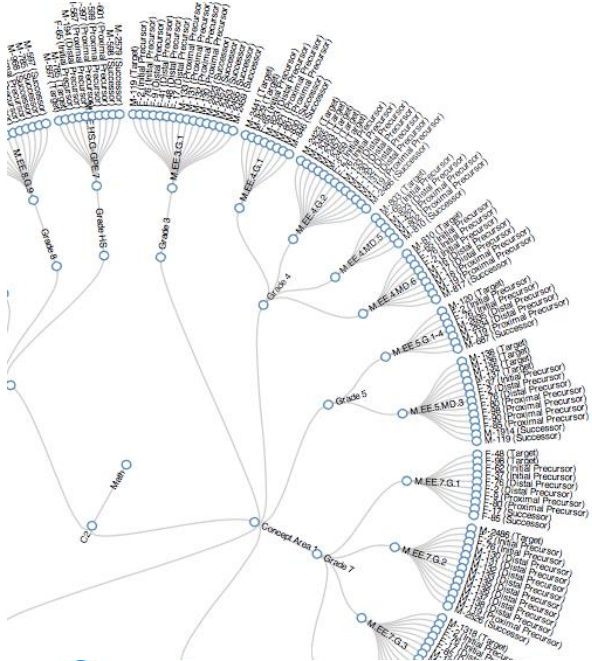


The School of Information and its Relationship to Computer Science at UC Berkeley

Zachary A. Pardos
 School of Information /
 Graduate School of Education

Andrew Joo Hun Nam
 Computer Science / Economics



Economics | **EECS**



Motivation

What is an
Information

School?

Polymorphism
(nimbleness)



CS

*“We see the world through information
colored glasses”*

Motivation

DATA SCIENCE

Information

CS

iConference 2017

Workshop 4a: Information Science to Data Science: New Directions for iSchools

“What is our competitive advantage over CS?”

Workshop 1a: Undergraduate Education in iSchools

“What is the difference between a Data Mining course in CS and an iSchool?”

Motivation

STAT 89A

Information

SCIENCE

INFO 88A

DATA

CS

COGSCI 88

DATA SCIENCE AND THE MIND

MCB 88

IMMUNOTHERAPY OF CANCER: SUCCESS AND FAILURE

Berkeley Institute for Data Science



The Berkeley Institute for Data Science (BIDS) promotes scientific breakthrough interdisciplinary discovery.

[Berkeley Insti](#)

Simons Institute



The Simons Institute for the Theory of Computing brings together the world's

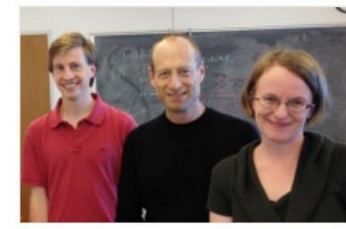
RISELab



The RISELab focuses on the development of data-intensive stems that provide real-time intelligence with secure execution (ISE).

[SELab](#)

Statistics



In addition to developing fundamental theory and methodology, many faculty in the Department of Statistics have strong interdisciplinary links.

[Statistics](#)

Center for Long-Term Cybersecurity



The Center for Long-Term Cybersecurity supports research, current events, seminars, and research on the future of cybersecurity.

[Cer](#) **CS** [Cybersecurity](#)

Information

D-Lab



D-Lab helps Berkeley faculty, staff, and graduate students move forward with world-class research in data intensive social science.

[D-Lab](#)

SDAV Institute



The Scalable Data Management, Analysis, and Visualization (SDAV) effort supports breakthrough science by providing technical solutions that are broadly applicable.

[SDAV Institute](#)

EECS



Electrical Engineering and Computer Science (EECS) is a top-ranked department in the nation, with many outstanding faculty combining science with the power of engineering.

[EECS](#)

Berkeley Artificial Intelligence Research Lab (BAIR)



The Berkeley Artificial Intelligence Research (BAIR) Lab brings together UC Berkeley researchers across the areas of computer vision, machine learning, natural language processing, planning, and robotics.

School of Information



The School of Information is a research and education community offering graduate degrees in Information and Data Science and Information Management and Systems.

[School of Information](#)

Our approach: Enrollment data analysis

DATA SCIENCE

Information

CS

RQ1: How do the graduate units differ in the students drawn to their courses?

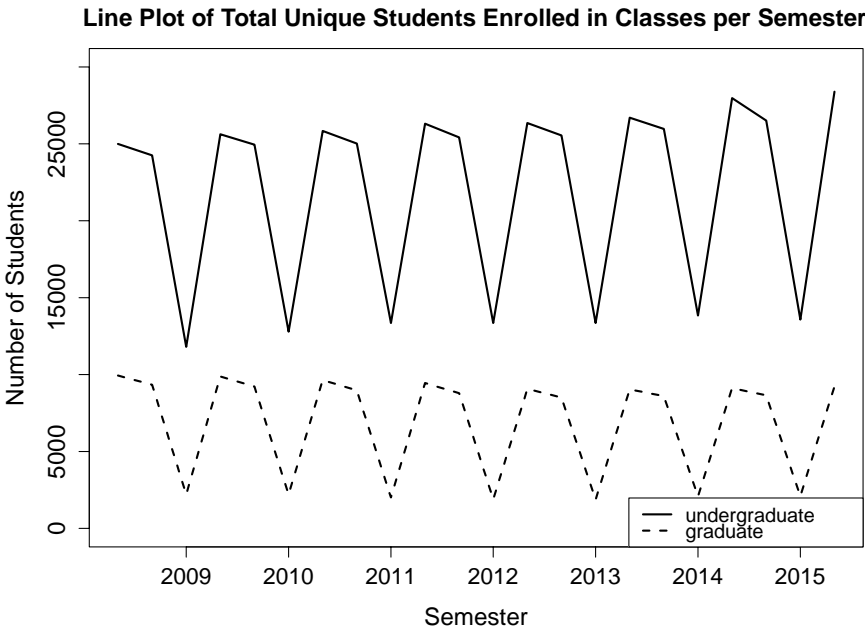
RQ2: How can we express the units as a composition of and proximity to other academic units?

Related Work

- Wu et al. (2012)
 - Categorize iSchools into: LIS, CS, Business Management Schools (UCB)
 - Consider faculty training, most common majors, journals, research keywords
- Wiggins & Sawyer (2012)
 - Studied faculty composition of iSchools, with Computing (UCB) being the top discipline of faculty, followed by Information, Library, and Social & Behavioral
- Thompson (2008)
 - Describes iSchools at a high level relating to philosophy and sociology
 - And at a low level, sharing principles with computer science and EE
- Other summaries of iSchools by Larsen (2009), Olson & Grudin (2009)

Dataset

- 3.6M enrollments at UCB from Fall '08 through Fall '15
- 110,335 undergraduates
- 38,147 graduates
- 9,038 unique lectures courses
 - across 17 colleges
 - 124 departments



Semester Year	STU ID (anon)	Undergraduate/ Graduate	Dept	Course Number	Grade	Major
Fall 2008	9984	Graduate	INFO	254	A	Econ
Fall 2008	9984	Graduate	INFO	290	A	Econ
Spring 2009	9984	Graduate	INFO	198	B	Econ
Spring 2014	282243	Undergrad	INFO	178	B	Law
Summer 2014	282243	Undergrad	CS	165	C	Law
Fall 2014	282243	Undergrad	CS	140	B	Law

Method 1: Descriptive statistics

- What is the distribution of outside majors (both undergrad and grad) enrolling in courses offered by the iSchool and CS?

Method 2: Vector embedding

- Which units are closest in cosine space to the iSchool and CS?
- What combination of two units is closest to the iSchool and CS?

Semester Year	STU ID (anon)	Undergraduate/ Graduate	Dept	Course Number	Grade	Major
Fall 2008	9984	Graduate	INFO	254	A	Econ
Fall 2008	9984	Graduate	INFO	290	A	Econ
Spring 2009	9984	Graduate	INFO	198	B	Econ
Spring 2014	282243	Undergrad	INFO	178	B	Law
Summer 2014	282243	Undergrad	CS	165	C	Law
Fall 2014	282243	Undergrad	CS	140	B	Law

Method 1 results

What majors are iSchool courses attracting?

Rank	Major	All Students	
1	<u>Electrical Engineering and Computer Science</u>	610	(8%)
2	Business Administration ¹⁰	162	(2%)
3	Public Policy	86	(1%)
4	<u>Mechanical Engineering</u> ¹	81	(1%)
5	Public Health	63	(1%)
6	Interdisciplinary Studies	54	(1%)
7	Education	50	(1%)
8	<u>Industrial Engineering and Operations Research</u> ⁶	49	(1%)
9	Architecture	46	(1%)
10	<u>Cognitive Science</u> ⁸	42	(1%)
	Total (Outside iSchool)	1,834	(24%)
	Total (Inside iSchool)	5,953	(76%)
	Total (Overall)	7,787	(100%)

- Six of the top 10 majors are non-STEM
- iSchool courses consist of 25% outside majors
- EECS is the top outside major served by the iSchool

- Underlined majors are considered STEM by USDHS

Method 1 results

What majors are Computer Science courses attracting?

Rank	Major	All Students	
1	<u>Mechanical Engineering</u> ⁶	274	(2%)
2	<u>Information</u>	215	(2%)
3	<u>Mathematics</u>	135	(1%)
4	<u>Statistics</u>	122	(1%)
5	<u>Applied Mathematics</u>	121	(1%)
6	<u>Industrial Engineering and Operations Research</u> ⁸	102	(1%)
7	<u>Civil and Environmental Engineering</u>	97	(1%)
8	<u>Cognitive Science</u> ¹⁰	91	(1%)
9	<u>Physics</u>	64	(1%)
10	Business Administration ²	52	(0%)
29	Education	11	(0%)
	Total (Outside EECS)	1,858	(15%)
	Total (Inside EECS)	10,300	(85%)
	Total (Overall)	12,158	(100%)

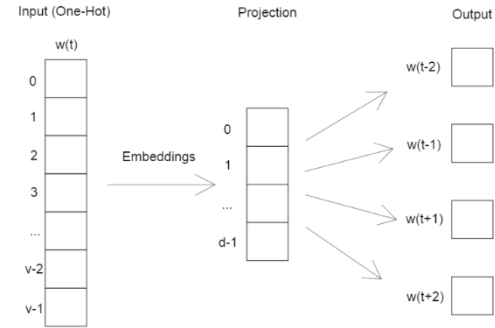
- The top nine outside majors in CS are STEM
- CS courses consist of 15% outside majors
- Information is the 2nd most served outside major

- Underlined majors are considered STEM by USDHS

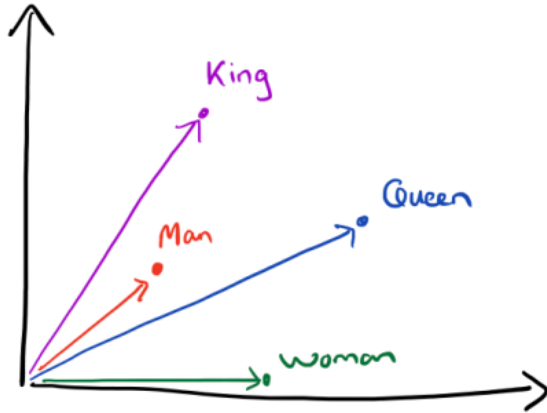
Method 2

Similarity

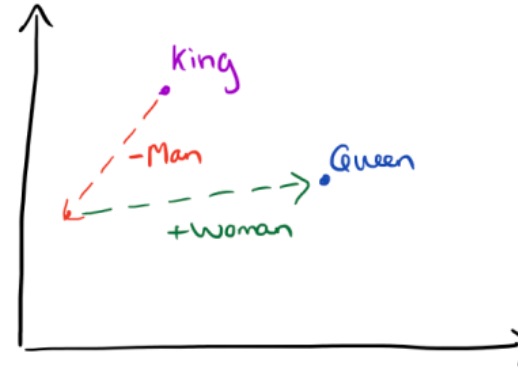
$$\cos(\mathbf{a}, \mathbf{b}) = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\|_2 \|\mathbf{b}\|_2}$$



Skip-gram model



Word
Vectors



Vector
Composition

Le, Q. V., & Mikolov, T. (2014, June). Distributed Representations of Sentences and Documents. In *ICML* (Vol. 14, pp. 1188-1196).
Adrian Colyer, <https://blog.acolyer.org/2016/04/21/the-amazing-power-of-word-vectors/>

Method 2 results

What departments are cosine closest to the iSchool?

Rank	Department	Similarity
1	New Media	0.610
2	Business Administration (MBA)	0.541
3	Development Practice	0.518
4	Electrical Engineering ₁	0.494
5	Computer Science	0.491
6	Industrial Engineering and Operations Research ₄	0.488
7	Public Policy	0.439
8	Vision Science	0.425
9	Education	0.410
10	Mechanical Engineering ₉	0.403

- Non-STEM are the top three departments

Method 2 results

What departments combined together are closest to the iSchool?

Rank	Department 1	Department 2	Cosine Similarity
1	Development Practice	New Media	0.721
2	Business Administration (MBA)	Computer Science	0.710
3	Computer Science	Development Practice	0.706
4	Business Administration (MBA)	New Media	0.678
5	Computer Science	Public Policy	0.670
6	New Media	Public Policy	0.665
7	Electrical Engineering	New Media	0.656
8	Industrial Engineering and Operations Research	New Media	0.650
9	Computer Science	Journalism	0.646
10	Computer Science	Education	0.640

- Two sides to the iSchool
 - Non-STEM combination of New Media and Dev Practice
 - Other side: Business Administration and Computer Science

Method 2 results

What departments are cosine closest to Computer Science?

Rank	Department	Similarity
1	Electrical Engineering ₄	0.649
2	Information	0.491
3	Mathematics	0.489
4	Industrial Engineering and Operations Research ₆	0.485
5	Engineering	0.475
6	Statistics	0.465
7	New Media ₁	0.420
8	Cognitive Science	0.386
9	Physics	0.382
10	Mechanical Engineering ₁₀	0.325
77	Education	-0.011

- Distant from social sciences and humanities like Education

Method 2 results

What departments combined together are closest to Computer Science?

Rank	Department 1	Department 2	Cosine Similarity
1	Information	Mathematics	0.633
2	Mathematics	New Media	0.614
3	Cognitive Science	Industrial Engineering and Operations Research	0.602
4	New Media	Statistics	0.597
5	Mathematics	Industrial Engineering and Operations Research	0.593
6	Linguistics	Industrial Engineering and Operations Research	0.587
7	Information	Statistics	0.580
8	Information	Industrial Engineering and Operations Research	0.566
9	New Media	Industrial Engineering and Operations Research	0.565
10	Cognitive Science	Mathematics	0.562

- Computer Science = Information + Mathematics

Conclusions

- The UCB iSchool has the distinguishing characteristic of bridging technology and the social science and humanities
 - iSchool outside majors are 1/3rd non-STEM vs. CS which are 1/20th
 - Education is 9th cosine most similar to the iSchool but 77th to CS
- The iSchool and CS have much value exchanged between one another
 - CS is the iSchool's top most served outside major and the iSchool is CS's 2nd
- The UCB iSchool is 2nd closest to a combination of CS and Business, as suggested by Wu et al. (2012), and CS is closest to a combination of Information and Mathematics.

Thank You!

Zach Pardos (zp@berkeley.edu)

Andrew Nam (jhnam95@berkeley.edu)