

Yeounoh Chung

yeounoh\_chung@brown.edu

Scott Houde

scott\_houde@brown.edu

Johannes Novotny

johannes\_novotny@brown.edu

Erfan Zamanian

erfanz@cs.brown.edu

**Goal:** To provide an intuitive web site merging three main data sets together that allows potential applicants to find the right Computer Science doctoral program for them.

### Citation Network Dataset<sup>1</sup>

Extracted from DBLP, ACM and other citation sources, contains more than 4,354,534 citations. We reduced this down by faculty and transformed paper abstracts into a TF/IDF format.



Entity resolution was tricky and not all professors were in the dataset!

### Faculty Professor Dataset

Generated as part of a class assignment contains 2,195 faculty from 55 different universities.<sup>2</sup> We enhanced this by adding a generated faculty ID and existing Institution ID.<sup>3</sup>

Unique Faculty ID's

IPEDS Institution ID's

### Integrated Postsecondary Education Data System (IPEDS)<sup>4</sup>

Every US Institution is required to submit this data every year to the government. Highly accurate but required filtering and interpretation. We used Python scripts to reduce the data down to institution ID's and award levels we are interested in.

### Final Dataset

#### TF/IDF Data

For every unigram and bigram in every abstract for our faculty we store the term, the faculty ID and the frequency, sorted from most frequent to least.

Faculty Name,  
Institution, Sub-Field,  
Join Year, PhD earned  
Inst, etc...

#### All Tables by Institution ID

General Institution fields

Current Enrollment by student level, gender and race

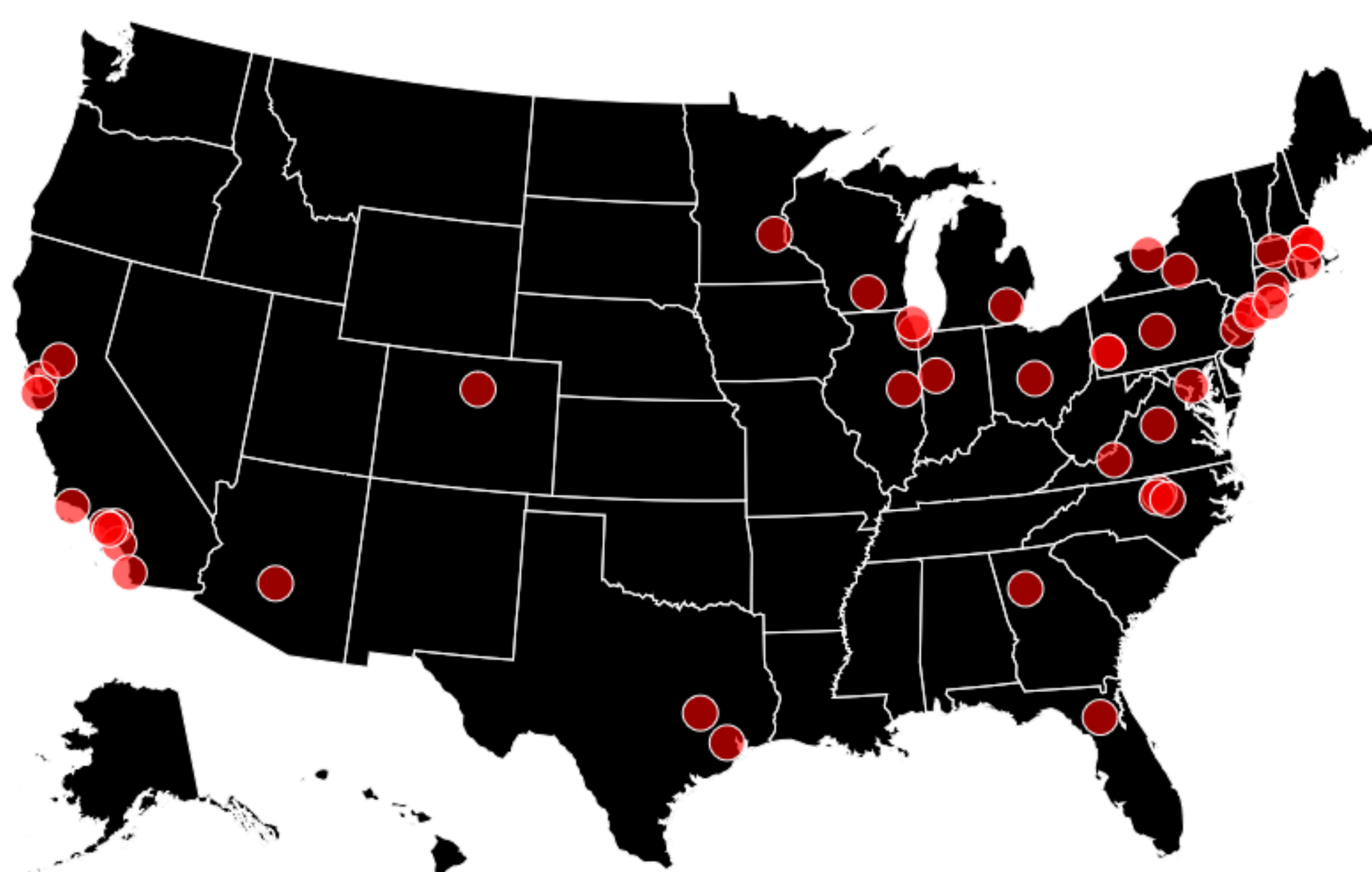
Computer Science doctoral completions, for multiple years

### Visualization and Interactive Search

#### Main Page



Prof Finder

 Search


#### Search Results

Prof Finder

 Search

Ugur Cetintemel (Professor)

Brown University (Providence, RI)  
Subfields: Database Systems, Mobile and Ubiquitous Computing, (+3 more)  
Recent Publication: The End of Slow Networks: It's Time for a Redesign (2015), arXiv preprint arXiv:1504.01048 (42 Publications found)

Avi Silberschatz (Professor)

Yale University (New Haven, CT)  
Subfields: Database Systems, Operating Systems, Network Management, (+1 more)  
Recent Publication: The End of Slow Networks: It's Time for a Redesign (2015), arXiv preprint arXiv:1504.01048 (42 Publications found)

Tim Kraska (Assistant Professor)

Brown University (Providence, RI)  
Subfields: Database Systems, Cloud Computing  
Recent Publication: The End of Slow Networks: It's Time for a Redesign (2015), arXiv preprint arXiv:1504.01048 (42 Publications found)

Page < 1 >

#### Faculty Detail

Prof Finder

 Search

Ugur Cetintemel (Professor)

Research Fields: Database Systems, Mobile and Ubiquitous Computing, Operating Systems and Distributed Systems, Big Data Tools and Systems, Real-time, Continuous Analytics

Institution

Brown University - CS Department

Providence, Rhode Island 02912

USA

Publications

Andrew Croft, Alex Galakatos, Kayhan Dursun, Tim Kraska, Ugur Cetintemel, Stanley B. Zdonik

Topicware: "Big" Data, Big Analytics, Smart Closures, CDR 2015

The End of Slow Networks: It's Time for a Redesign, CoRR abs/1503.01143 (2015)

Elhanan Dadasheh, Ugur Cetintemel, Tim Kraska

Putting Analytics on the Spot: Or How to Lower the Cost for Analytics, IEEE Internet Computing 18(5): 70-73 (2014)

Ugur Cetintemel, Jang Du, Tim Kraska, Samuel Madden, David Maier, John Meenan, Andrew Pavlo, Michael Stonebraker, Erik Sutherland, Nesime Tatbul, Kristin Tufte, Hao Wang, Stanley B. Zdonik

S-Store: A Streaming MapReduce System for Big Velocity Applications, PVLDB 7(13): 1633-1636 (2014)

Z. Meral, Ozgur Ugur, Ugur Cetintemel, Nishu N. Dalvi, Hank Korth, Anthony K. H. Tung

Special issue on best papers of VLDB 2012, VLDB J. 25(2): 173-174 (2014)

Jen-Chieh Duggan, Olga Papadimitriou, Ugur Cetintemel, Elhanan Dadasheh

Contentware: A Resource Mapping Approach for Concurrent Query Performance Prediction, EDBT 2014: 109-120

1)Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, and Zhong Su. ArnetMiner: Extraction and Mining of Academic Social Networks. In [Proceedings of the Fourteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining \(SIGKDD'2008\)](#). pp.990-998.

3) Postsecondary Institutional Codes crosswalk table [http://www.pesc.org/interior.php?page\\_id=145](http://www.pesc.org/interior.php?page_id=145)

2) Analysis of Over 2,000 Computer Science Professors at Top Universities , [http://jeffhuang.com/computer\\_science\\_professors.html](http://jeffhuang.com/computer_science_professors.html)

4) National Center of Education Statistics, <http://nces.ed.gov/ipeds/datacenter/DataFiles.aspx>