Independent Thinkers and Scientific Progress

An Analysis of Superstar Influence on Computer Science Research Dynamics

1. Introduction

In the scientific community, a few prominent researchers, often referred to as "superstars," dominate attention, citations, and resources. However, it is unclear whether they promote true innovation. This study replicates and extends the work of Kelty et al. (2023) on superstar influence in scientific innovation, focusing on the field of computer science. We examine how collaboration with highly influential researchers impacts research output, diversity, and innovation. Our aim is to understand the role of superstars in shaping computer science research.

2. Methods

We analyzed 718,355 computer science papers from the Semantic Scholar database. We computed several metrics including Shannon Entropy, Citation/Reference Diversity, Innovation, and a new metric we propose called Pairwise Diversity. These metrics capture different aspects of research diversity and novelty. Our analysis focused on three main areas: (1) comparing superstars (defined as the top 0.1% researchers by h-index) to non-superstars across these metrics, (2) examining how frequently citing superstars affects researchers' output and impact, and (3) investigating the career trajectories of early-career researchers who either collaborated frequently with superstars or innovated independently.

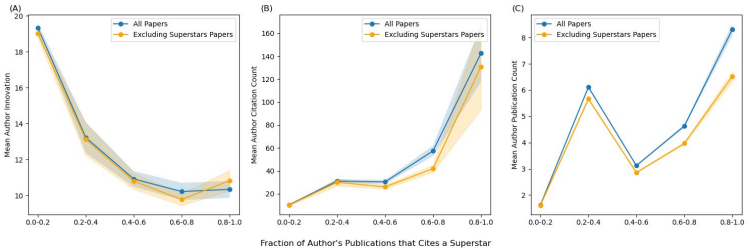
3. Results and Discussion

Our first analysis compares the performance of superstars to non-superstars across various metrics. Table 1 presents these results:

Metric	Our Work (%)	Kelty et al. (2023) (%)
Entropy (LDA)	5%	2%
Entropy (SPECTER)	2%	-
Citation Diversity (LDA)	67%	15%
Citation Diversity (SPECTER)	64%	-
Reference Diversity (LDA)	14%	20%
Reference Diversity (SPECTER)	4%	-
Pairwise Diversity (LDA)	179%	-
Pairwise Diversity (SPECTER)	450%	-
Innovation	135%	900%

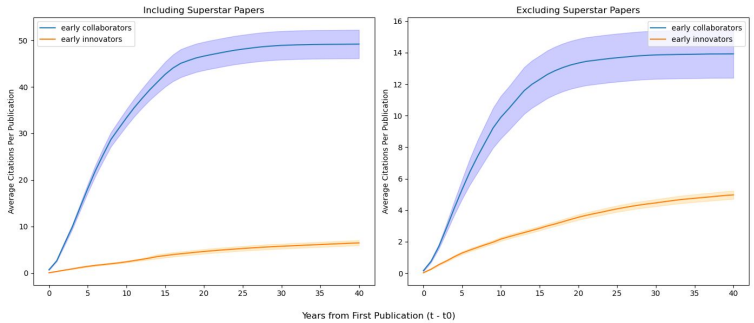
As shown in Table 1, superstars consistently outperform non-superstars across all metrics, aligning with Kelty et al.'s findings. However, the innovation gap is smaller in computer science, suggesting less concentrated innovation among top researchers. Our new Pairwise Diversity metric shows the largest effect size, indicating its potential as a robust indicator of research impact and diversity.

Figure 1 illustrates how citing superstars affects research output:



We see several important trends. Researchers who more frequently cite superstars tend to receive more citations and publish more papers. However, their work tends to be less innovative. These trends persist even when excluding papers co-authored with superstars. This suggests that the effect is not solely due to direct collaboration, but may reflect broader patterns of engagement with mainstream, highly-cited work. While this approach may boost visibility and productivity, it appears to come at the cost of originality.

Finally, we analyzed the career trajectories of early-career researchers, comparing those who frequently collaborate with superstars to those who produce highly innovative work independently. Figure 2 presents these findings:



As seen in Figure 2, early-career researchers who frequently collaborate with superstars initially receive more citations. However, when we exclude papers co-authored with superstars, this advantage diminishes significantly. In contrast, early innovators, who produce highly innovative work without relying on superstar collaborations, show a slower plateau in citation counts over time. This suggests that while collaborating with superstars can provide an initial boost, developing independent, innovative research paths may lead to more sustained impact.

4. Conclusion

Our study replicates and extends Kelty et al.'s work on the influence of superstars in computer science using the Semantic Scholar Academic Graph dataset. We found that superstars consistently outperform non-superstars in metrics such as innovation and diversity, though the innovation gap is smaller in computer science than in physics.

Frequent citation of superstars is linked to higher citation and publication counts but lower innovation. This pattern remains even when excluding papers co-authored with superstars, suggesting that researchers focusing on highly cited work may sacrifice originality.

Analyzing early-career researchers, we found that those collaborating frequently with superstars initially gain more citations but see a significant drop when excluding superstar collaborations. Conversely, early innovators maintain stable citation counts, highlighting the need to support independent researchers for long-term innovation. Our new Pairwise Diversity metric showed potential but requires further exploration. Future research should refine data preprocessing methods and validate our findings across other scientific disciplines. Overall, our study highlights the need for policies that support both superstars and independent, innovative researchers.

Filip Płonka
f.j.plonka@student.tudelft.nl
https://github.com/fplonka/cse3000
Supervisors: Hayley Hung, Chenxu Hao, Vandana Agarwal