The Selfish Rewards of Multidisciplinarity

Nathan Franz

Abstract

Introduction

Questions:

Are papers that cite outside their own fields more successful than those that don't? "Fields that don't share a parent field?"

Are authors that publish in multiple fields more successful than those that don't?

Why this question?

Low hanging fruit The university

The syncretic act of

"Carving nature at its joints" as Plato put it in the *Phaedrus*

Mental models

Why is this area important?

This project is concerned with scientometrics – the measurement and analysis of scientific publications. (This project is actually concerned with *academic* pulications more generally, including publications in the arts and humanities.) At first glance, this meta-scientific field may seem like so much navel-gazing on the part of the scientific community. However, it is important for the purpose of evaluating the effectiveness of scientific publishing and identifying ways in which it could be better.

What is the question?

In this project, I hope to answer whether papers that cite outside their own fields are more successful than those that don't.

The metric of success used here is the number of citations as reported by the Microsoft Academic Knowledge Graph. Something like betweenness or Page rank would be preferable, but calculating such statistics for the tens of millions of papers in the sample was infeasible.

Why this question?

What other work has been done on this question?

What data is being used to answer this question?

For this project I used the Microsoft Academic Knowledge Graph, a data set featuring over 200,000,000 papers published in over 48,000 journals by over 250,000,000 authors affiliated with over 25,000 institutions.

Defining "success"

This is a hard problem.

Take a random sample of papers in each field that has a parent field but has no

Methods

What methods were used to collect the data?

The Microsoft Academic Knowledge Graph is encoded as a Resource Description Framework (RDF) graph, accessible via a SPARQL Protocol and RDF Query Language (SPARQL) end point.

What methods were used to analyze the data?

What methods were used to present the data?

The scientometric methods

I used the software packages RStudio for analysis of the data and Gephi for visualization of the networks. I used a Gephi plug-in called "Semantic Web Import" for importing the RDF data with an implementation of SPARQL Protocol and RDF Query Language (SPARQL).

Results

Conclusion

Restatement of Work

More Work to Be Done

This work could be extended with more refined measures of paper success than the crude number of citations. (Such as betweenness or Page rank.) It could also, following in the work of Heinze & Bauer (2007),

Citations

Färber M. (2019) The Microsoft Academic Knowledge Graph: A Linked Data Source with 8 Billion Triples of Scholarly Data. In: Ghidini C. et al. (eds) The Semantic Web – ISWC 2019. ISWC 2019. Lecture Notes in Computer Science, vol 11779. Springer, Cham

Heinze, T., & Bauer, G. (2007). Characterizing creative scientists in nano-S&T: Productivity, multidisciplinarity, and network brokerage in a longitudinal perspective, Scientometrics, 70(3), 811-830. Retrieved May 4, 2020, from https://akjournals.com/view/journals/11192/70/3/article-p811.xml

SCImago, (n.d.). SJR — SCImago Journal & Country Rank [Portal]. Retrieved May 5, 2020, from http://www.scimagojr.com

Appendix