USU Class Network

Bennett DenBleyker
Brad Leavitt
Rachael Fisher

Introduction

Utah State University

- Colleges: 8
- Departments: 52
- Majors: 165
- Programs: 647
- Courses: 6,982

- How similar are the distinct programs at USU?
- Which majors or minors could you easily do within a certain program? Or which major could you transfer to? Which minors would be easy to accomplish with your declared major
- Was CS mistakenly moved out of the engineering program? Does CS belong more to the college of Science or the college of Engineering?

Data Collection

From USU Registrar (Kristi Swainston)

- Degree, Program, Term, and Students in each program for that term.

From usu.edu (Python `requests` for scraping, BeautifulSoup for parsing from university website)

- Programs Data: ID, Name, College, Department, and Related course IDs
- Course Data: ID, Name, and Description
- Sections Data: ID, Term, Course, Section number, Enrollment capacity and actual, Credit hours, Faculty, Instructional method

Data Collection


```
{
    "poid": "33681",
    "desc": "Computer Science - BS",
    "college": "College of Science",
    "department": "Computer Science",
    "courses": [
        "288278",
        "286641",
        "289138",
        "288262",
        "286665",
        "292580",
        "286632",
        "286067",
        "292578",
        "293724",
        "302015",
        "288265",
```


Programs in the department of Computer Science:

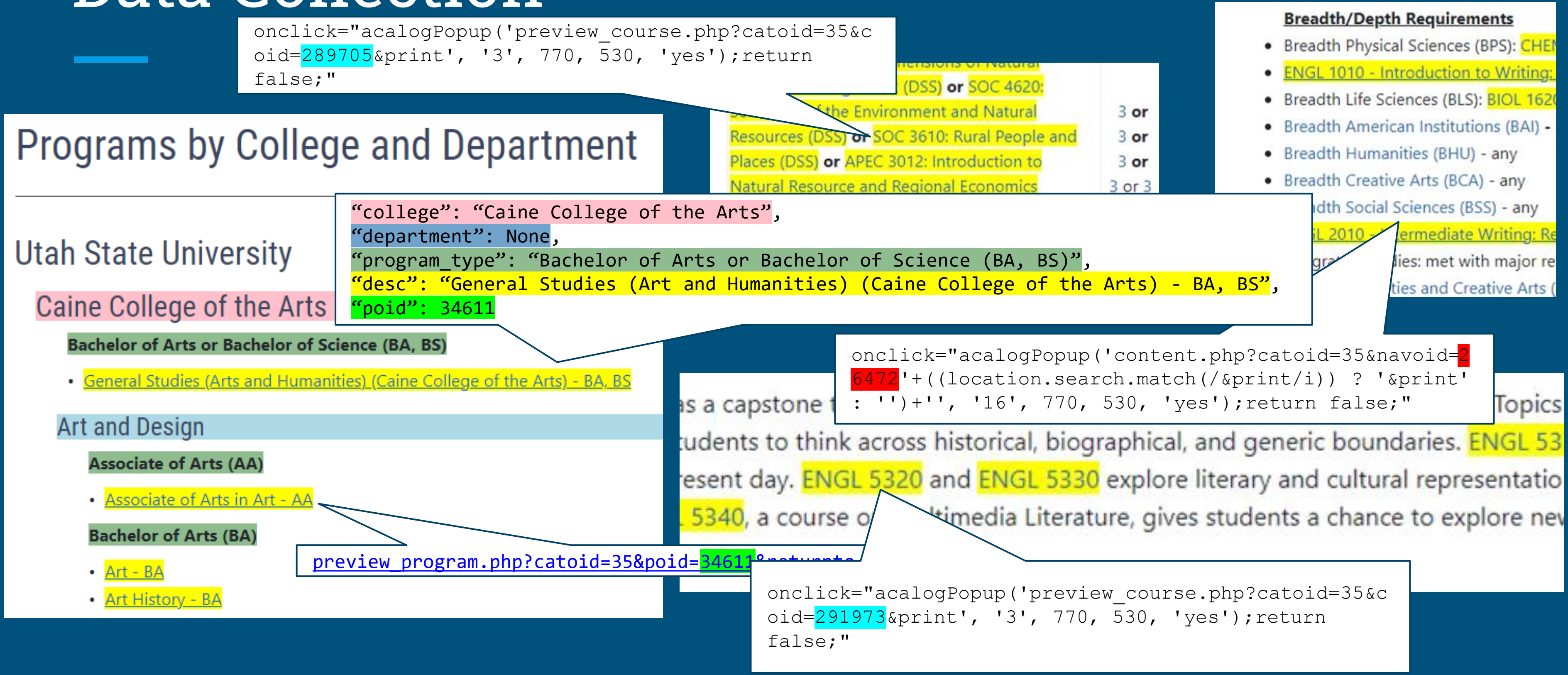
```
p33681 : Computer Science - BS
p33684 : Computer Science - PhD
p34334 : Computer Science
Institutional Certificates of
Proficiency
p33682 : Computer Science - Master
of Computer Science - MCS
p33683 : Computer Science - MS
p34264 : Data Science - MS
p33679 : Computer Science - Minor
p34260 : Computer Science Teaching
- Minor
```

Data Collection


```
{
  "courseID": "286630",
  "subNum": "CS 1400",
  "desc": "Introduction to Computer Science--CS 1"
},
{
  "courseID": "286632",
  "subNum": "CS 1410",
  "desc": "Introduction to Computer Science--CS 2 (QI)"
},
{
  "courseID": "291445",
  "subNum": "CS 1440",
  "desc": "Methods in Computer Science"
},
{
  "courseID": "286633",
  "subNum": "CS 2250",
  "desc": "Cooperative Work Experience"
},
{
  "courseID": "286634",
  "subNum": "CS 2410",
  "desc": "Introduction to Event Driven Programming and GUI's"
},
```


STUDENT_COUNT_IN_PROGRAM_FALL_202	SUBJECT	COURSE_NUMBER	COURSE_TITLE	STUDENT_COUNT_TAKEN_COURSE
30	ITLS	5500	INTEGRAT INNOVAT TECH IN EDUC	3
29	CS	1400	INTRO COMPUTER SCIENCE 1	3
29	CS	1400	INTRO COMPUTER SCIENCE 1	3
29	CS	1400	INTRO COMPUTER SCIENCE 1	3
29	ITLS	5500	INTEGRAT INNOVAT TECH IN EDUC	3
29	ITLS	5500	INTEGRAT INNOVAT TECH IN EDUC	3
29	ITLS	5500	INTEGRAT INNOVAT TECH IN EDUC	3
29	FIN	3200	FINANCIAL MANAGEMENT	2
29	FIN	3200	FINANCIAL MANAGEMENT	2
41	FIN	3200	FINANCIAL MANAGEMENT	7
41	FIN	3200	FINANCIAL MANAGEMENT	7
41	FIN	3200	FINANCIAL MANAGEMENT	7
41	FIN	3200	FINANCIAL MANAGEMENT	7
41	FIN	3200	FINANCIAL MANAGEMENT	7
41	FIN	3200	FINANCIAL MANAGEMENT	7
41	FIN	3200	FINANCIAL MANAGEMENT	7
41	FIN	3400	CORPORATE FINANCE (QI)	4
41	FIN	3400	CORPORATE FINANCE (QI)	4
41	FIN	3400	CORPORATE FINANCE (QI)	4
41	FIN	3400	CORPORATE FINANCE (QI)	4
41	ITLS	5500	INTEGRAT INNOVAT TECH IN EDUC	2

Data Collection



Algorithms

Networkx: jaccard similarity, SimRank similarity, Common Neighbor Centraility

- `jaccard_coefficient()`
- `simrank_similarity()`
- `common_neighbor_centrality()`
- `louvain_communities()`
- `panther_similarity()`

Algorithms Continued

Jaccard similarity

- Is a score ranking based on the shared neighbors of two nodes. We found that this was one of the most useful measures because similar programs would have shared courses students can take to satisfy program requirements.

Simrank similarity

- Is a score that says that "two objects are considered to be similar if they are referenced by similar objects." as described by networkx documentation
- It's a cyclical definition so it eventually converges where each node gets its similarity based on its neighbors.

Algorithms Continued

Common Neighbor Centrality

- Score is based on how many shared neighbors two nodes have and their centrality based on how far away the two nodes are from each other. Though we thought another neighbor based similarity measure would yield good results, the actual results were more entertaining than useful

Panther similarity

- “two objects are considered to be similar if they frequently appear on the same paths.” - networkx. We used it to find the most path similar college with computer science

Results and Analysis

The Most Similar Major to Computer Science

- The program with the highest similarity score to Computer science was Mathematics: Computational - BA, BS with a similarity score of 0.181102.

- Computational Mathematics and Computer Science share the following required courses - Math 1210, Math 1220, Math 3310, Math 2270, CS 1400, CS 1410, CS 2410, CS 2420

Mathematics Major Requirements

- MATH 1210 - Calculus I (QL) 4 credit(s)
- MATH 1220 - Calculus II (QL) 4 credit(s)
- MATH 2210 - Multivariable Calculus (QI) 3 credit(s)
- MATH 2270 - Linear Algebra (QI) 3 credit(s)
- MATH 2280 - Ordinary Differential Equations (QI) 3 credit(s)
- MATH 4200 - Foundations of Analysis (CI) 3 credit(s)

Requirements

- CS 1400 - Introduction to Computer Science–CS 1 4 credit(s)
- CS 1410 - Introduction to Computer Science–CS 2 (QI) 3 credi
- CS 2410 - Introduction to Event Driven Programming and GU
- CS 2420 - Algorithms and Data Structures–CS 3 (QI) 3 credit(s
- MATH 3310 - Discrete Mathematics 3 credit(s)
- MATH 5210 - Introduction to Analysis I 3 credit(s)

- MATH 4610 - Fundamentals of Numerical Analysis 3 credit(s)
- MATH 5610 - Computational Linear Algebra 2 credit(s)

- MATH 5620 - Numerical Algorithms for Approximate Solution
- MATH 5710 - Introduction to Probability 3 credit(s)

Required for Admission to the Professional Prog

A GPA of 2.5 or higher and a grade of C- or better is required in the pre-professional c

- CS 1400 - Introduction to Computer Science–CS 1 4 credit(s)
- CS 1410 - Introduction to Computer Science–CS 2 (QI) 3 credit(s)
- CS 1440 - Methods in Computer Science 3 credit(s)
- CS 2410 - Introduction to Event Driven Programming and GUI's 3 credit(s)
- CS 2420 - Algorithms and Data Structures–CS 3 (QI) 3 credit(s)
- CS 2610 - Developing Dynamic, Database-Driven, Web Applications 3 credit(s)
- MATH 1210 - Calculus I (QL) 4 credit(s)
- MATH 3310 - Discrete Mathematics 3 credit(s)

- BIOL 1610 - Biology I 3 credit(s) **and**
- BIOL 1620 - Biology II (BLS) 3 credit(s)

or

- CHEM 1210 - Principles of Chemistry I 4 credit(s) **and**
- CHEM 1220 - Principles of Chemistry II (BPS) 4 credit(s)

or

- GEO 1110 - Physical Geology (BPS) 3 credit(s)

and

- GEO 1115 - Physical Geology Laboratory 1 credit(s)

and

- GEO 2200 - The Earth Through Time 3 credit(s)

or

- PHYS 2110 - General Physics - Life Sciences I 4 credit(s) a
- PHYS 2120 - General Physics - Life Sciences II (BPS) 4 cre

or

- PHYS 2210 - Physics for Scientists and Engineers I (BPS/C
- PHYS 2220 - Physics for Scientists and Engineers II (BPS/C

r Science Required Courses

ire a grade of C- or better.

omputer Systems Organization and Architecture 3 credit(s)

perating Systems and Concurrency 3 credit(s)

ntroduction to Software Engineering (CI) 4 credit(s)

rogramming Languages 3 credit(s) **or**

ompiler Construction 4 credit(s)

heory of Computability 3 credit(s) **or**

dvanced Algorithms 3 credit(s)

The Best Minor(s)

With **Computer Science - BS**:

1. **Computer Science - Minor** (13.043% similar)
2. **Biomathematics - Minor** (9.420% similar)
   - Assuming you take the Biology classes for your Gen-Ed classes
3. **Math Education - Minor** (7.031% similar)

➢ **Mathematics - Minor** (5.882% similar)

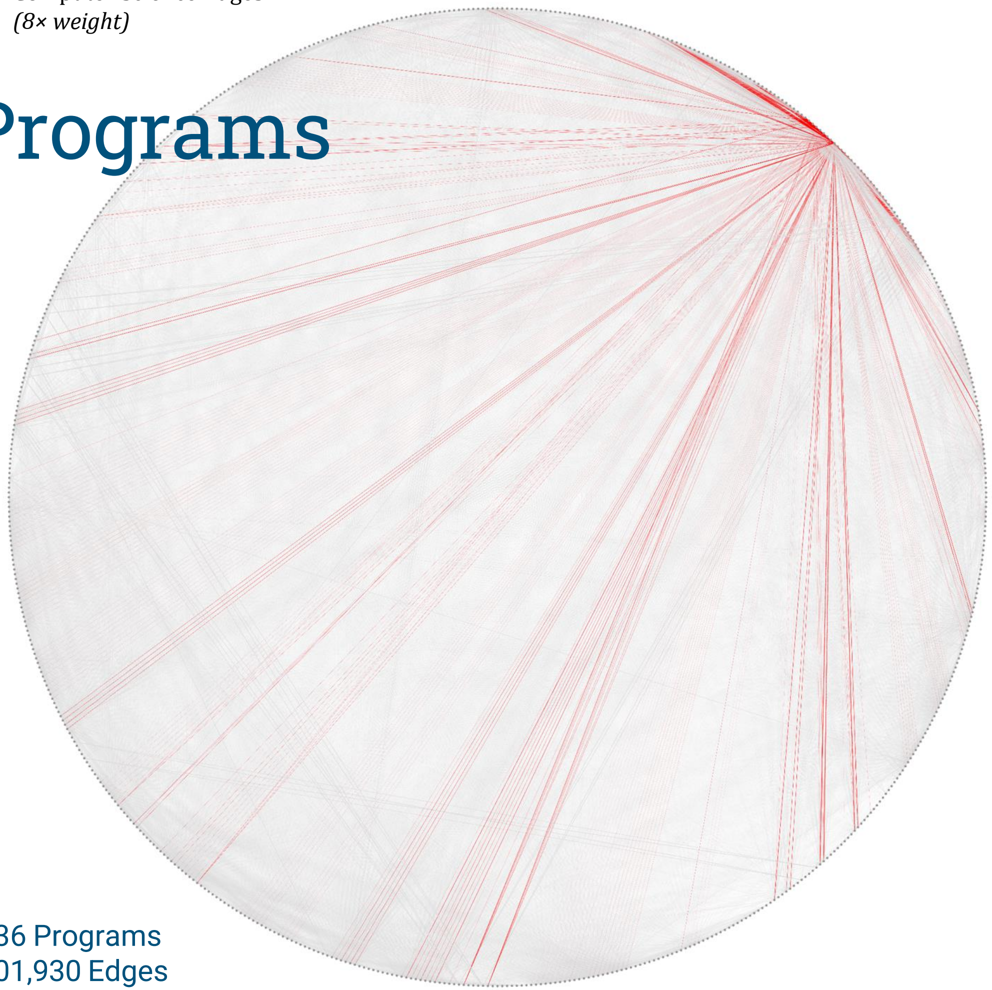
Similarity Between Programs

Jaccard Similarity

$$\frac{|\Gamma(u) \cap \Gamma(v)|}{|\Gamma(u) \cup \Gamma(v)|}$$

$$\Gamma(x_i): \{x_j \in A \mid A_{ij} > 0\}$$

Measures the similarity between the neighborhoods of two nodes



636 Programs
201,930 Edges
(Complete Graph)

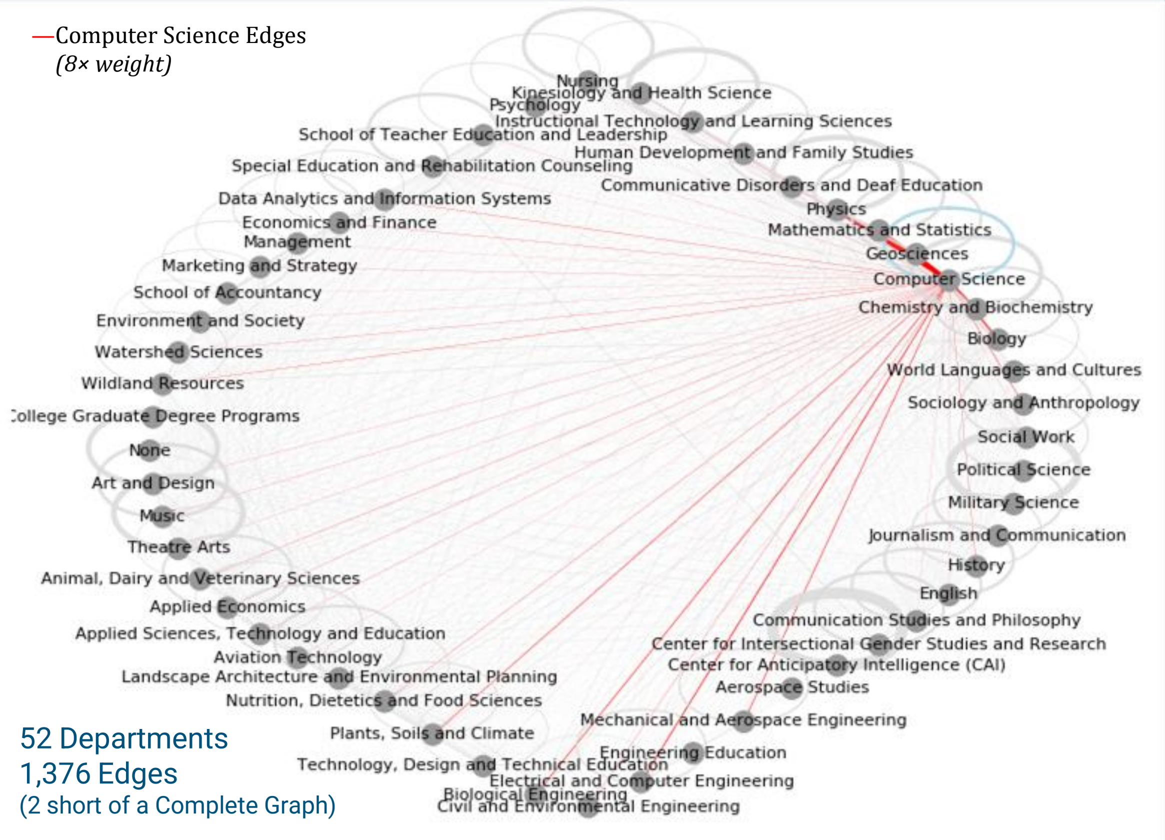
Similarity Between Programs

Jaccard Similarity

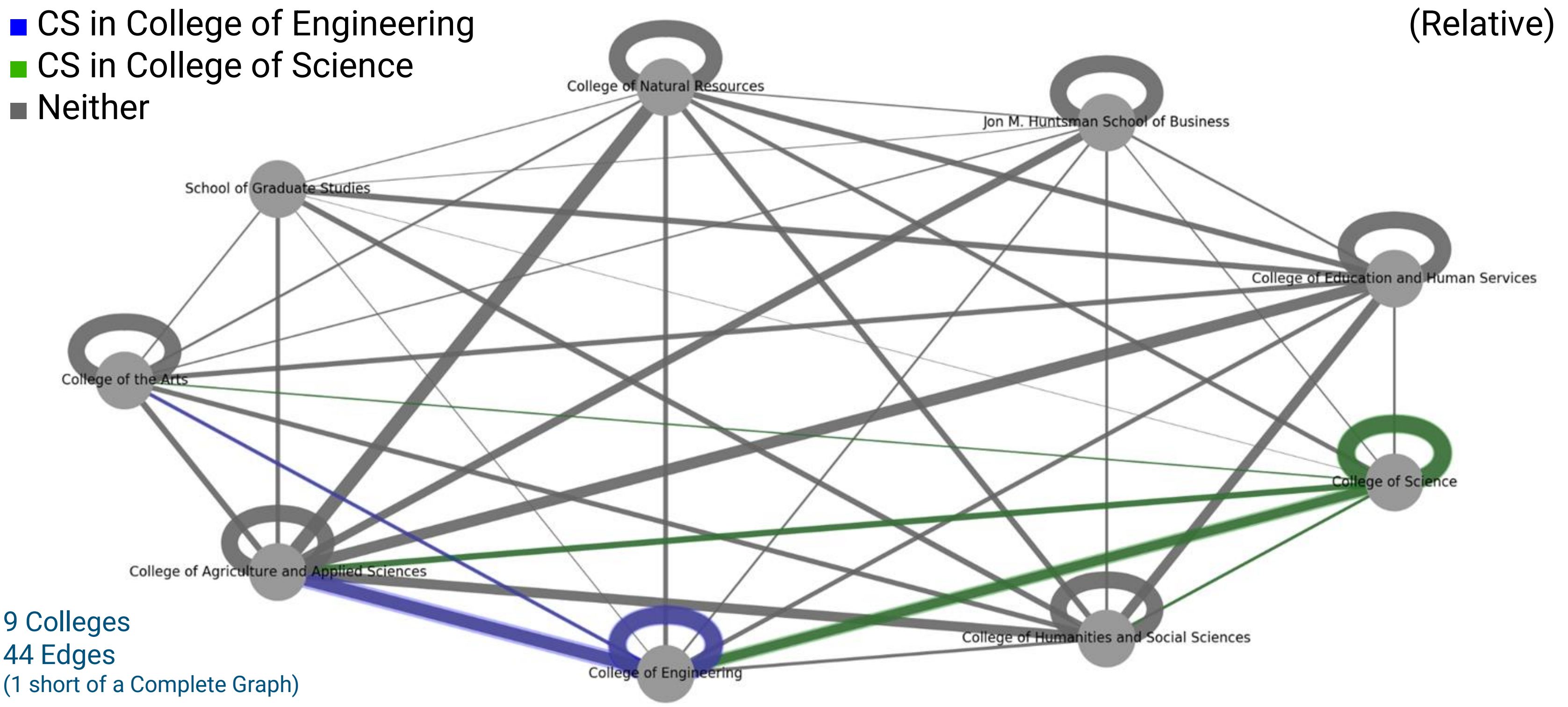
$$\frac{|\Gamma(u) \cap \Gamma(v)|}{|\Gamma(u) \cup \Gamma(v)|}$$

$$\Gamma(x_i): \{x_j \in A \mid A_{ij} > 0\}$$

Measures the similarity between the neighborhoods of two nodes



Similarity Between Programs



9 Colleges
44 Edges
(1 short of a Complete Graph)

Similarity Be

Louvaine Grouping

Greedy algorithm for m

Special Education and Rehabilitation Counseling: 45.2%

- Disability Studies - Minor
- Special Education - BA, BS
- Special Education - Minor
- Special Education Online Practical Teacher Training (OPTT): Birth to Age 5 Emphasis - BA, BS
- Special Education Online Practical Teacher Training (OPTT): Mild/Moderate Emphasis - BA, BS
- Special Education Online Practical Teacher Training (OPTT): Severe Emphasis - BA, BS
- Special Education: Birth to 5 Emphasis - BA, BS
- Special Education: Mild/Moderate & Birth to 5 Dual Emphasis - BA, BS
- Special Education: Mild/Moderate & Severe Dual Emphasis - BA, BS
- Special Education: Mild/Moderate Emphasis - BA, BS
- Special Education: Mild/Moderate Emphasis/Elementary Education Composite - BA, BS

Plants, Soils and Climate: 59.3%

- Agronomy - Minor
- Climate Science - BS
- Crop Biotechnology - Minor
- Horticulture - BS
- Horticulture - Minor
- Landscape Management Certificate
- Ornamental Horticulture - AAS
- Ornamental Horticulture - Certificate of Completion
- Ornamental Horticulture - Minor
- Plant Science: Horticulture and Cropping Systems Emphasis - BS
- Plant Science: Research Emphasis - BS
- Residential Landscape Design - Minor
- Residential Landscape Design and Construction - BS
- Soil Science - Minor
- Soils and Sustainable Land Systems: Environmental Soil Science Option - BA, BS
- Soils and Sustainable Land Systems: Sustainable Food Production Option - BA, BS

Applied Sciences, Technology and Education: 29.6%

- Agricultural Communication - BS
- Agricultural Communication and Journalism - BS
- Agricultural Education - BS, Community-Based
- Agricultural Education - BS, School-Based
- Agricultural Machinery Technology - AAS
- Agricultural Machinery Technology - Certificate
- Agricultural Systems Technology - BS
- Agricultural Systems Technology - Minor

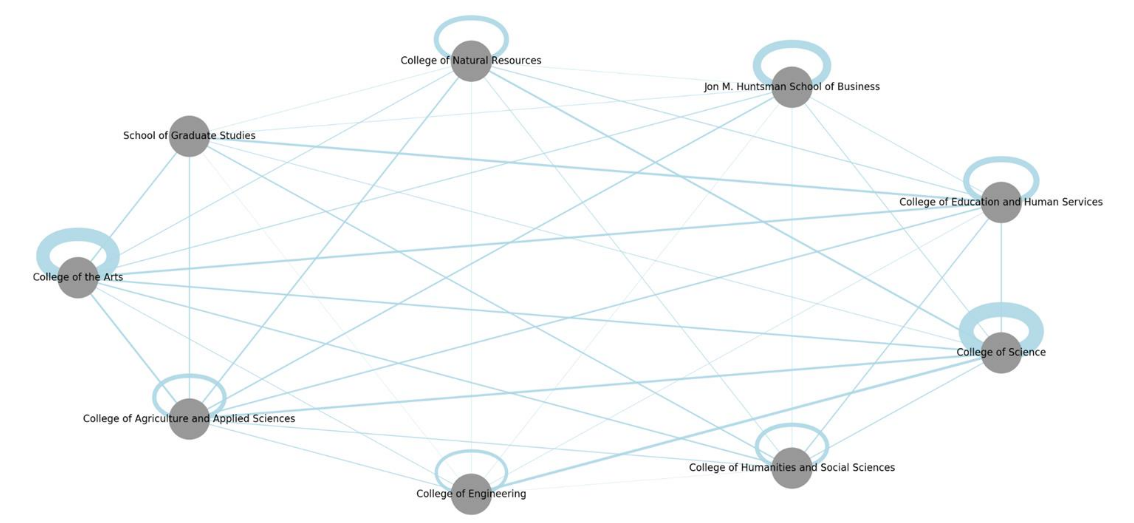
Animal, Dairy and Veterinary Sciences: 11.1%

- Animal and Dairy Sciences - Minor
- Equine - Minor
- Ranch Horse Management - Certificate of Completion

What College For *Computer Science*?

Jaccard Similarity

- Comparing absolute vs. relative values



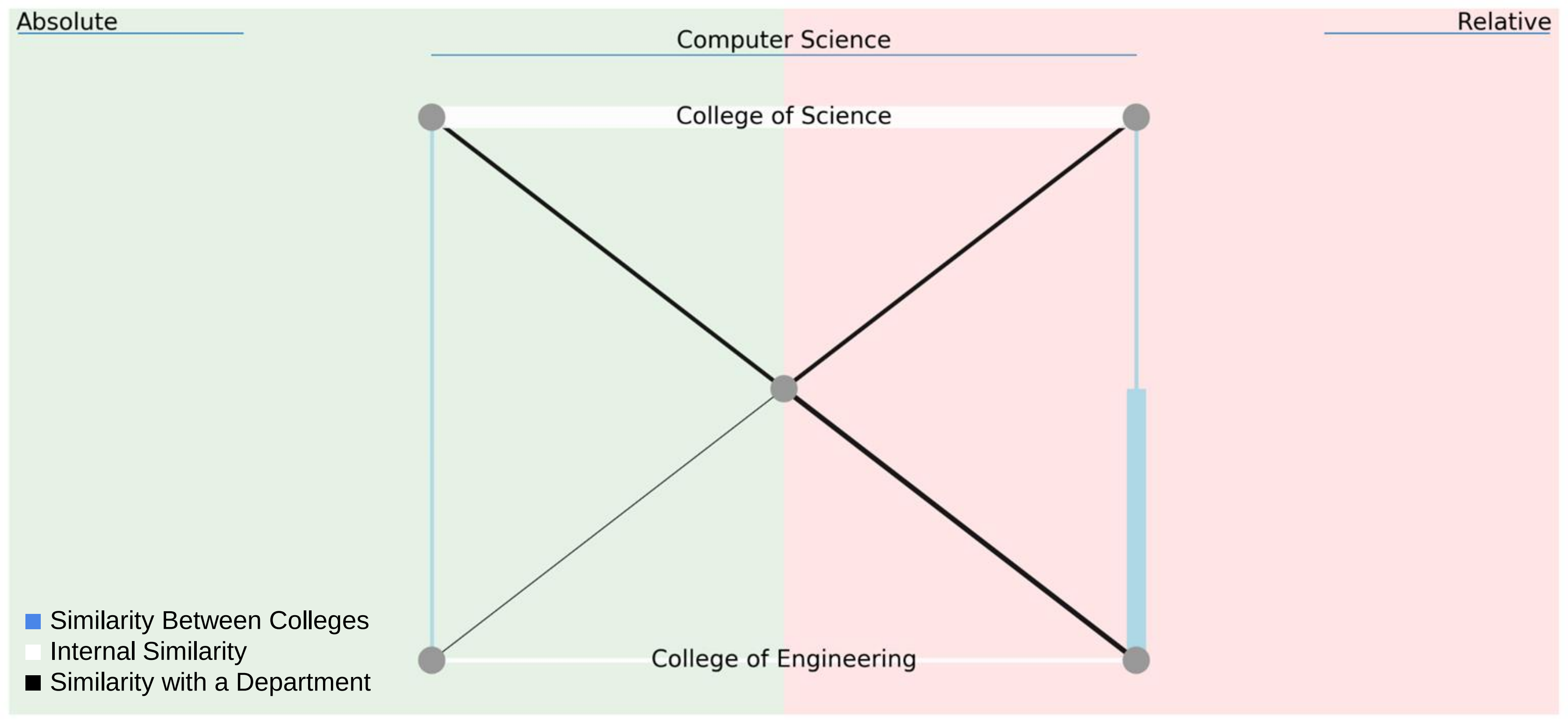
What College For *Computer Science*?

Jaccard Similarity

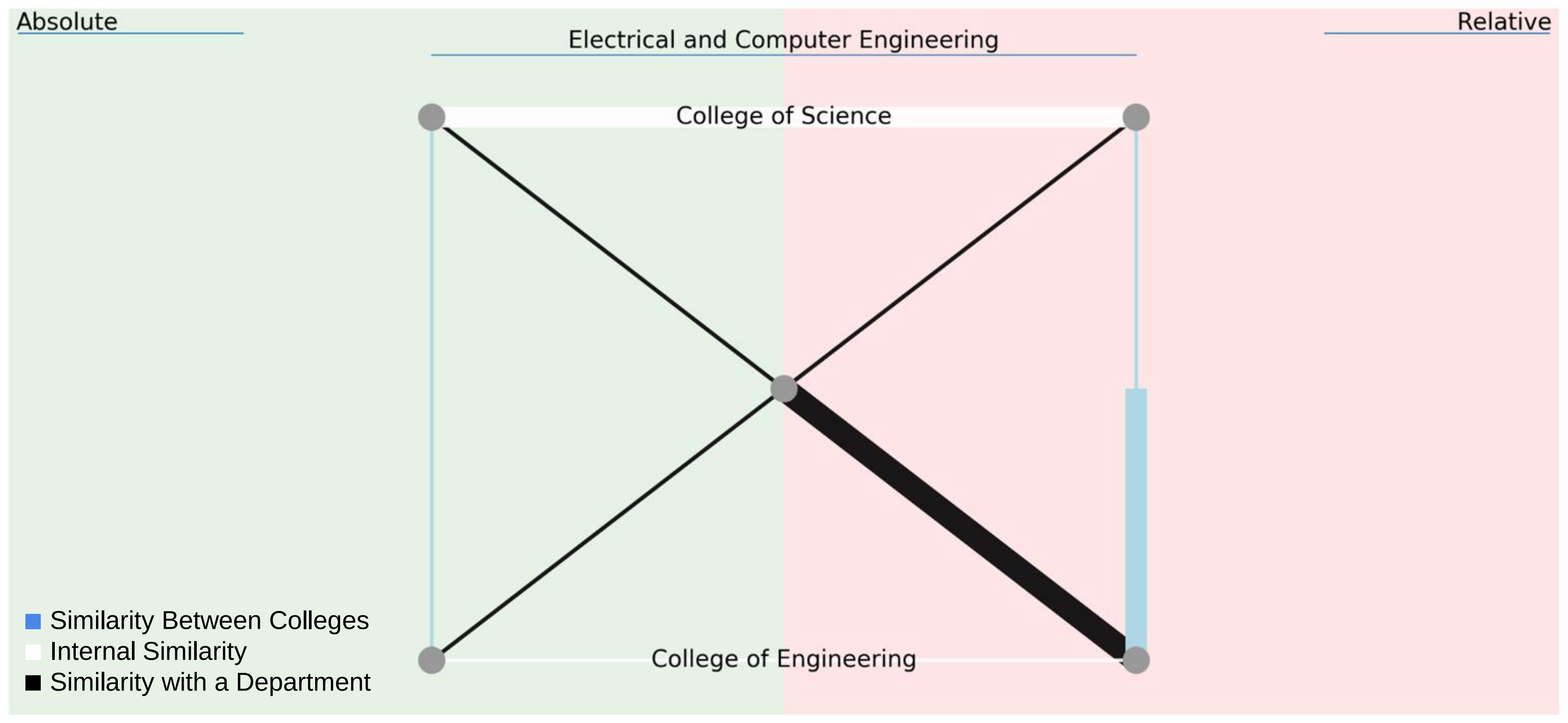
- Comparing absolute vs. relative values

Three programs:

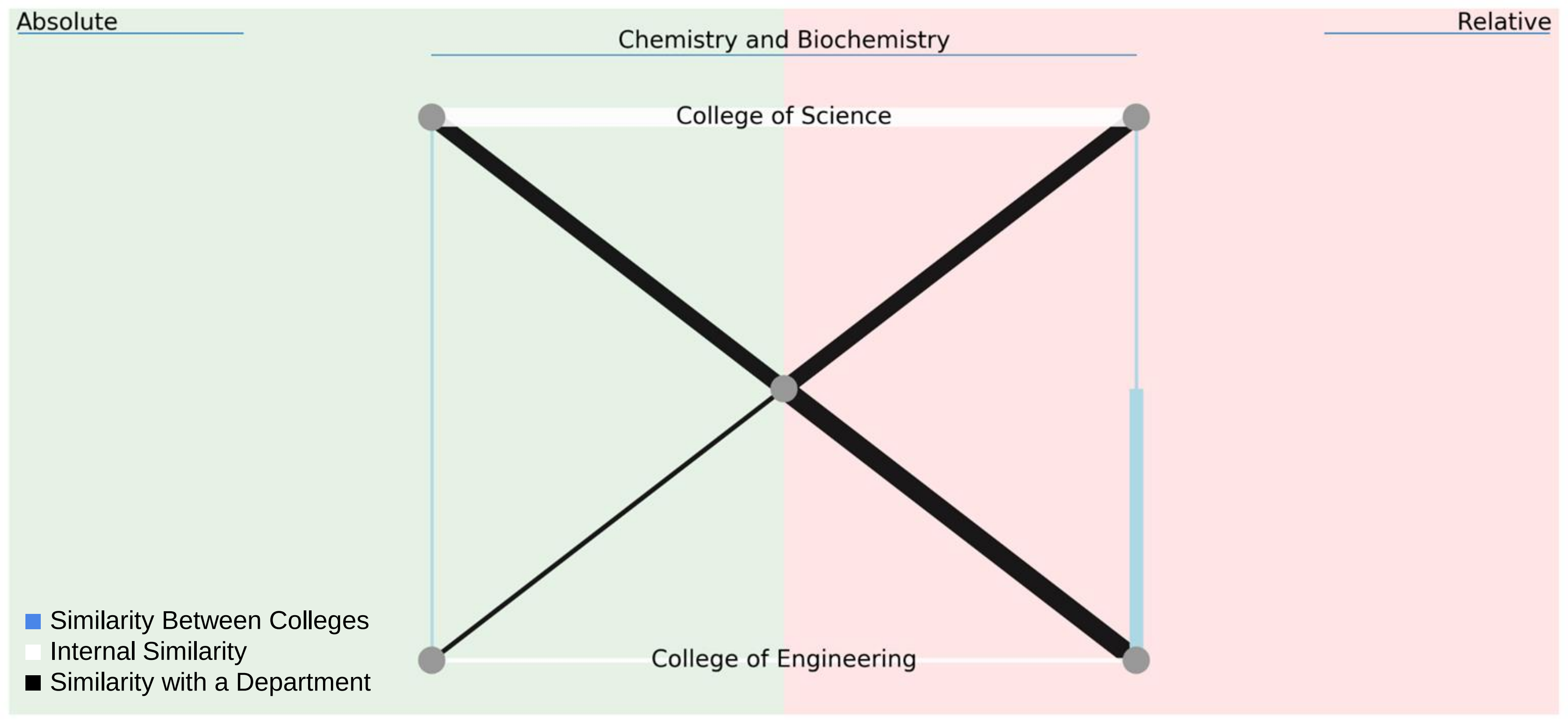
1. *Electrical and Computer Engineering* - ***College of Engineering***
2. *Chemistry and Biochemistry* - ***College of Science***
3. *Computer Science* - ***College of ?????***



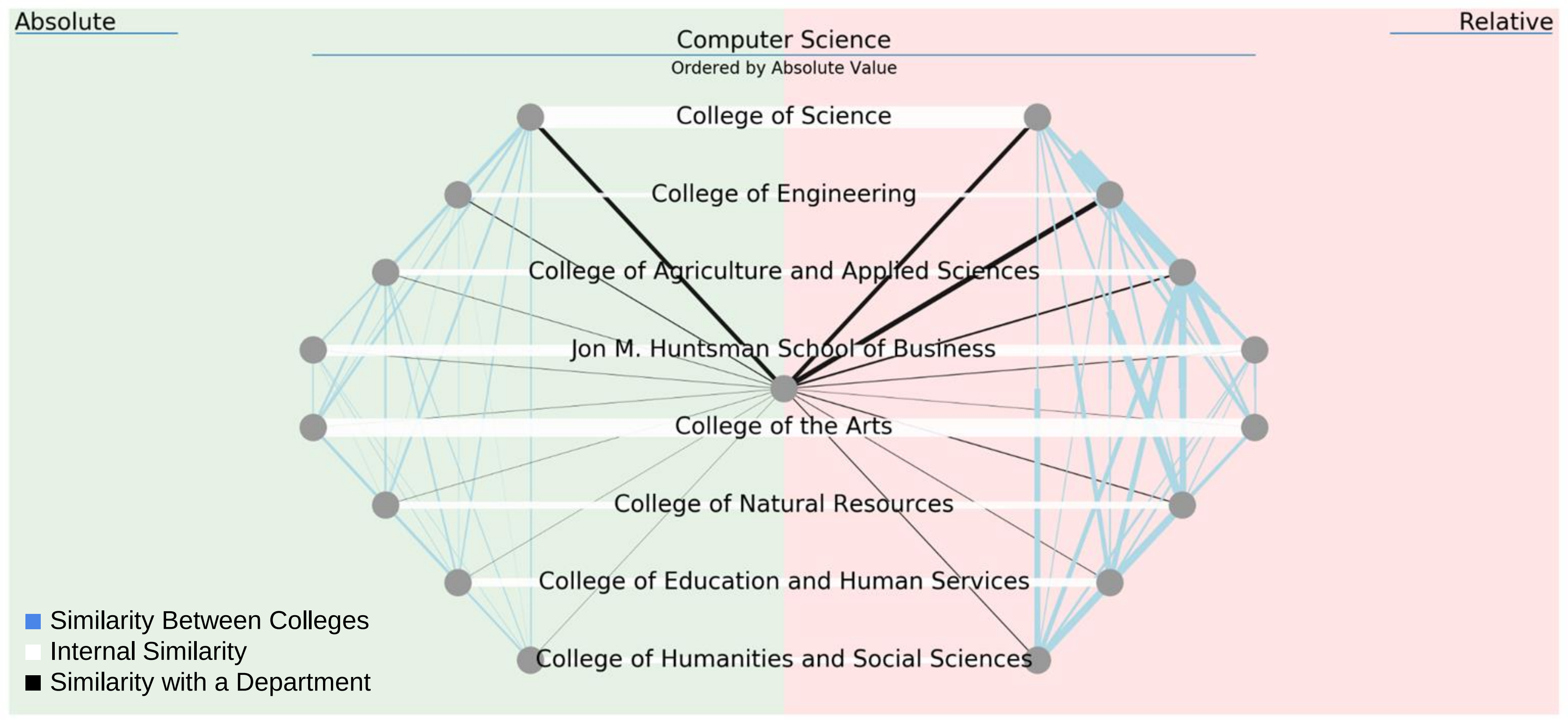
College of Science vs. College of Engineering - Jaccard Similarity



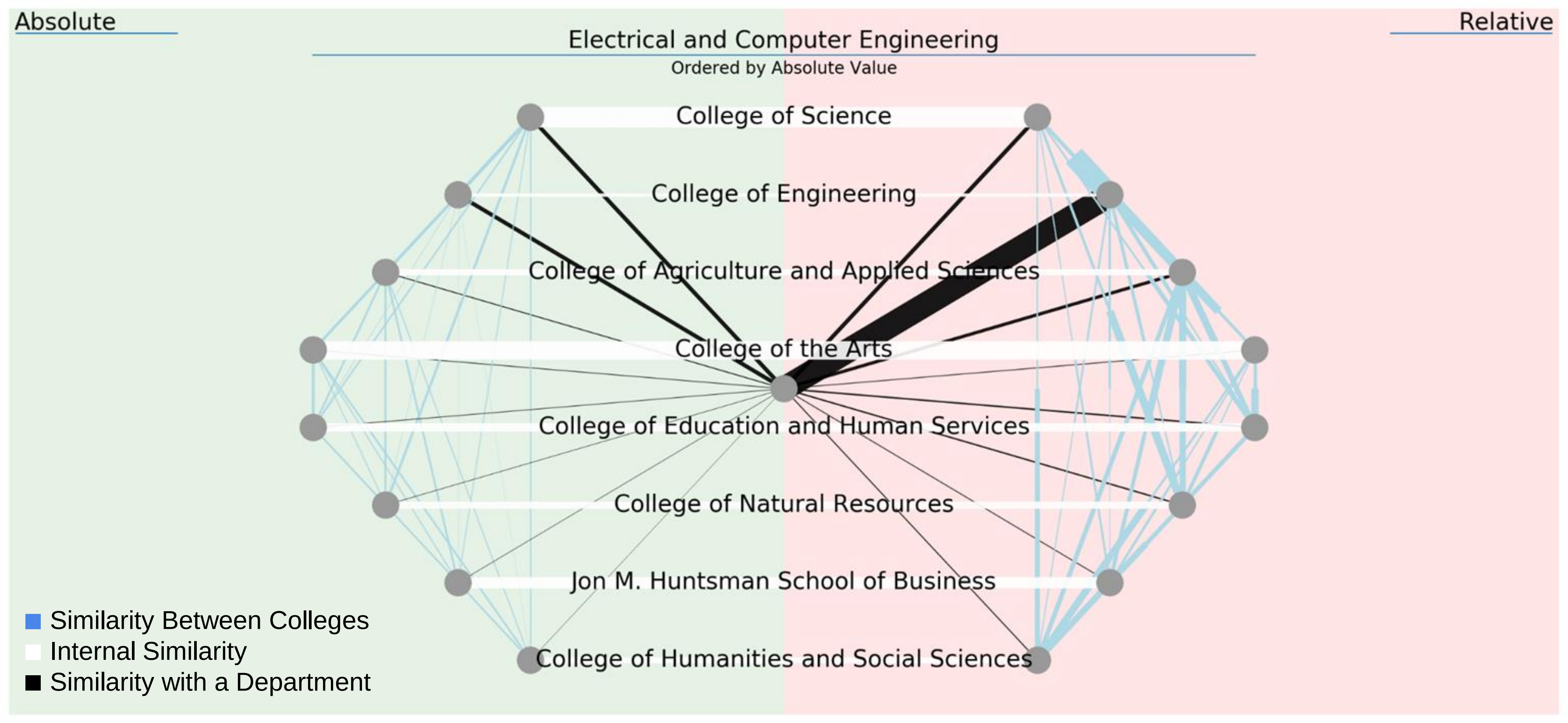
College of Science vs. College of Engineering - Jaccard Similarity



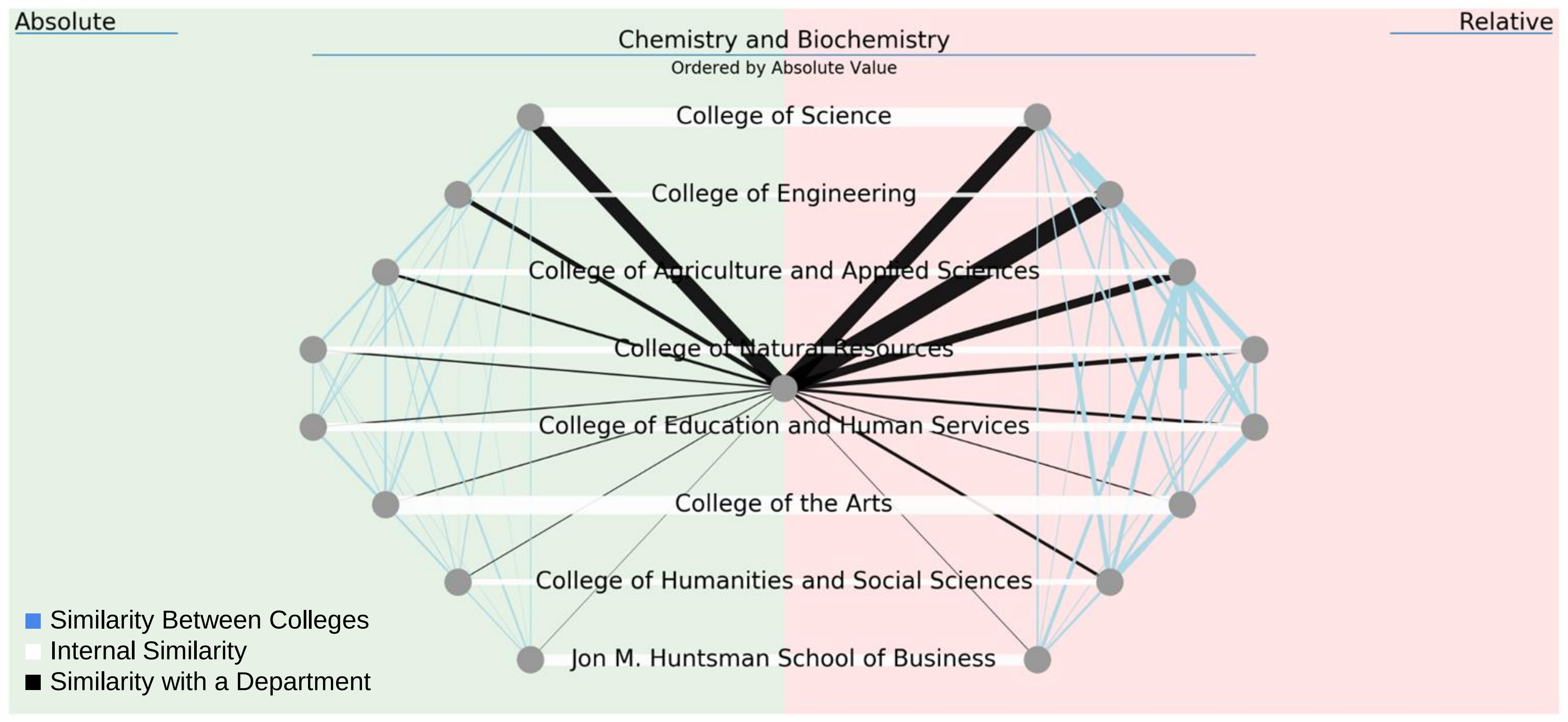
College of Science vs. College of Engineering - Jaccard Similarity



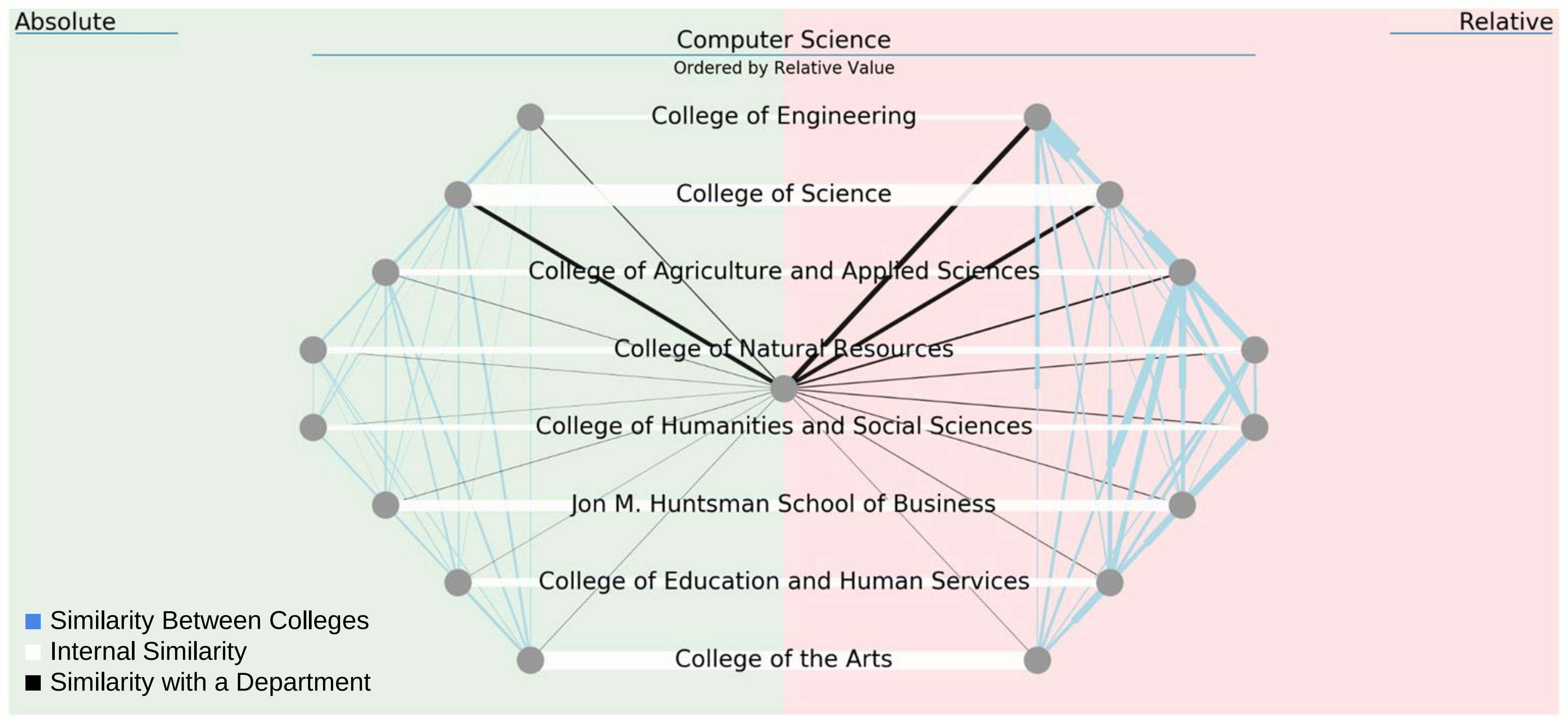
Jaccard Similarity (Ordered by Absolute Similarity)



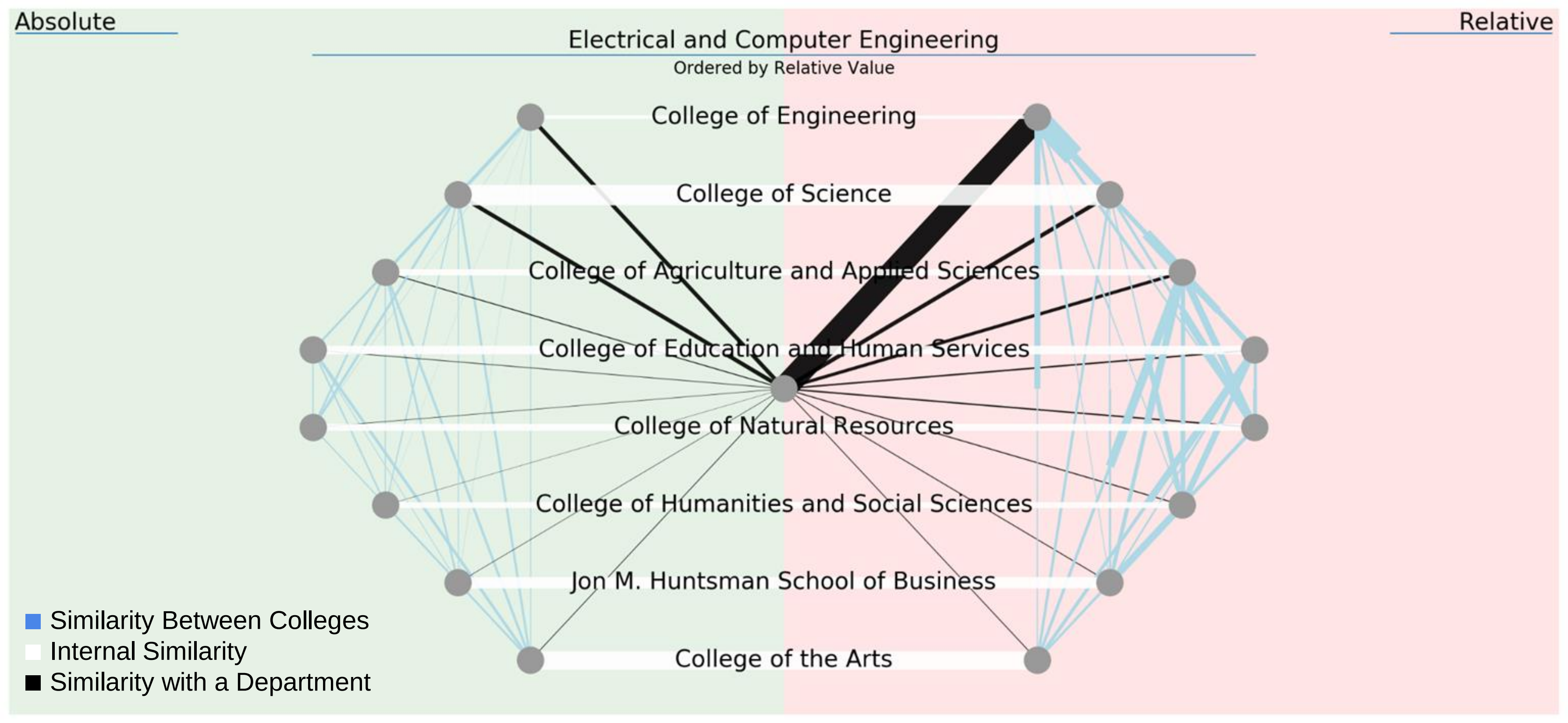
Jaccard Similarity (Ordered by Absolute Similarity)



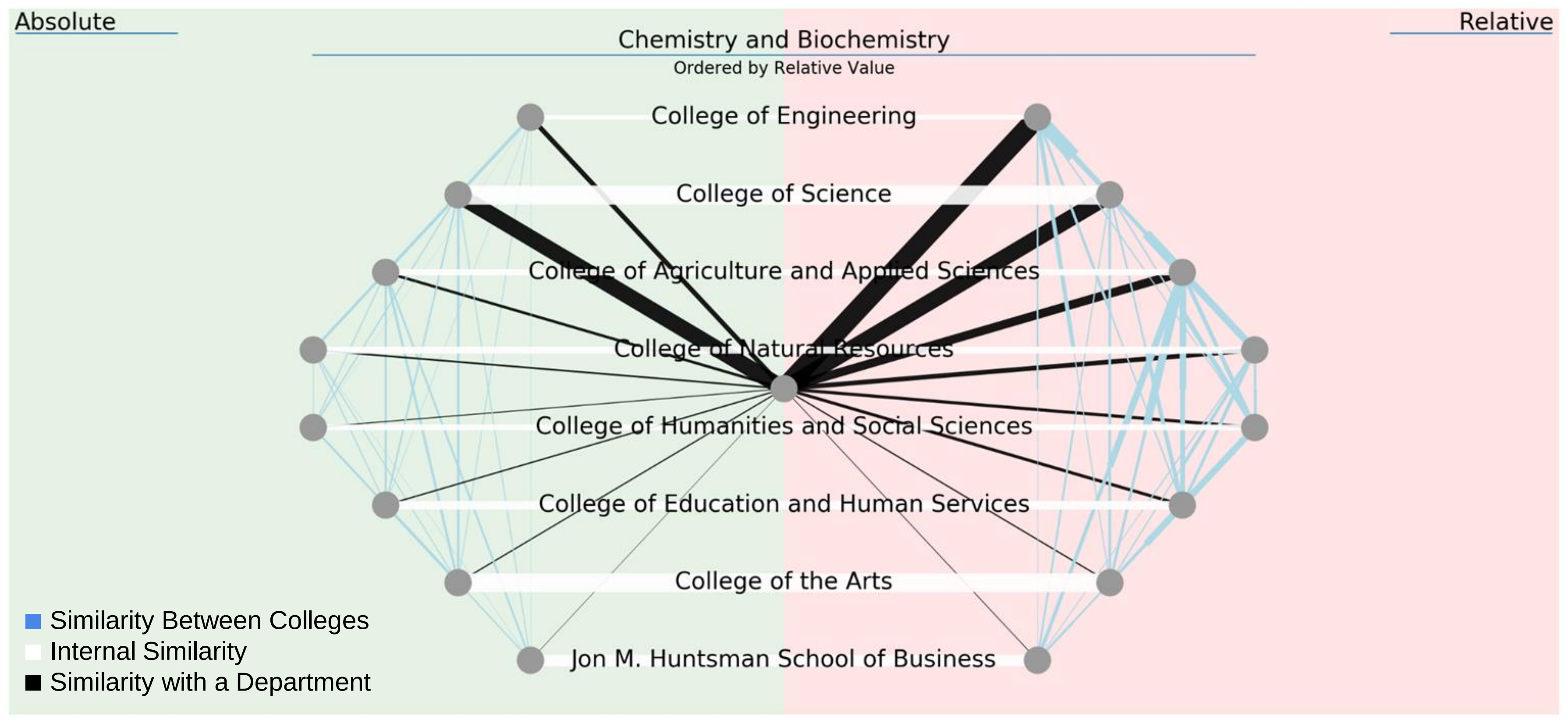
Jaccard Similarity (Ordered by Absolute Similarity)



Jaccard Similarity (Ordered by Relative Similarity)



Jaccard Similarity (Ordered by Relative Similarity)



Jaccard Similarity (Ordered by Relative Similarity)

Results : Simrank Similarity on departments

Between departments the simrank similarity of CS with others is shown in this top 10 chart.

Showing the same evidence that while CS is most similar with the math dept. The within the top five are engineering departments. It also has many college of science depts. as well.

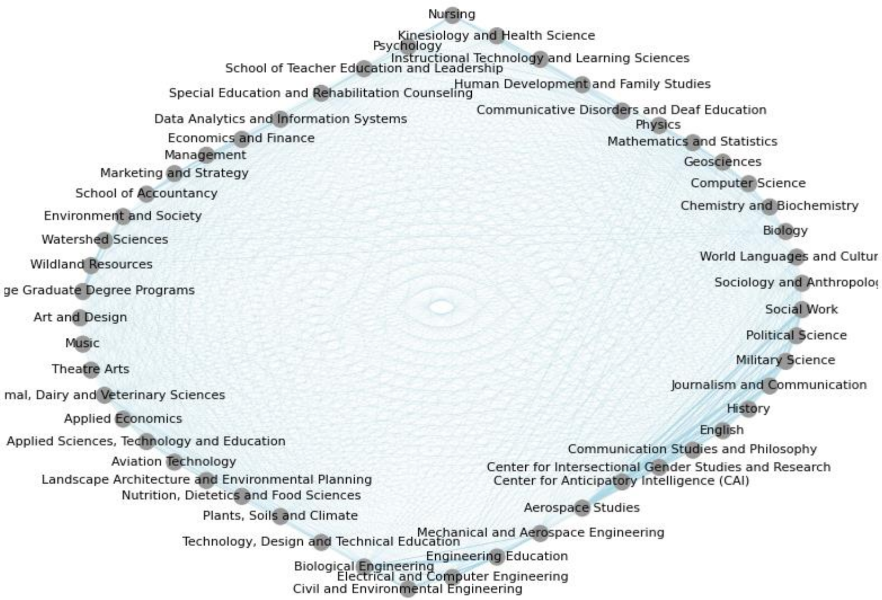
Note: department nodes start with a character ‘d’

departments	similarity score
dMathematics and Statistics	0.028097
dElectrical and Computer Engineering	0.026339
dPhysics	0.021381
dMechanical and Aerospace Engineering	0.020187
dGeosciences	0.017547
dChemistry and Biochemistry	0.016342
dBiological Engineering	0.015676
dPlants, Soils and Climate	0.015141
dBiology	0.015002

Simrank Cont.

It’s slightly visible that CS (upper right quadrant) and departments like math, engineering and science have some similarity compared to others.

Another point to note is observing the bottom right quadrant. (Aerospace, gender studies and CAI)



Hierarchical Clustering from Simrank

Using the similarities from the Simrank algorithm. We were able begin clustering like nodes together. Single linkage clustering is used to decide what clustered sets can combine. We found by simrank:

- First to group: Mechanical and Aerospace Engineering, and Electrical and Computer Engineering. Being the most similar departments
- College Graduate Degree Programs, Environment and Society, Wildland Resources, and Watershed Sciences are the first group of size 4.
- CS finally comes in at the 11th grouping in the cluster:
- Mathematics and Statistics, Physics, Computer Science, Mechanical and Aerospace Engineering, and Electrical and Computer Engineering

Other Results

Common Neighbor Similarity

- Was the least successful.
- Gave nearly all departments the same highest similarity score. So when grouped there appeared to be little to no intuitive similarities.
- Came from converging values in graph cycles.

Panther similarity

- Limited workability, difficult to separate different node types. So we grabbed the first college that it produced
    - It was the college of science.

Conclusion

With the results given by our experiments how were our questions answered?

- Does CS lean towards college of Engineering or Science?
  - There was plenty of evidence for both but it still is unclear given the results on our experiments.
- What programs would be easy to switch to if needed?
  - With our results its safe to say (and intuitive) that switching to a program within your current department or college would be fairly easy
  - According to similarity the departments in Science and Engineering are the first to cluster together. It's not too rare for students to switch from CS to Math or engineering and vice versa.
  - Minors in CS Math and Engineering also have some support across those departments.

Lessons Learned

- Learned a lot more about graph algorithms in practice
  - Learning new algorithms, Simrank, louvain communities
  - Tying in ones we know. Cliques, Jaccard coefficient
- Graph structure is important
  - Most graph algorithms depend on structure
  - It’s important to know if the graph has been connected correctly
- Data can be messy/dirty
  - Many isolated nodes unconnected to graph
  - Even with uniform sites data was hard to extract

Future work

- Using the registrar data
  - Has a lot more info that would be good to parse through
  - Deeper data set for node labelling
- Including IDEA survey data
  - Learn more of the quality of course/section nodes
- Possible machine learning
  - Graph neural networks (GNN)
  - Course path finding
  - Schedule suggestions