The dissociation of researchers from superstars through a new metric

Problem

Associating with superstar researchers tends to stifle innovation. Individualism tends to stimulate creativity instead. How can we understand this discrepancy?

Superstar - Those who are among the .1% in terms of their H-index [1]

H-Index - Count the number of publications for which the author has been cited by other authors the least number of times.

Innovation - The degree to which the document adds topics in new combination to the literature [1]. A mixture of subjectivity and objectivity that drives progress across one or various fields

The H-index works in terms of determining superstar researchers and academic notoriety but doesn't determine the dissociation of an author between academic notoriety and innovation

Can we develop a new metric to efficiently assess the effect of the dissociation of researchers from their superstar researchers?

More specifically - How can we quantify the dissociation of an author in terms of innovation and academic notoriety and how can we understand what an author is more inclined on

Background

1. Investigation of the dissociation of researchers and their superstar researchers [1]
2. The entire S2ORC corpus database of academic research papers [2]
3. How the H-index works as a reference [3]

How are these papers relevant for devising a metric?

Data Pipeline

Requirements - Extract superstars and non-superstars from the computer science related field

Modelling

Multiple Linear Regression - Build the dissociation index for each author

For each author (Y) the following formula is applied

$Y = \beta 0 + \beta 1 \times X1 + \beta 2 \times X2 + \ldots + \beta n \times Xn + \varepsilon$

We will consider the innovation side to be the innovation score and the number of publications, and the H-index and the number of times he has been cited to be the exposure side.

Ordinary Least Squares - Optimize the coefficients

Independent Variables	Author Criteria
X_1	H-index
X_2	Innovation Score
X_3	Number of publications
X_4	Number of times he has been cited

Values	β_1	β_2	β_3	β_4
Min	-0.535	0.487	0.205	-0.073
Optimal	**-0.440**	**0.534**	**0.299**	**-0.050**
Max	-0.311	0.589	0.351	-0.031

Metric Development

1. Have a clear definition of the superstar and innovation criteria
2. Perform a Multiple Linear Regression Analysis
3. Optimize the coefficients with Ordinary Least Squares
4. Consider Bias and Optimize the Result
5. Normalize the metric

Conclusion

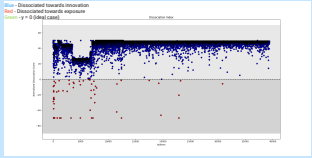
Bridge the gap between researchers and superstar researchers by developing a new metric to measure an author's dissociation. The findings suggest that while collaborating with superstar researchers can significantly boost exposure and citation impact, it may also stifle individual innovation

Finally this metric has its objective to balance the ongoing gap, and to bring fairness in quantifying authors' dissociation, by giving a more fair overview of each researcher.

Future Work

1. Improve the metric with a more extensive dataset
2. The analysis can be extended beyond the computer science field
3. Innovation is partially subjective and can always be adapted
4. Bias reduction

Graph Representation



Analysis and Discussion

Selection Bias - random sampling from the database for fair selection

Cross Validation - Use the metric on multiple random sampled datasets.

Limited scope and coverage bias - The S2ORC database may not cover everything in the Computer Science field and some authors may be underrepresented. This database may not cover all forms of scholar output such as patents, technical reports and software repositories.

Comparative Analysis - Distinguishing itself from conventional innovation metrics, the proposed metric offers a more comprehensive evaluation by considering various author-related factors outlined in the MLR model.

Responsible Research - The aim is to develop a metric that offers an impartial and unbiased assessment of innovation, free from any preconceptions or inherent biases.

TUDelft

Filip Marchidan | f.t.marchidan@student.tudelft.nl
Responsible Professor | Hayley Hung
Supervisor | Vandana Agarwal, Chenxu Hao

[1] Sean Kelty, Raiyan Abdul Baten, Adiba Mahbub Proma, Ehsan Hoque, Johan Bollen, and Gourab Ghoshal. Don't follow the leader: Independent thinkers create scientific innovation, 2023.

[2] Kyle Lo, Lucy Lu Wang, Mark Neumann, Rodney Kinney, and Daniel Weld. S2ORC: The semantic scholar open research corpus. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault, editors, Proceedings of the 58th Annual Meeting of the Associa- tion for Computational Linguistics, pages 4969–4983, Online, July 2020.

[3] Leif Engqvist and Joachim G Frommen. The h-index and self-citations. Trends in ecology & evolution, 23(5):250–252, 2008.