CS 452 Report

David Bauch

2025-10-21

Table of contents

CS 452 Report

This report contains the results of various database queries and analyses performed as part of the CS 452 course.

1. Graph of the Bacon Path for Nora Ephron

A screen shot of the Neo4j Browser with the graph is sufficient.

```
MATCH p=shortestPath((nora:Person {name: "Nora Ephron"})-[*..999]-(keanu:Person {name: "Keanu
RETURN p
```

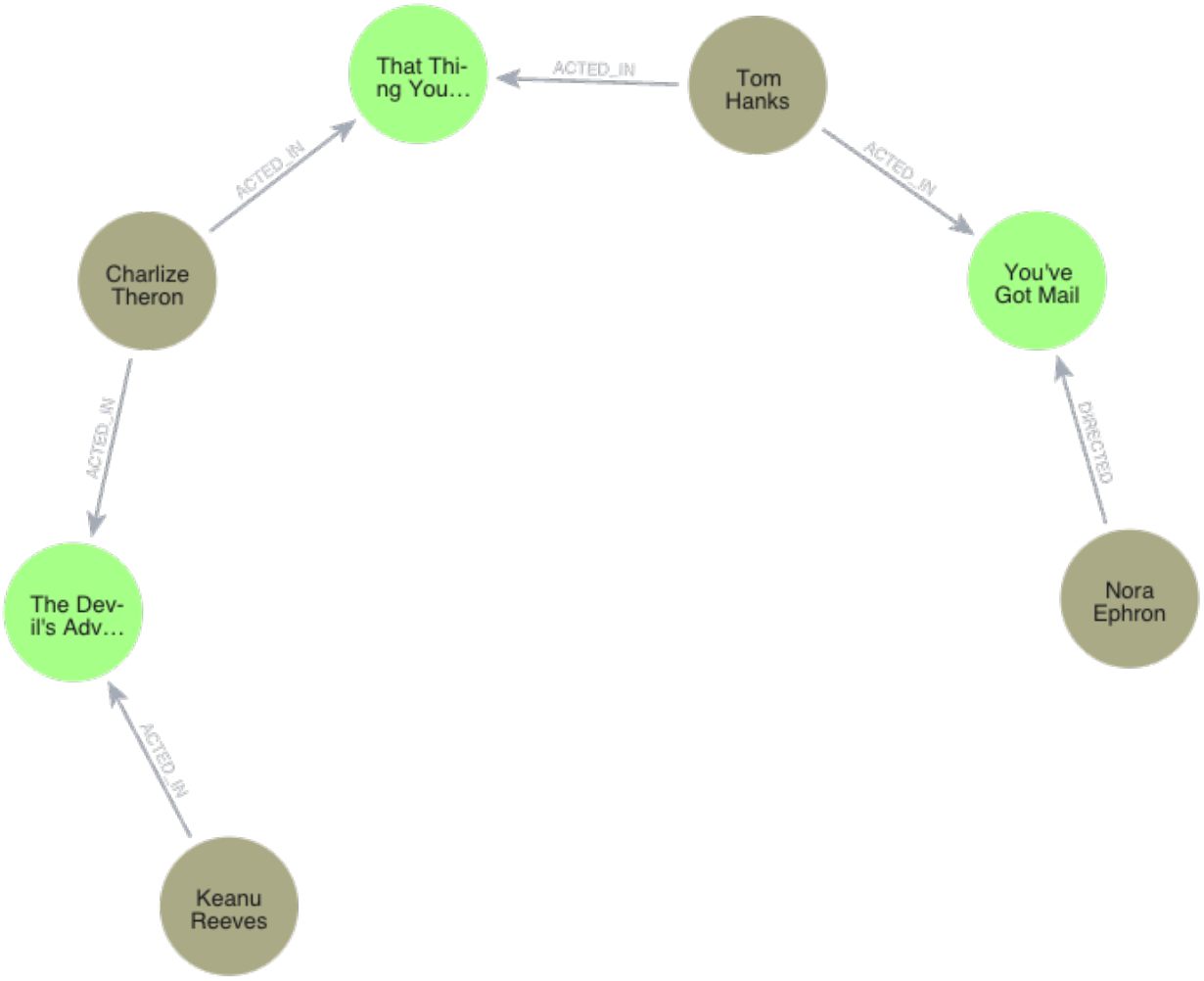

Figure 1: Bacon Path

Figure: Bacon path between Nora Ephron and Keanu Reeves in the Neo4j Movies database.

2. InDegree vs Category Name Chart

A screen shot of the inDegree vs Category Name chart from your working jupyter notebook for the ArXiv tutorial (linked at the beginning). This is a bar chart, not a graph of all the papers.

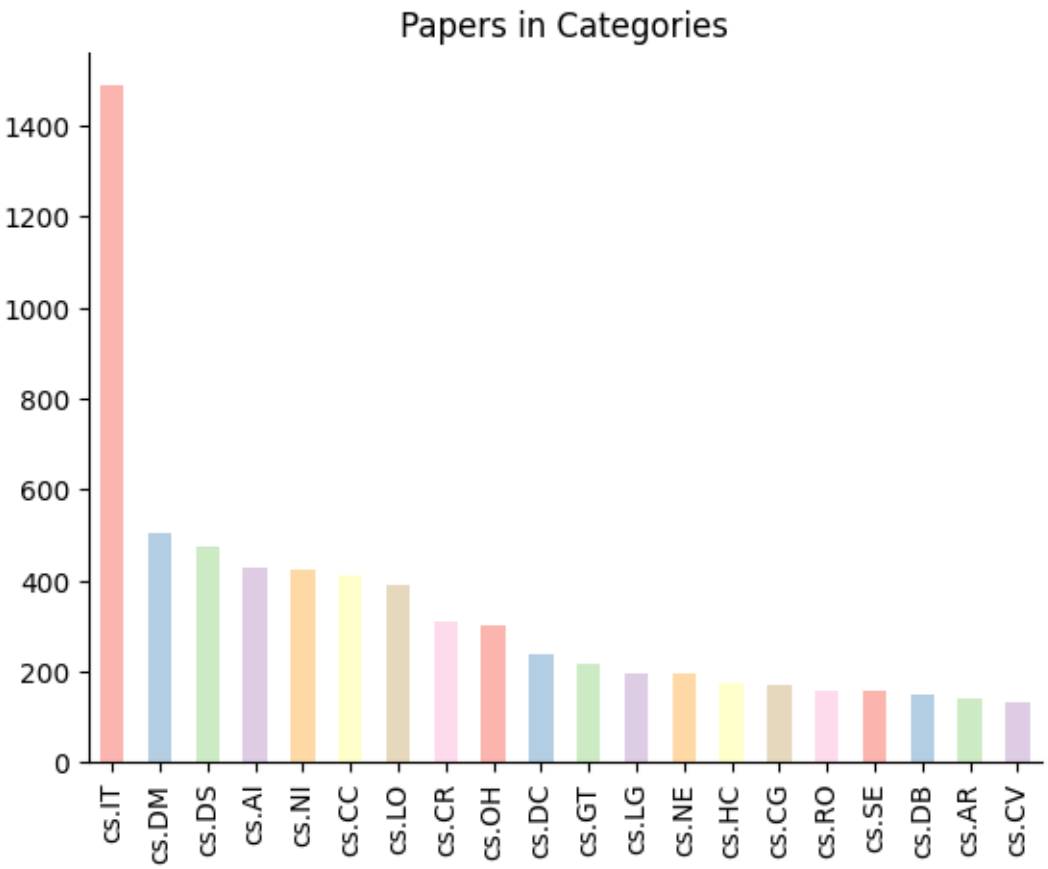

Figure 2: InDegree vs Category Name

Figure: Bar chart showing in-degree vs category name from the ArXiv dataset.

3. Total Number of Papers in CS Categories

Report the total number of papers from all cs categories as one total amount.

- The query is similar to the query used in the ArXiv database tutorial but you will need to use:
  - the where clause
  - the starts with operator
  - either the count() or the sum() aggregation (maybe distinct?)

Total: `5812` papers

4. Query for Total Number of Papers in CS Categories

The query used to get the total number of papers in all cs categories as one total amount from number 3.

- Explain whether you wrote a query of all the unique papers in CS categories or whether you wrote a query that additionally counts a paper for each CS category it participates in (either is fine, I just want you to explain which one you did)! Maybe do it the other way for fun.

```
MATCH (p:Paper)-[:IN_CATEGORY]->(c:Category)
WHERE c.category STARTS WITH 'cs.'
RETURN count(DISTINCT p) AS csPapers
```

This query counts unique papers in CS categories. It:

- Finds papers connected to categories via the IN_CATEGORY relationship
- Filters for categories that start with "cs." using the WHERE clause
- Uses count(DISTINCT p) to ensure each paper is only counted once, even if it belongs to multiple CS categories

5. Additional Queries and Analysis

Have fun and experiment with this technology, do something, or query something that would be interesting to you!

Most Collaborative Authors

```
MATCH (a:Author)-[:AUTHORED]->(p:Paper)
RETURN a.name, count(p) as paperCount
ORDER BY paperCount DESC
LIMIT 10
```

Table 1: Top 10 most collaborative authors in the ArXiv dataset.

a.name	paperCount
"H. Vincent Poor"	105
"Damien Chablat"	99
"Philippe Wenger"	70
"Uwe Aickelin"	45
"Tshilidzi Marwala"	32
"B. Sundar Rajan"	26
"Jérôme Darmont"	25
"Shlomo Shamai"	22
"Olivier Finkel"	19
"Jocelyne Troccaz"	19

Most Interdisciplinary Papers

```
MATCH (p:Paper)-[:IN_CATEGORY]->(c:Category)
WHERE c.category STARTS WITH 'cs.'
RETURN p.title, count(c) as categoryCount
ORDER BY categoryCount DESC
LIMIT 10
```

Table 2: Top 10 most interdisciplinary papers in the ArXiv dataset.

p.title	categoryCount
"The Cyborg Astrobiologist: Porting from a wearable computer to the Astrobiology Phone-cam"	7
"A Novel Clustering Algorithm Based on Quantum Games"	6
"Personalizing Image Search Results on Flickr"	5
"Compressed Counting"	5
"Executable Set Theory and Arithmetic Encodings in Prolog"	5
"TRUST-TECH based Methods for Optimization and Learning"	5
"SAFIUS - A secure and accountable filesystem over untrusted storage"	5
"Knowledge Technologies"	4
"Shallow Models for Non-Iterative Modal Logics"	4
"Beyond Nash Equilibrium: Solution Concepts for the 21st Century"	4