Visualizing Collaboration with Superstars

A Novel Approach to Visualizing Collaboration | Preston Hull | 25-06-2024 | TU Delft

TU Delft

Step 1: Preparation

- We first looked at the S2ORC dataset used in the paper by Kelty (et al.)[1]
- After further examination, we found that the S2AG dataset, by the same group, is more suitable for our use-case
- We processed the authors, citations, and papers datasets into a PostgreSQL database.

Authors
- author_entry
  - Author ID
  - Author name
  - Paper count
  - Citation count
  - H-Index

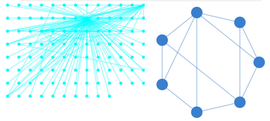
Citations
- citation_entry
  - Citation ID
  - Citing paper
  - Cited paper
  - Is Influential
- citation_context
  - Citation ID
  - Context Text
  - Intent Text

Papers
- paper_entry
  - Paper ID
  - External IDs
  - URL / Title / Year
  - Reference Counts
- paper_entry_paperauthor
  - Paper ID
  - Author name
  - Author ID
- paper_entry_paperfield
  - Paper ID
  - Field of study
  - Source

Step 2: Methodology

- With a pre-processed and queryable dataset, the next step is to produce a graph dataset with authors as nodes and papers as edges
- I wrote a script to query the pre-processed dataset into an efficient JSON graph format
- The graph and visualization use standard web programming languages, including HTML, CSS, and JavaScript, to maximize compatibility and reduce complexity
- For this visualization, the Cytoscape library was used and the *cose* (Compound Spring Embedder) force-directed graph layout was chosen to optimize for graph usability heuristics

```
{
  nodes: {
    <author_id>: {...}
  },
  edges: {
    <author_id_lower>: {
      <author_id_upper>: [
        {...}, {...}
      ]
    }
  }
}
```



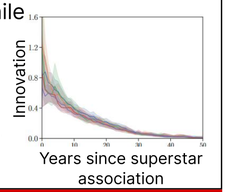
Step 3: Conclusions

- Some disciplines have far more collaborators per associate than others
- There does not appear to be a strong correlation between prominence and collaborators per associate
- This disparity visualizes in the differing graphs per discipline
- The reasons for this can be evaluated further on a per-discipline level

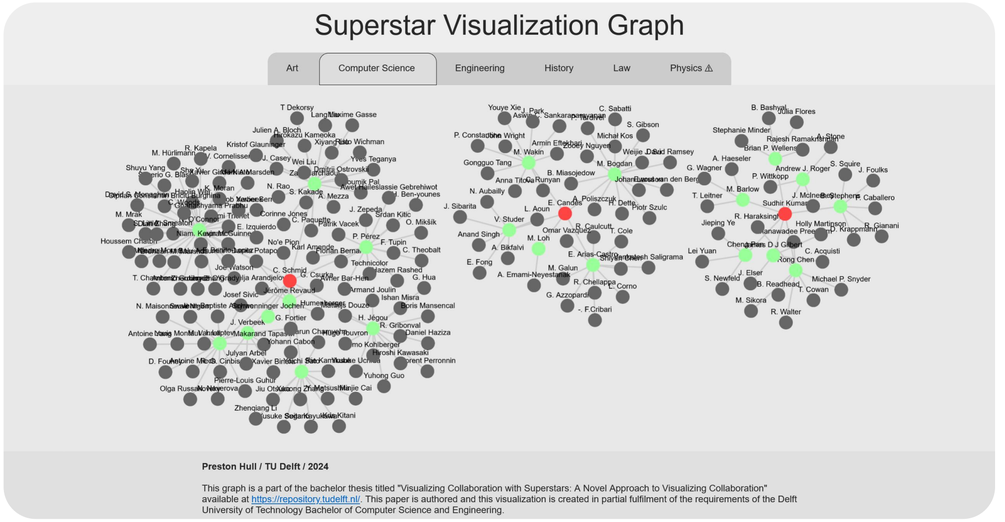
Discipline	Prominence	Associates	Collaborators	Collaborators per associate
Art	1	1	1	1
Computer Science	20.1	77	10356	134.5
Engineering	10.7	55	4281	77.8
History	1.6	24	3993	166.4
Law	2.1	56	32109	573.4
Physics	33.6	177	109698	619.8

Step 0: Introduction

- In academia, some papers are far, far more successful than average. Authors of these papers are known as “superstar” authors.
- New researchers often seek guidance and collaboration with superstars as a way to advance their career.
- Collaboration with superstars is believed to increase research output while
- decreasing innovation[1].
- This project visualizes the effect of collaboration with a superstar on the later career of an associate.



Superstar Visualization Graph



Preston Hull / TU Delft / 2024

This graph is a part of the bachelor thesis titled "Visualizing Collaboration with Superstars: A Novel Approach to Visualizing Collaboration" available at https://repository.tudelft.nl/. This paper is authored and this visualization is created in partial fulfilment of the requirements of the Delft University of Technology Bachelor of Computer Science and Engineering.

Step 4: Evaluation

- The graph is evaluated based on usability heuristics by Bennett (et al).[2]
- Overall, the graph performs well against edge heuristics at any size.
- At large sizes, the visualization struggles with node heuristics, such as maintaining a minimum distance between the nodes and preventing overlap
- The graph is dynamically generated and fits its container.

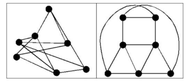


Figure 2: *The effect of applying aesthetic heuristics.*

[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. arXiv preprint arXiv:2301.02396, 2023.

[2] Chris Bennett, Jody Ryall, Leo Spalteholz, and Amy Gooch. The aesthetics of graph visualization. In CAe, pages 57–64, 2007.