The location of academic organizations in Germany.

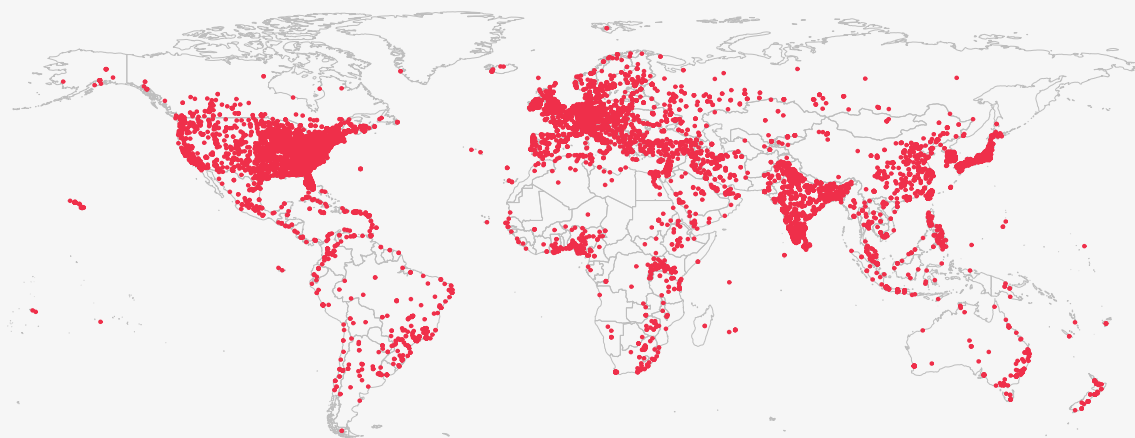


Figure 3: Location of academic organization on world map.

Authors

Fields for author information, academic history, and impact measures are included in the authors' table. This table contains 253144301 rows. The fields are listed in the table below.

Authors Table Schema

Field Name	Field Type	Description	Example
id	string/long	author id	53f42f36dabfaedce54dcd0c
name	string	author name	Jiawei Han
normalized_name	string	normalized author name	jiawei han
orgs	list of strings	author affiliations	["Department of Computer Science, University of Illinois at Urbana-Champaign"]
org	string	author organization	Department of Computer Science, University of Illinois at Urbana-Champaign
last_known_aff_id	long	last known affiliation	157725225
position	string	author position	professor
n_pubs	int	the number of author publications	1217
n_citation	int	author citation count	191526
h_index	int	author h-index	175
tags.t	string	research interests	"data mining"
tags.w	int	weight of interests	243
pubs.i	string/long	author paper id	53e9b9fbb7602d97045f7bb8
pubs.r	int	author order in the paper	0

Citation Overview

The Top 10 Frequency of Total Authors' Citations

Number of Citation	Frequency	Percent Share
0	159758931	63.1%
1	20144876	8.0%
2	11136954	4.4%
3	7359595	2.9%
4	5347158	2.1%
5	4136283	1.6%
6	3338694	1.3%
7	2791633	1.1%
8	2365922	0.9%
9	2049786	0.8%

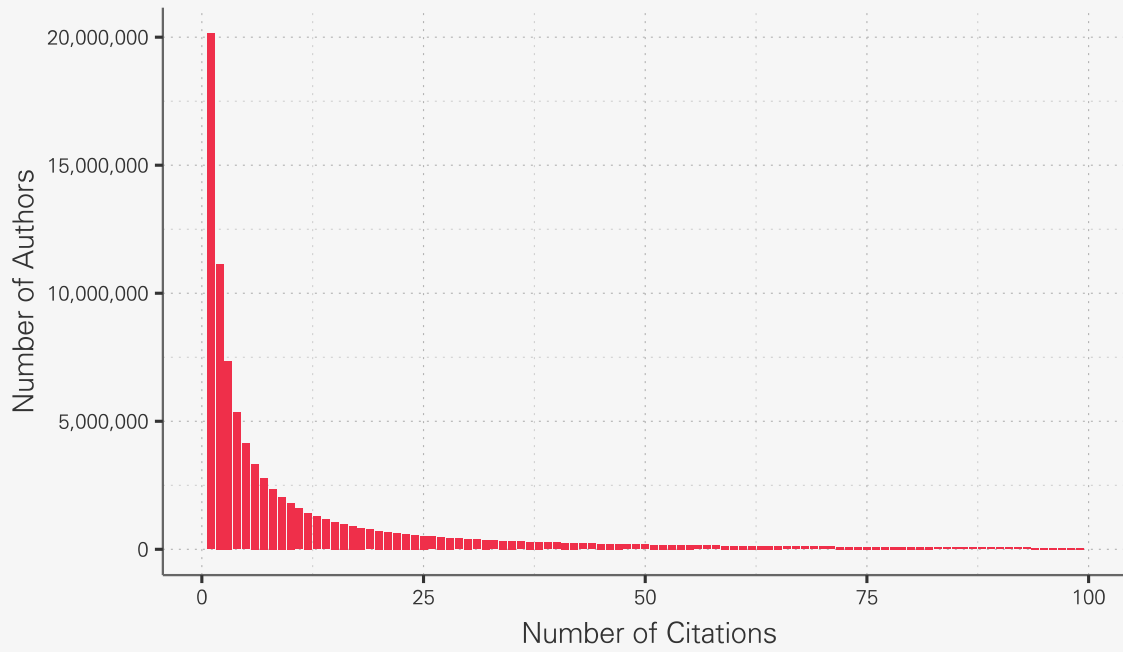


Figure 4: The Frequency of Authors by Number of Total Citations

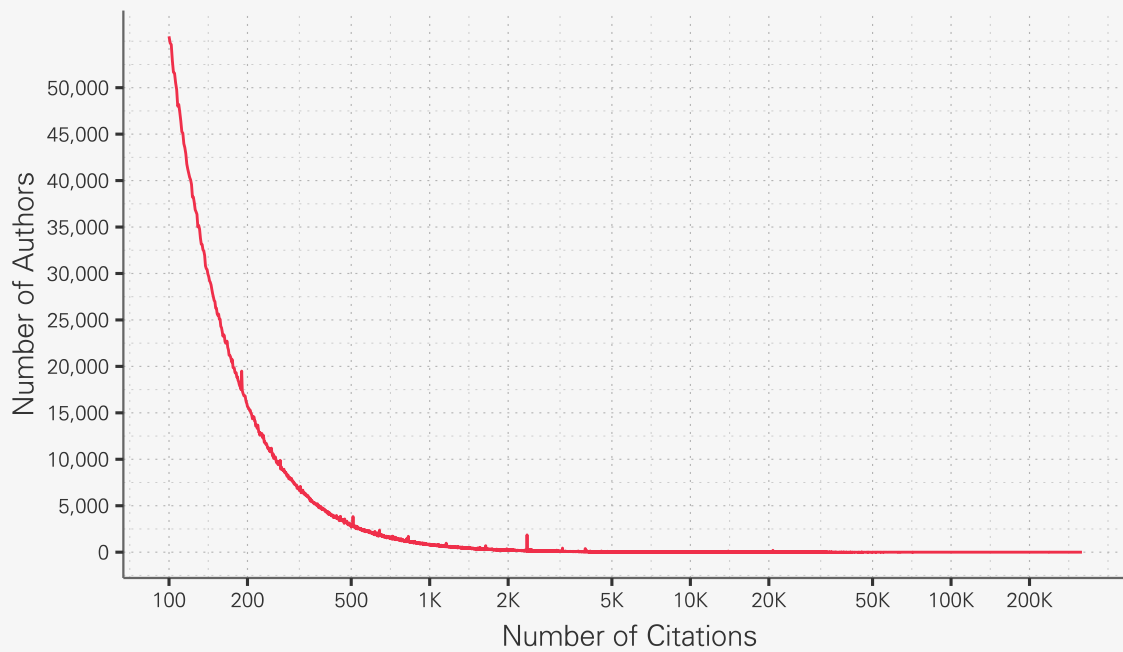


Figure 5: The Frequency of Authors by Number over Total Citations 100

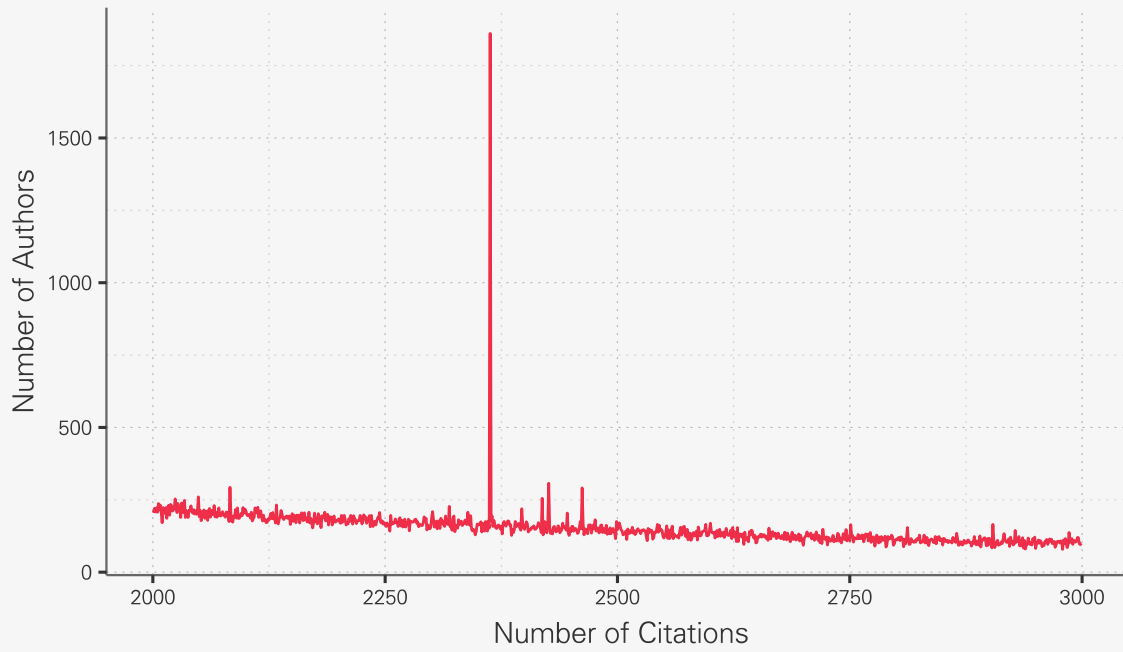


Figure 6: Strange Jump in Total Citations Density Function

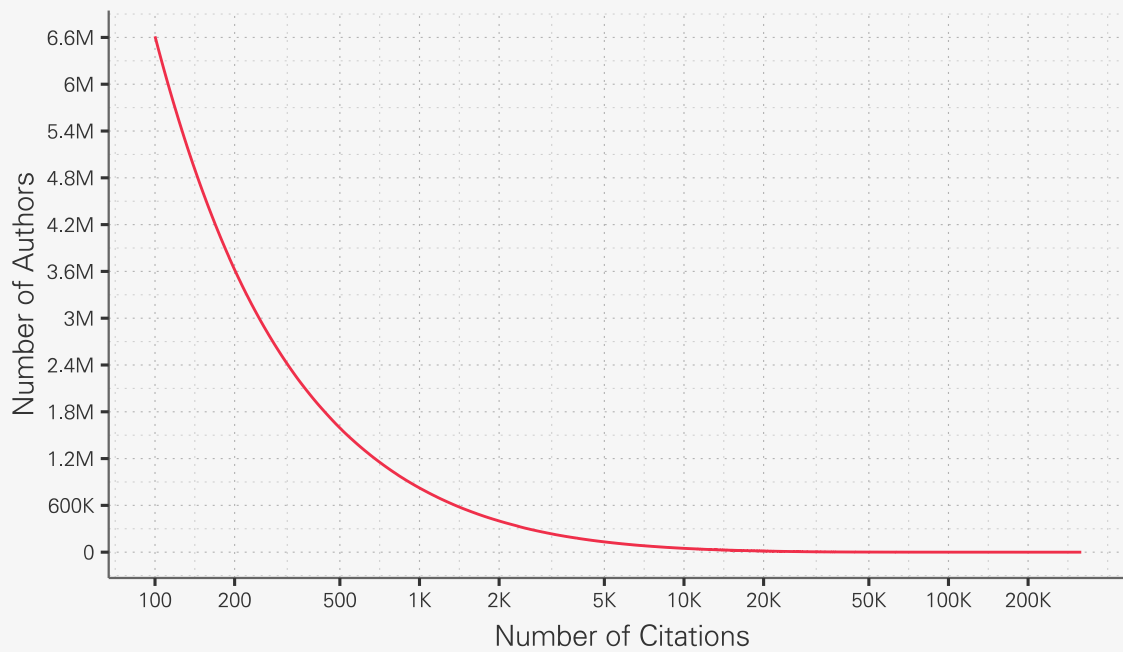


Figure 7: Tail Distribution of Authors over Total Citations 100

Number of Publication

The Top 10 Frequency of Total Number of Authors' Publications.

Number of Publication	Frequency	Percent Share
1	210702425	83.2%
2	18059438	7.1%
3	7251838	2.9%
4	3950886	1.6%
5	2477118	1.0%
6	1718516	0.7%
7	1248849	0.5%
8	954450	0.4%
9	748627	0.3%
10	605597	0.2%

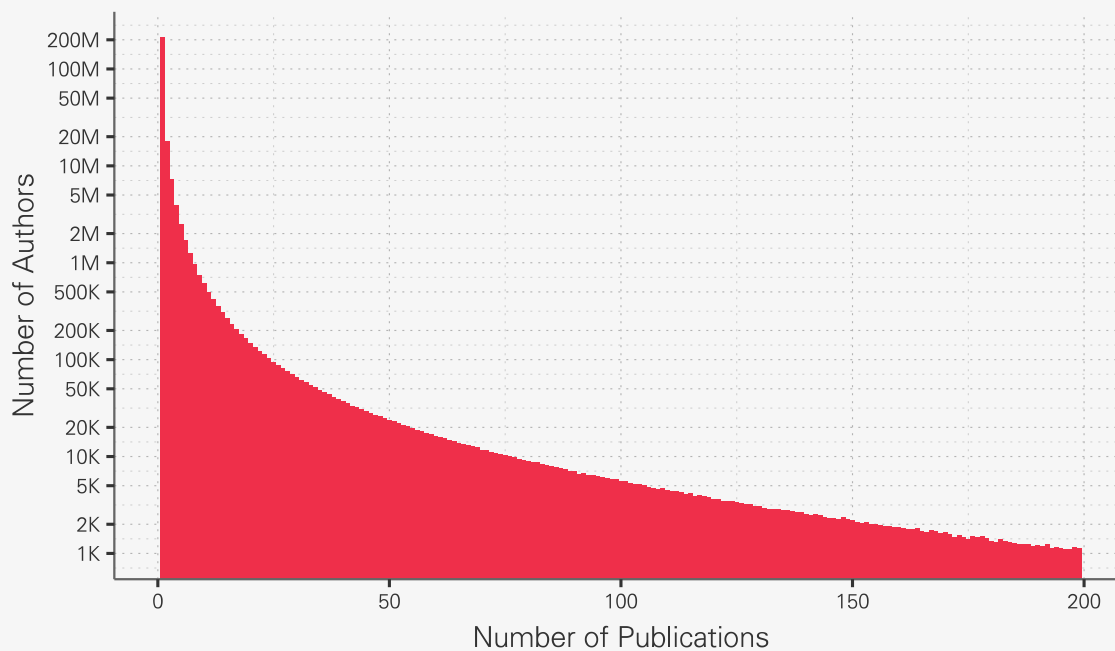


Figure 8: The Frequency of Authors by Number of Total Publications

