

Conference Speaker Biography Analysis

Shiau Chu Heng, Shide Adibi, Chuyi Sun, Md Kamal Hossain

Q: Do the self-presentation of male and female researchers differ? In what ways?

Basic Methodology

- **Data:** Biographies of speakers in computer science conferences
 - A concise self-introduction to be read by fellow academics
 - Should contain information the speaker considers to be important about their career
 - No unified format across conferences
- **Metrics:**
 - Cosine similarity: simple way to quantify difference between texts
 - Term frequency: what information gets emphasized over and over again?

Basic Methodology

- **Most research into gender representation in computer science focuses on the male-to-female ratio, and the larger societal trends that causes less women entering the field**
 - Many compiled statistical reports are also available from government organizations (Bureau of Labor etc...)
- **Most research are also focused on the computer science workforce instead of academia**
- **Much less research on how women already in computer science view their roles**
- **“Women in Technology” report: highlights the unique barriers facing women working in the technology sector**
 - Do the same barriers exist in academia?
 - Does this affect how women researchers presents themselves and how they are presented by others?

Data Collection

The screenshot shows the RECOMB/ISCB Conference website. The left sidebar contains a navigation menu with items like Overview, Schedule, Keynote, Papers, Invited Talks, Invited Speakers Biographies, Panels, Posters & Visualizations, Tutorials, Workshops, Birds-of-a-Feather, SCinet Research Sandbox, Awards, and Award Recipients. The main content area is titled 'Invited Speakers Biographies' and lists several speakers. The first speaker is William J. Harrod, DOE Office of Advanced Scientific Computing Research. The second is Henry Markram, Swiss Federal Institute for Technology (EPFL). The third is Steve Scott, Nvidia. The fourth is Mitsuo Yokokawa, RIKEN.

Home » Technical Program »

Invited Speakers Biographies

INVITED PLENARY SPEAKERS

William J. Harrod, DOE Office of Advanced Scientific Computing Research
Dr. William Harrod is the Research Division Director in the Advanced Scientific Computing Research Division at the Department of Energy (DOE). The mission of the Advanced Scientific Computing Research Division is to discover, develop, and deploy computational and networking capabilities to analyze, model, and simulate phenomena important to the Department of Energy (DOE). Before joining DOE, Dr. Harrod was the Information Processing Technology Office (IPTO) in the Defense Advanced Research Projects Agency. He received his PhD in Mathematics from the University of Tennessee at Knoxville.

Henry Markram, Swiss Federal Institute for Technology (EPFL)
Henry Markram is a professor of neuroscience at the Swiss Federal Institute of Technology (EPFL). He is the founder of the Brain Mind Institute, founder and director of the Blue Brain Project, and co-founder and president of Frontiers. His work focuses on understanding the human brain, and he has been recognized as one of the 3 of the most fundamental laws governing plasticity of synaptic connections, the microcircuitry of the neocortex, co-developed a new theory of how the brain works, and co-developed the Intense World Theory of Autism. In the Blue Brain Project, Henry and his team are working to simulate the human brain, cell for cell, synapse for synapse, molecule for molecule. He is one of the highest cited neuroscientists in the world in his age group.

Steve Scott, Nvidia
Dr. Steve Scott is Chief Technology Officer of the Tesla business unit at NVIDIA, where he is responsible for the NVIDIA's GPU computing roadmap. Prior to joining NVIDIA in August 2011, Steve spent 19 years at Intel, where he was the Chief Architect of multiple systems at Intel, architected the routers for the Cray and led the Cray Cascade project funded by the DARPA High Productivity Computing Systems program. He has received numerous US patents, and has served on numerous advisory boards and program committees. He was the recipient of the 2004 Maurice Wilkes Award and the 2005 IEEE Seymour Cray Computer Engineering Award. He received his PhD in Computer Architecture in 1992 from the University of Wisconsin at Madison, where he was a Wisconsin Hertz Foundation Fellow.

Mitsuo Yokokawa, RIKEN
Mitsuo Yokokawa is the director of the Operations and Computer Technologies Division, RIKEN Advanced Institute for Computational Science at present. After he received his B.Sci and M.Eng from University of Tsukuba, he joined the Computer Science Department at the University of Tsukuba.

The screenshot shows the RECOMB/ISCB Conference website. The top navigation bar includes links for JOIN ISCB, KEY DATES, NEWS, and REGISTER. The main content area is titled 'Invited Speakers Biographies' and lists several speakers. The first speaker is Trey Ideker, University of California, San Diego. The second is Steve Scott, Nvidia. The third is Mitsuo Yokokawa, RIKEN.

JOIN ISCB KEY DATES NEWS REGISTER

RECOMB/ISCB Conference on Regulatory and Systems Genomics, with DREAM Challenges

KEYNOTE SPEAKER ABSTRACTS & BIOGRAPHIES

updated Oct 27, 2013

DREAM Challenges

Trey Ideker
University of California, San Diego
United States

Turning Networks Into Ontologies: Towards A Data-driven Gene Ontology

Abstract: Ontologies have been very useful for capturing knowledge as a hierarchy of concepts and their interrelationships. In biology, a prime challenge has been to develop ontologies of gene function given only partial biological knowledge and inconsistency in how this knowledge is curated by experts. I will discuss how large networks of gene and protein interaction, as are being mapped systematically for many species, can be transformed to assemble an ontology with equivalent coverage and power to the manually-curated Gene Ontology (GO). Our network-extracted ontology contains 4,123 biological concepts and 5,766 relations, capturing the majority of known cellular components as well as many additional concepts, triggering subsequent updates to GO. Using genetic interaction profiling we provide further support for novel concepts related to protein trafficking, including a link between Nnf2 and YEL043W. This work enables a shift from using ontologies to evaluate data to using data to construct and evaluate ontologies.

Biography: Trey Ideker, Ph. D. is Professor of Medicine at the University of California at San Diego. He serves as Division Chief of Medical Genetics and Director of the National Resource for Network Biology, as well as being Adjunct Professor of Bioengineering and Computer Science and Member of the Moores UCSD Cancer Center. Ideker received Bachelor's and Master's degrees from MIT in Electrical Engineering and Computer Science and his Ph.D. from the University of Washington in Molecular Biology under the supervision of Dr. Leroy Hood. He is a pioneer in assembling genome-scale measurements to construct network models of cellular processes and disease. His recent research activities include assembly of networks governing the response to DNA damage; development of the Cytoscape and NetworkBLAST software packages for biological network visualization and cross-species network comparison; and methods for identifying network-based biomarkers in development and disease. Ideker serves on the Editorial Boards for *Bioinformatics* and *PLoS Computational Biology*, is on the Scientific Advisory Boards of the Sanford-Burnham Medical Research Institute and the Institute for Systems Biology, and is a regular consultant for companies such as Monsanto and Mental Rintechology. He was named one of the Top 100

invited speakers: 5

Invited speakers: 19

- Collects biography for each scientist from the conference websites.

Data Collection

- Extract all the significant information out of each conference such as biography, gender, publish year, Name of speaker,

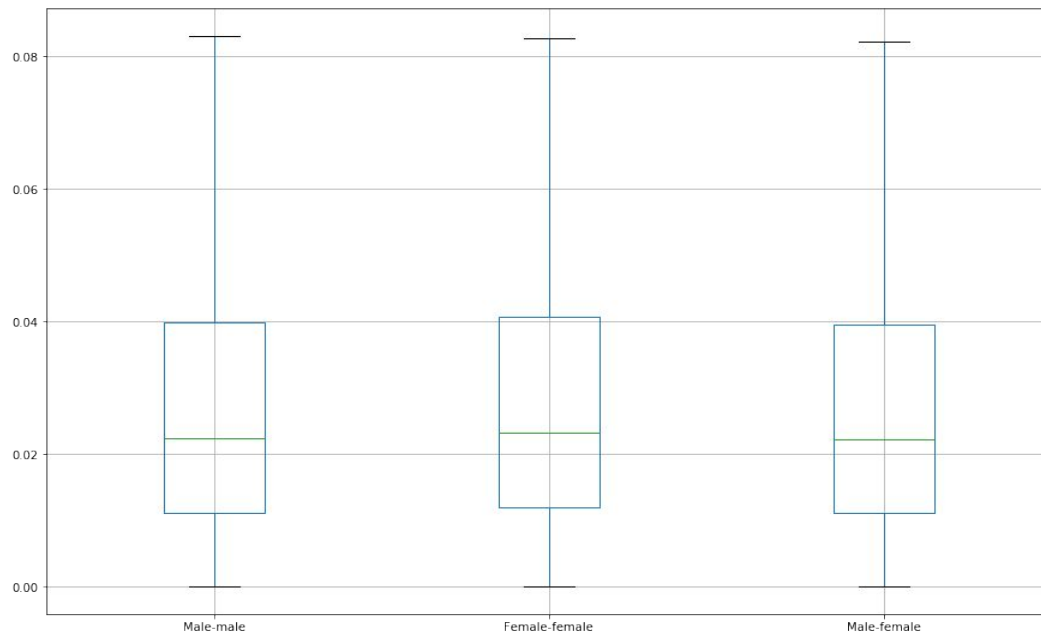
	A	B	C	D
1	Bio	Gender	Year	Name
2	Kathryn S. McKinley is a Principal Research F	F	2016	Kathryn S McKinley
3	Dr Larry Persons PhD is on the Faculty at S M	M	2018	Larry Persons
4	Dr Arthur Shelley is an independent educato	M	2018	Arthur Shelley
5	Shane McCarthy Shane is the CEO of Blue	M	2018	Shane McCarthy
6	Stephen O'Leary is managing director at Oly	M	2018	Stephen O'Leary
7	Olav Lysne is a director of Simula Metropolit	M	2018	Olav Lysne
8	Wallace Chigona is a Professor in Informati	M	2018	Wallace Chigona
9	Dr .Iohannes Cronié is the Dean of Informati	M	2018	Dr .Iohannes Cronié

Preliminary Results in June

- Calculated cosine similarity on a small data set of 94 biographies (74 male, 20 female), with primitive preprocessing (small list of stop words, no stemming)
- Average cosine similarity of male-male pairs: 0.0331
- Average cosine similarity of female-female pairs: 0.0588
 - High similarity due to small sample size?
- Average cosine similarity of male-female pairs: 0.0255
- There does seem to be a slight difference between male and female biographies

Result: Cosine Similarity

- Cosine similarity calculated over a set of 191 bios (153 male, 38 female)
- Improved preprocessing via NLTK



Result: Cosine Similarity

- Average cosine similarity of male-male pairs: 0.031
- Average cosine similarity of female-female pairs: 0.032
- Average cosine similarity of male-female pairs: 0.030
- Similar distribution
- **No significant difference!**

Result: Term Frequency

- Frequency of word stems in male and female bios

Male

comput	353
research	297
univers	280
scienc	242
professor	130
system	114
award	110
algorithm	98
work	93
includ	89

Female

research	84
comput	71
scienc	66
univers	64
award	44
professor	42
engin	27
receiv	25
learn	24
data	23

Result: Term Frequency

- Term frequency separated by part-of-speech tags (noun, adjective, verb, adverb) showed similar results: significant overlap of most common terms across genders.
- Calculated without removing stop words and stemming to ensure accuracy of tagger

Result: Document Frequency

- Frequency of word stems by documents in male and female bios

Male

research	121
univers	115
scienc	112
comput	109
professor	93
includ	65
work	60
receiv	58
institut	55
award	54

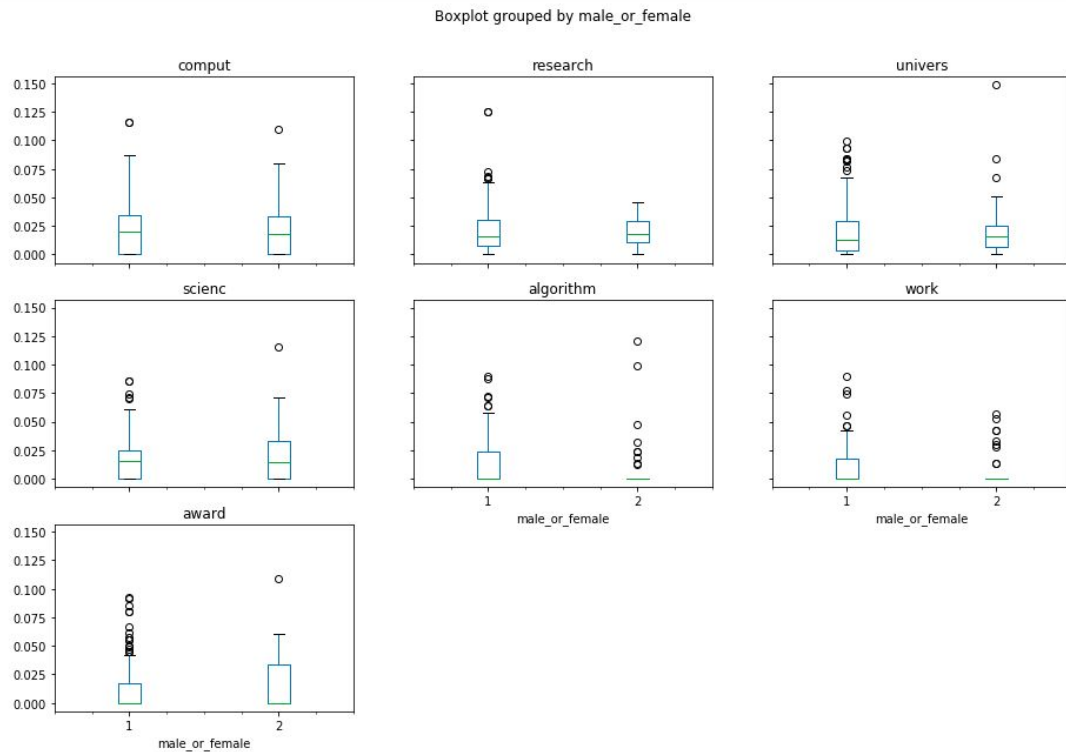
Female

research	33
univers	30
professor	28
comput	27
scienc	25
receiv	19
associ	16
engin	16
fellow	15
interest	14

Result: Document Frequency

- Document frequency separated by part-of-speech tags (noun, adjective, verb, adverb) showed similar results: significant overlap of most common terms across genders
- Again, calculated with minimal preprocessing

Result: TFIDF



Result: Mutual Information

➤ Top pairs of terms representing Male class

Term 1	Term 2
'aachen'	'cross'
'aachen'	'envoy'
'aberration'	'array'
'ababa'	'age'
'abstract'	'deeper'
'abstract'	'pure'
'academics'	'brothers'
'accenture'	'watch'
'adapted',	'serious'
address'	'bacterial'

➤ Top pairs of terms representing Female class

Term 1	Term 2
'ability',	'alzheimers'
'ability'	'anita',
'ability'	'disease'
'about'	'failure'
'about'	'smart'
'accounting'	'assist'
'accounting'	'algorithm'
'activities'	'it'
'activities',	'leadership'

Conclusion: No significant difference exists in the language used by male and female computer scientists in their conference biographies.

Final Thoughts

- **Related research shows large gender differences in the tech industry**
 - Academic advancement (based on blind peer reviews) more objective than metrics for climbing the “corporate ladder”?
- **More significant differences across conferences than across gender?**
 - Individual “conference traditions”, since no unified **format across conferences exist?**
- **Does the result hold for older conferences?**
 - Would the presumably larger gender gap make this harder to calculate?
- **The importance of having a large enough data set**
- **Format standards (or lack thereof) sometimes have a larger influence than the data they carry!**
- **Different preprocessing procedures for different calculations**



Thanks for your
attention