Faculty Fellows Inspiring the Next Generation of Computer Scientists

Moderator: Rane Johnson-Stempson
Lucy Sanders, CEO NCWIT
Emma Brunskill, Carnegie Mellon
Magdalena Balazinska, University of Washington
Miriah Meyer, University if Utah
Wei Wang, University of California, Los Angeles

July 16, 2012

Agenda

Topic	Speaker
Overview current situation	Rane Johnson and Lucy Sanders
Introduction & Research Agenda	Emma Brunskill, Carnegie Mellon
Introduction & Research Agenda	Magdalena Balazinska, University of Washington
Introduction & Research Agenda	Miriah Meyer, University if Utah
Introduction & Research Agenda	Wei Wang, University of California, Los Angeles
Panel Discussion	Moderated by Rane Johnson
Call to Action	Emma, Magdalena, Miriah, Wei, Lucy and Rane
Audience Q & A	All attendees

Why Women in Computing Matters?

1.4 MILLION
COMPUTING JOBS WILL OPEN IN THE U.S. BETWEEN 2008-2018

ONLY **18%** OF UNDERGRADUATE **COMPUTING** AND INFORMATION SCIENCES DEGREES WERE AWARDED TO **WOMEN**

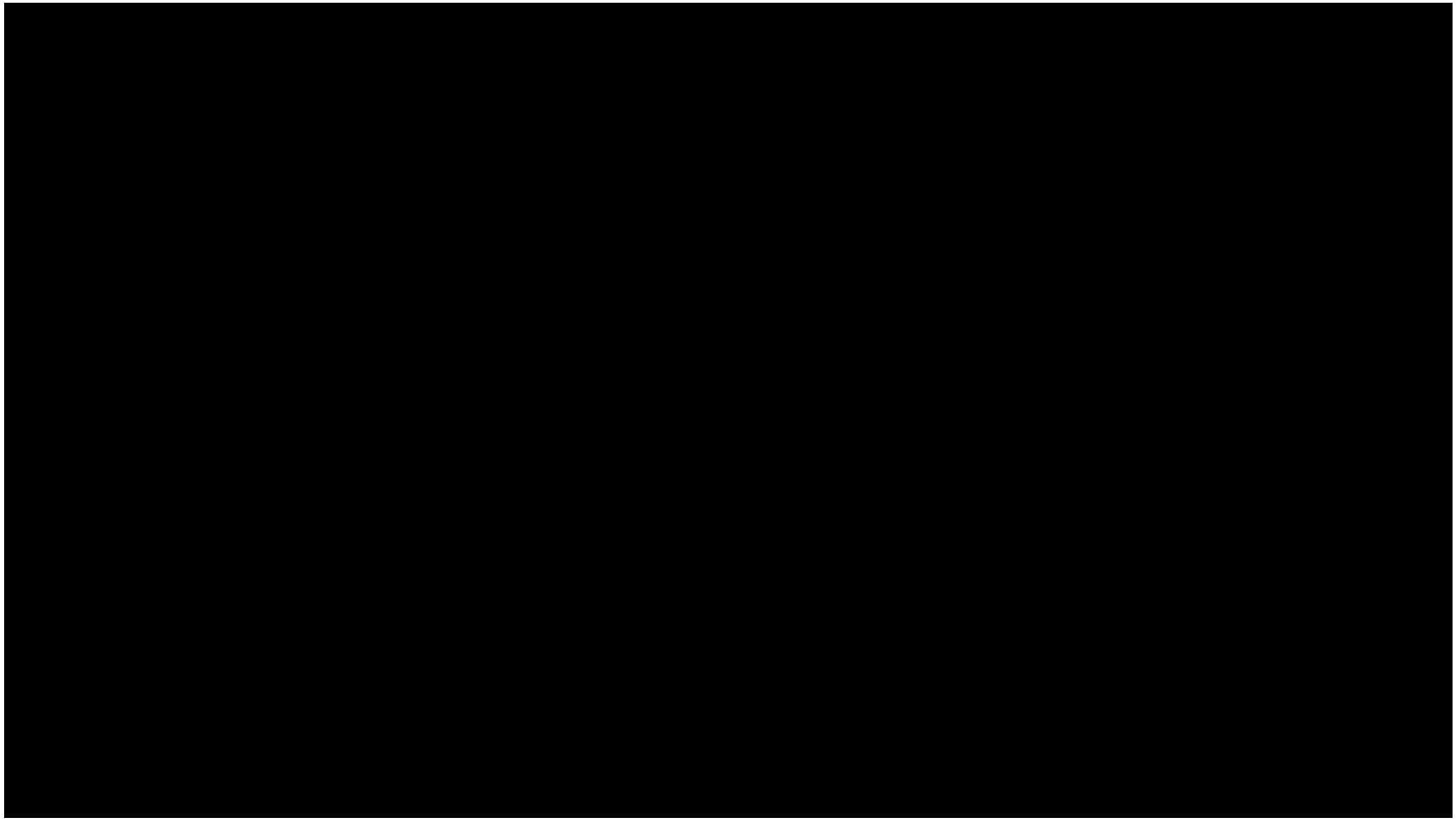
LESS THAN **30%** OF **WOMEN** COMPLETE **MASTERS** OR **DOCTORATE** DEGREES

46% OF **AP** TEST TAKERS ARE **GIRLS** ONLY **19%** TAKE **CS AP**

2x **WOMEN** LEAVE COMPUTING CAREERS THEN **MALES**

34% HIGHER RETURN ON INVESTMENT WHEN **WOMEN** ARE IN **MANAGEMENT**

Faculty Fellows



Microsoft Research Faculty Fellowship Program

Each year since 2005, Microsoft Research has recognized innovative, promising new faculty members from a number of research institutions to join the ranks of Microsoft Research Faculty Fellows.

This program now encompasses more than 50 academic researchers whose exceptional talent for research and innovation in computer science identifies them as emerging leaders in their fields.

The selected professors are exploring breakthrough, high-impact research that has the potential to help solve some of today's most challenging societal problems.

Each fellowship includes a cash award and access to other Microsoft resources such as software, invitations to conferences, and engagements with Microsoft Research.







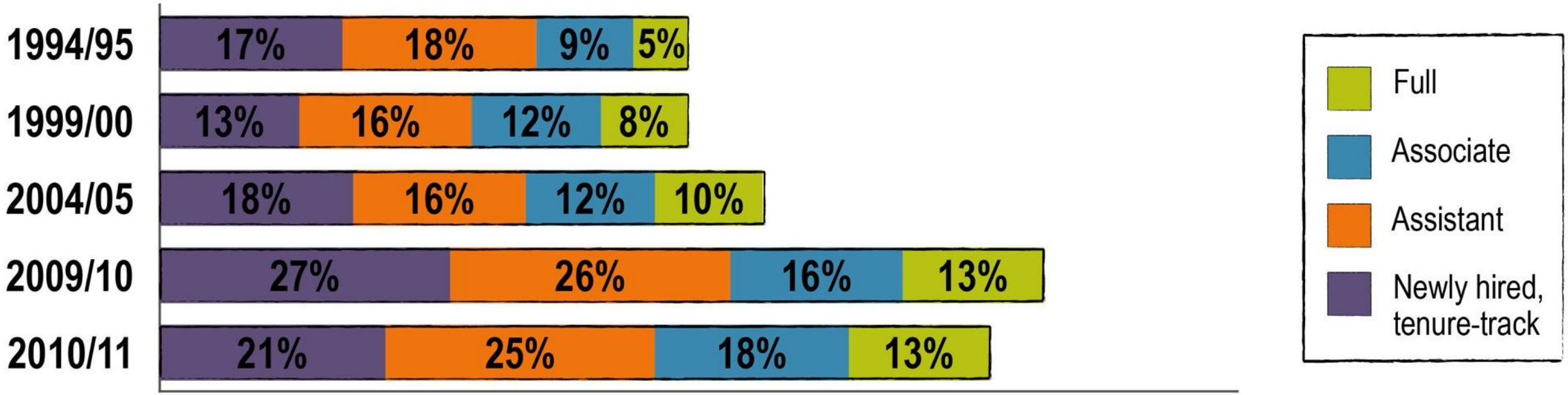
Faculty Fellows Inspiring the Next Generation of Computer Scientists

Lucy Sanders, CEO NCWIT

July 16, 2012

Female Computing Faculty: The Stats

FEMALE PERCENTAGE OF COMPUTER SCIENCE FACULTY AT PHD-GRANTING INSTITUTIONS, 1995-2011



 Source: CRA Taulbee Survey, 1994-95, 1999-00, 2004-05, 2009-10, 2010-11.

The Pipeline to Professorship

24% of CIS PhDs were awarded to women in 2010 (**19.8%** of Taulbee schools).

More than **50% of women** in CS or CE doctoral programs think about leaving before the end of their second year (vs. 35% of men.) Women who observed or experienced sexism **leave at a 10X greater rate**.

Sources: IPEDS; 2012 CRA Taulbee Survey; CRA-W report on CRA-W Grad Cohort, conducted by Joanne Cohoon and Holly Lord.

The Pipeline to Tenure

- Women's authorship of technical conference papers increased from **7%** in 1967 to **27%** in 2009.
- Institutions are putting programs into place to support advancement of women faculty (**stopping tenure track for family leave, mentoring**, etc.).
- Women still get asked to do more service for things-related-to-women and spend more time taking care of home and family.

Source: Cohoon, J.M., Nigai, S., & Kaye, J. (2011, August). Gender and computing conference papers. *Communications of the ACM, 54*(8). http://cacm.acm.org/magazines/2011/8/114945-gender-and-computing-conference-papers/fulltext

Case Study – Georgia Tech

Mentoring Program for Faculty Advancement

www.ncwit.org/gatechmentoring

1) Interactive tool, "Navigate Your Career," developed by Carol Colatrella

2) Mentoring and career coaching for faculty women

3) Cross-college workshops (leadership, obtaining and managing individual and collaborative grants, and life-work balance)

4) Integrated evaluation. Chief criterion of success is "faculty advancement" at GT.

Departments Can Take Action

Use clear criteria for hiring and promotion.

Unconscious gender bias can mislead both men and women to make inaccurate judgments in hiring, performance evaluation, and tenure promotion.

Identify and dismantle institutional biases.

Removing institutional barriers (“just the way things are around here”) benefits all of an organization's members, not just the minority group.

Offer a support network to women (and men.)

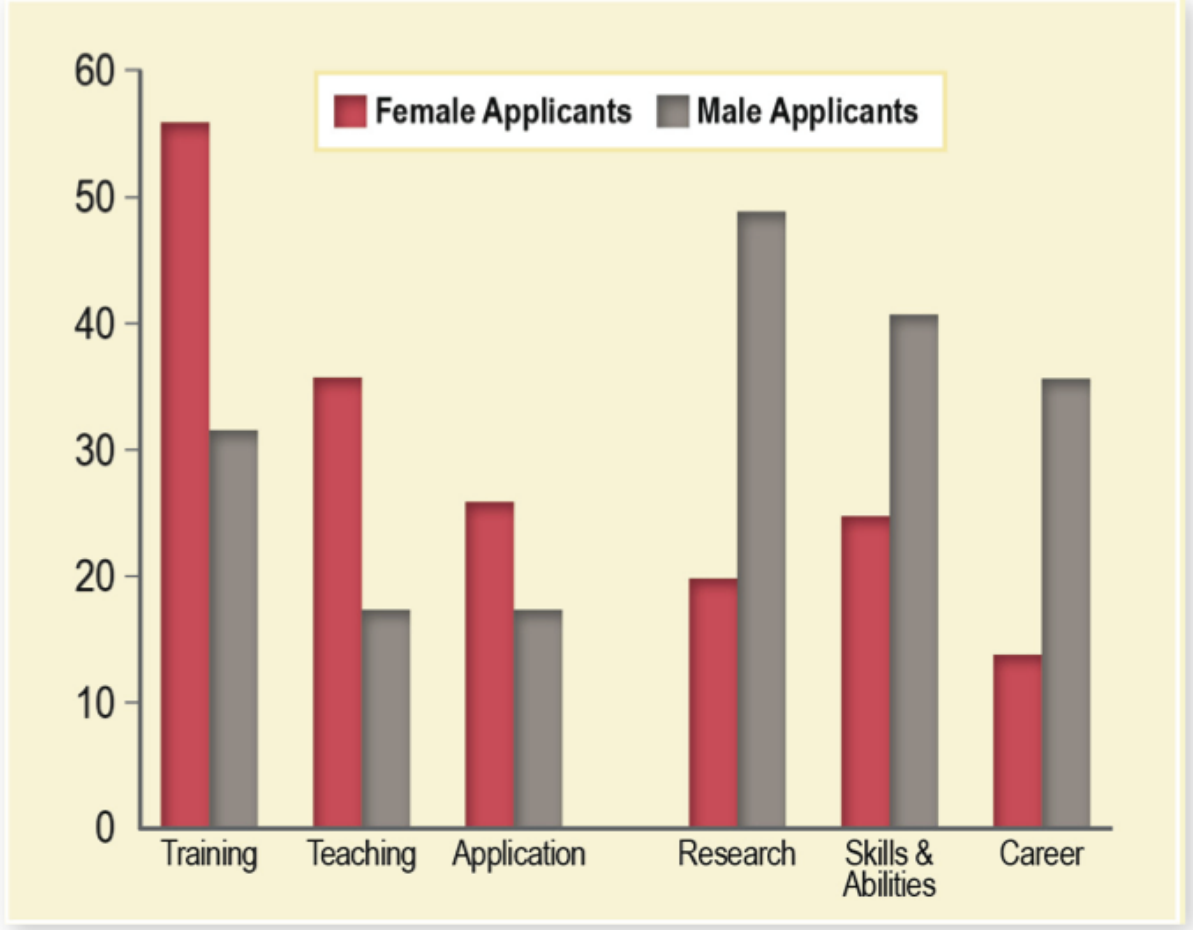
Mentoring, policies for flexibility, and administrative support for gender diversity can have significant impacts for both female and male faculty.

Understand the Research

Recommendation Letters

"Her" – more often followed by "training," "teaching," or "personal life"

"His" more often followed by "research," "skills," and "publications"



Source: Trix, F. & Psenka, C. (2003). "Exploring the color of glass: Letters of recommendation for female and male medical faculty." Discourse & Society, 14(2), 191-220.

Organizational Resources

CRA / CRA-W

Advanced Career Mentoring Workshops (CAPP), REU programs, and ProfessHers mailing list encourage recruitment and retention.

ACM / ACM-W

ACM Women in Computing and ACM newsletters featuring female faculty topics keep women's careers top of mind. Regional conferences provide encouragement.

NSF ADVANCE

Developing systemic approaches to increasing representation and advancement of women in STEM careers; 30 Institutional Transformation sites.

NCWIT Resources

Programs-in-a-Box

Mentoring-in-a-Box for faculty women, REU-in-a-Box, and Supervising-in-a-Box series provide tips for improving recruitment and retention.

Practices

Practices for combating sexism, creating neutral spaces, and reducing bias in letters of recommendation, hiring, and promotion.

Talking Points and Top Ten Ways

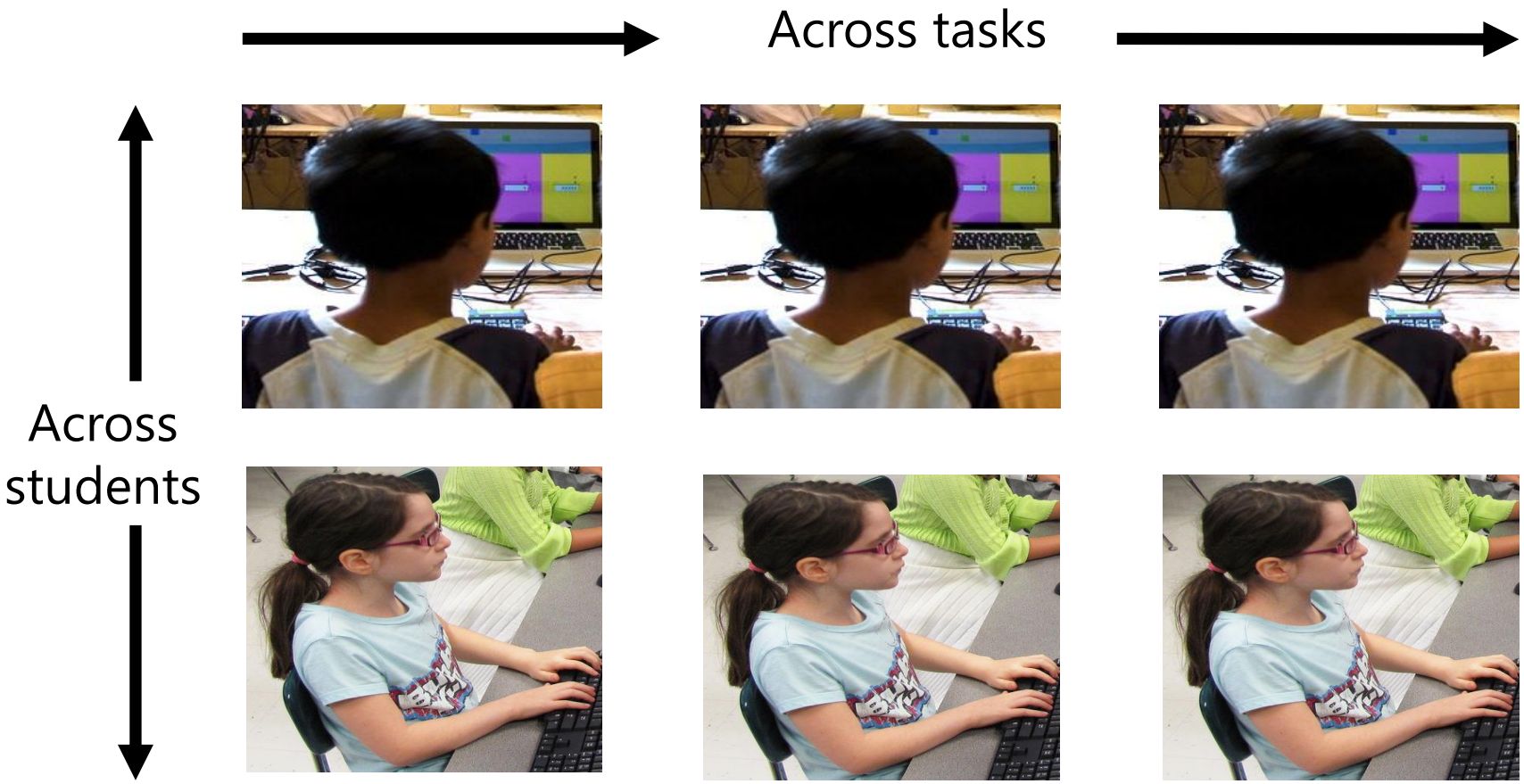
Identifying institutional barriers, and recognizing that men in positions of power can help create positive change.

Emma Brunskill

Carnegie Mellon


Education = Opportunity

Tutors & Lifelong AI Agents



Computational Thinking
in
K-12 Education

MSR Faculty Fellow Mom

Magdalena Balazinska

Computer Science and Engineering

University of Washington

http://www.cs.washington.edu/people/faculty/magda

Magdalena Balazinska

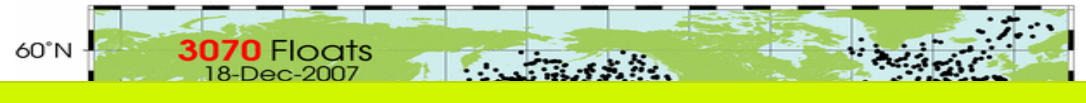
- Assistant Prof… soon to be Associate Prof. ☺
- U. of Washington - Computer Science Department
- PhD from MIT in 2006
- Research in Database Systems
  - Distributed stream processing (Borealis)
  - Imprecise, sensor data management (RFID Ecosystem+ Lahar)
  - Big-data analytics, focus on science (Nuage,SciDB)
  - Database usability (CQMS)
  - Interactions between pricing and data mgmt (Data Eco$y$tem)
- Some awards and honors
  - NSF CAREER 2009
  - 10 year most influential paper award 2010
  - A couple of best-paper awards 2002, 2010, 2011, 2011
  - Microsoft faculty fellow 2007

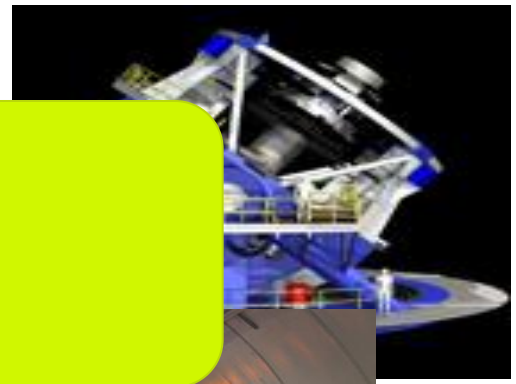
Science is Facing a Data Deluge!

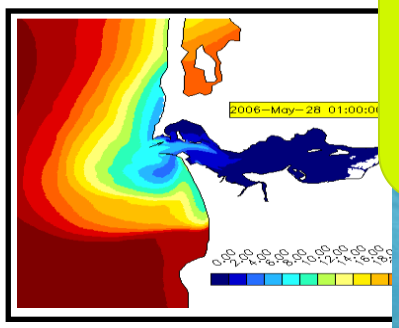
- Astronomy: High-resolution, high-frequency sky surveys (SDSS, LSST)
- Medicine: ubic
- Biology: lab au
- Oceanograph
- Etc.

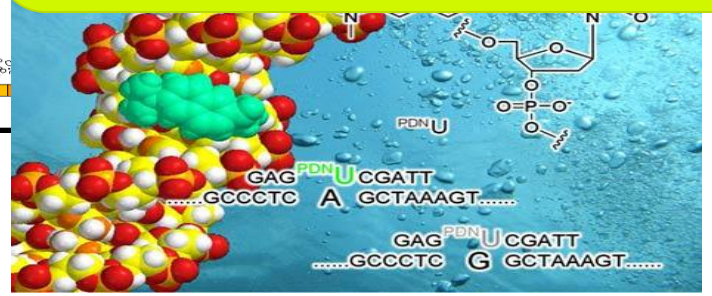
Scientists need new tools and techniques to effectively analyze all this data!

This need extends beyond science!



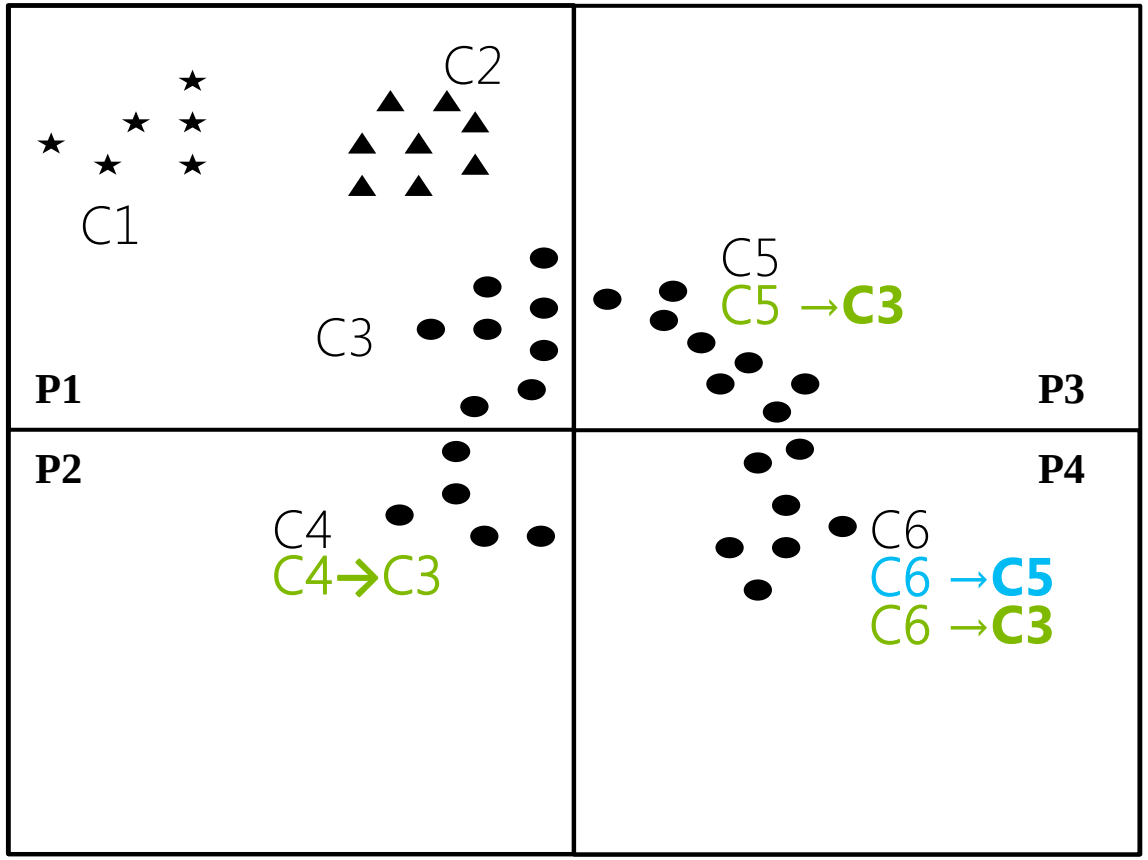




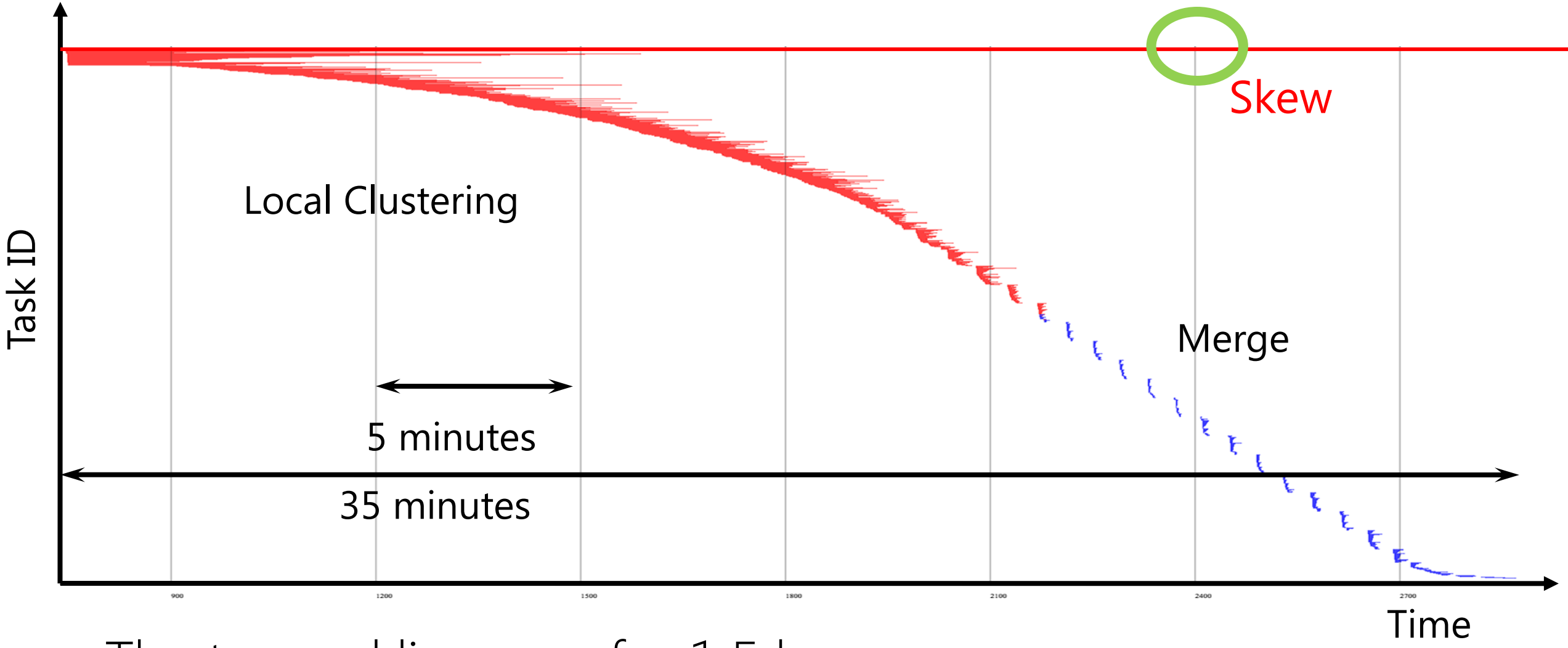


Why is This Hard: Example

- Partition
- Local clustering
- Merge
  - P1-P2
  - P3-P4
- Merge
  - P1-P2-P3-P4
- Finalize
  - Annotate original data



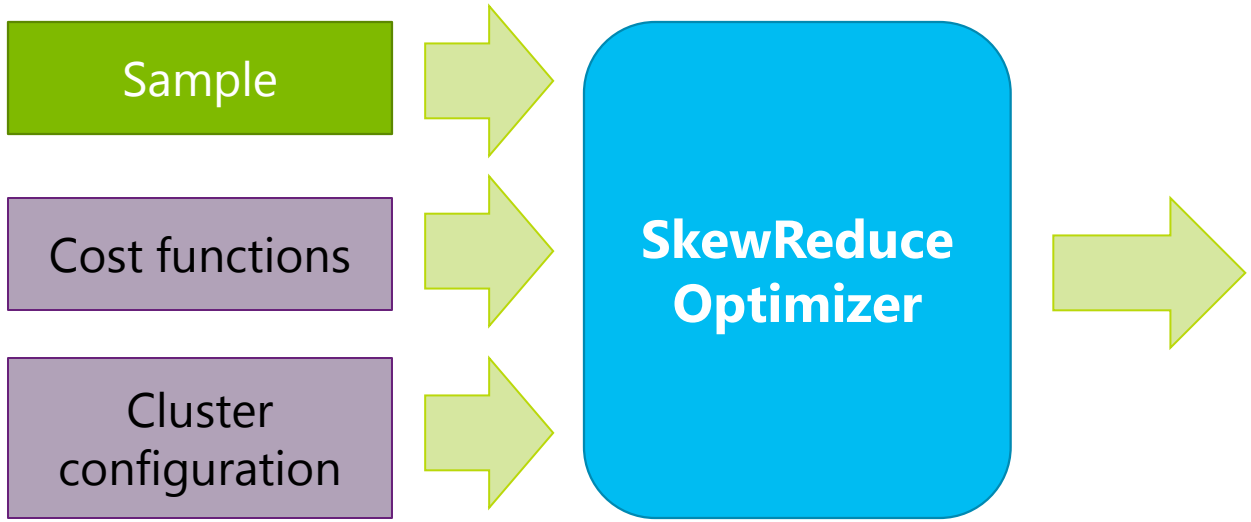
Problem: Skew

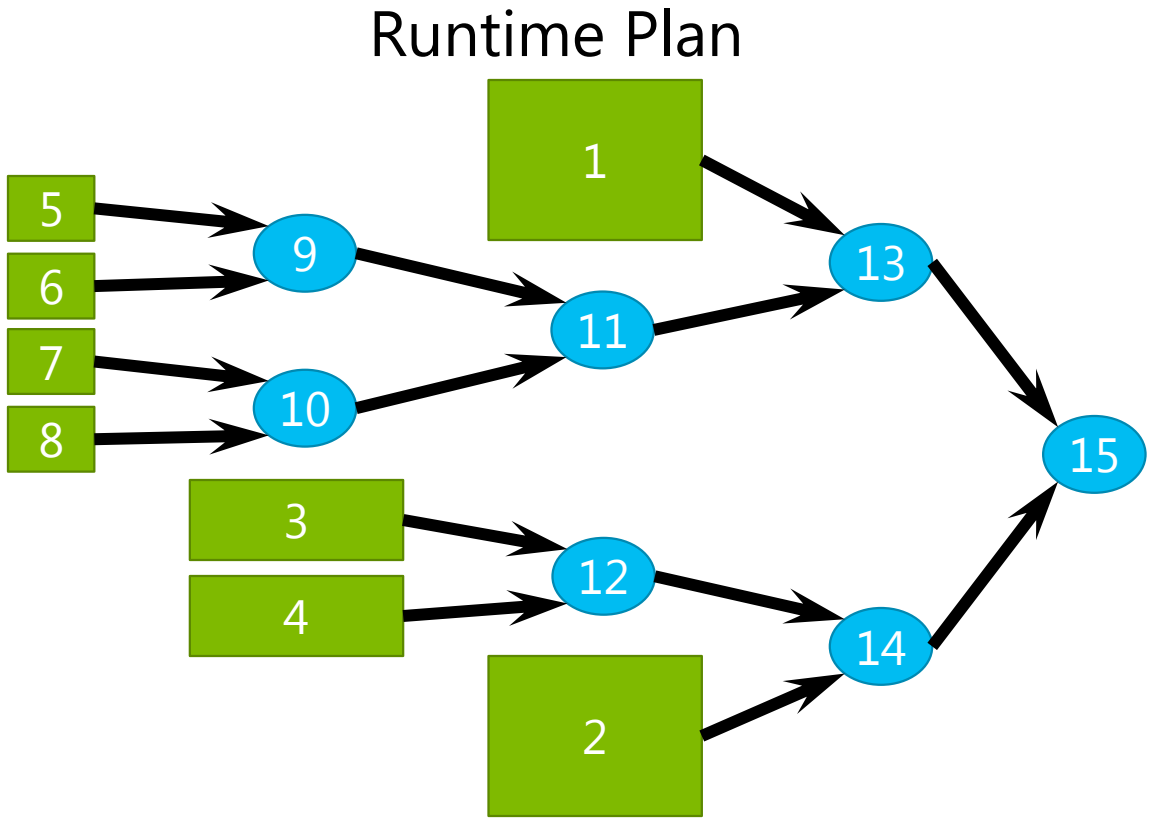


- The top red line runs for 1.5 hours

SkewReduce

[SOCC'10]





- *Goal: minimize expected total runtime*
- SkewReduce runtime plan
  - Bounding boxes for data partitions
  - Schedule

6X to 8X runtime improvement

SkewTune

[SIGMOD'12]

- Key ideas:
  - Ask nothing from the developer
  - Make skew handling completely transparent
  - Mitigate skew at runtime
- Key approach:
  - As long as everyone is doing useful work, do nothing
  - If resources idle, re-partition longest task remaining

HaLoop

[VLDB'10]

- Support for iterative processing in Hadoop
- If system is aware of iterations:
  - Can make developer's life easier with better API
  - Can improve performance
- Performance improvements:
  - Smart scheduling
  - Various forms of data caching

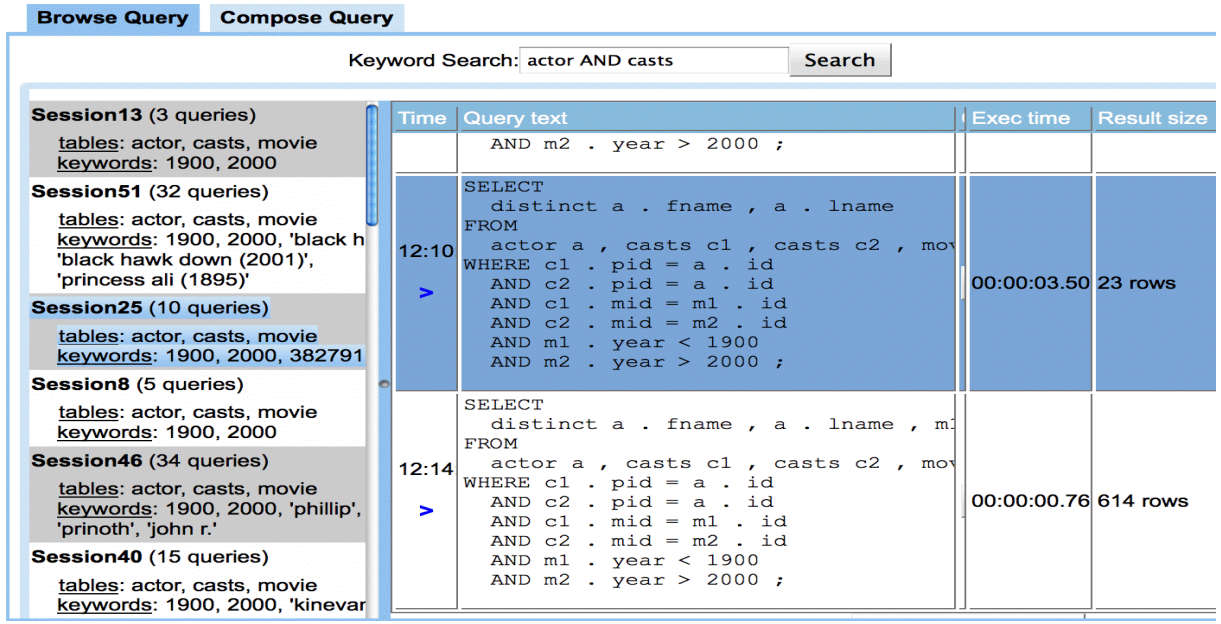
2X runtime improvement

SnipSuggest and SQB

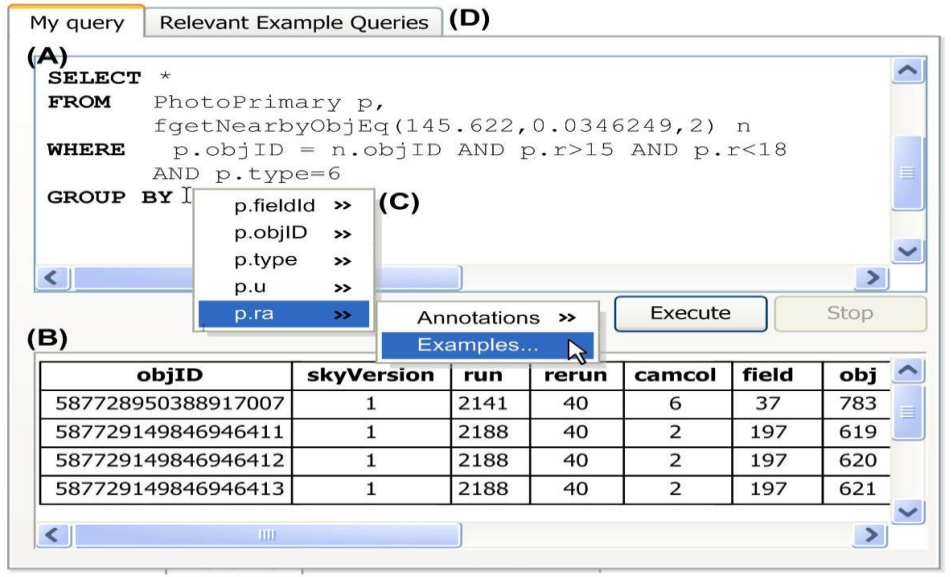
[VLDB'11]

- SQL Auto-complete and Query Browsing:
  - Takes current context into account
  - History-based domain-specific recommendations

SQB: Session-based SQL query browsing



SnipSuggest: SQL Autocomplete [VLDB'11]



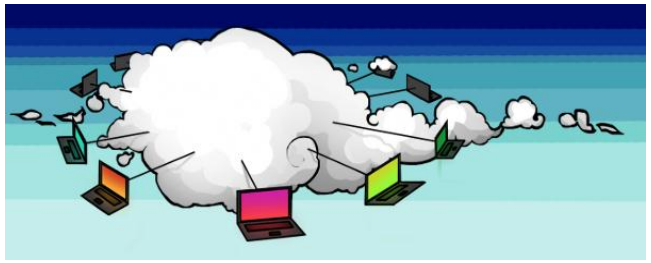
Summary of Current Big-Data Projects

High-Performance

Big-Data Analytics and

Cloud Computing (Nuage)

http://nuage.cs.washington.edu/



Easier Big-Data Analytics

(CQMS/Nuage)

http://cqms.cs.washington.edu/

Pricing and Data Management

(Data Eco$y$tem)

http://cloud-data-pricing.cs.washington.edu/

Other Stuff I Recently Did

New house (2010)

Built from scratch!

Two daughters (2007 and 2010)

How I Manage?

I get help from anyone who's willing to help!

I mix work and life

Business trip to UCSB

Conference in Greece

NSF Panel in DC

I do what I have to do
I ask for what I need

I don't try to be perfect
I just do what I can

Conference in France

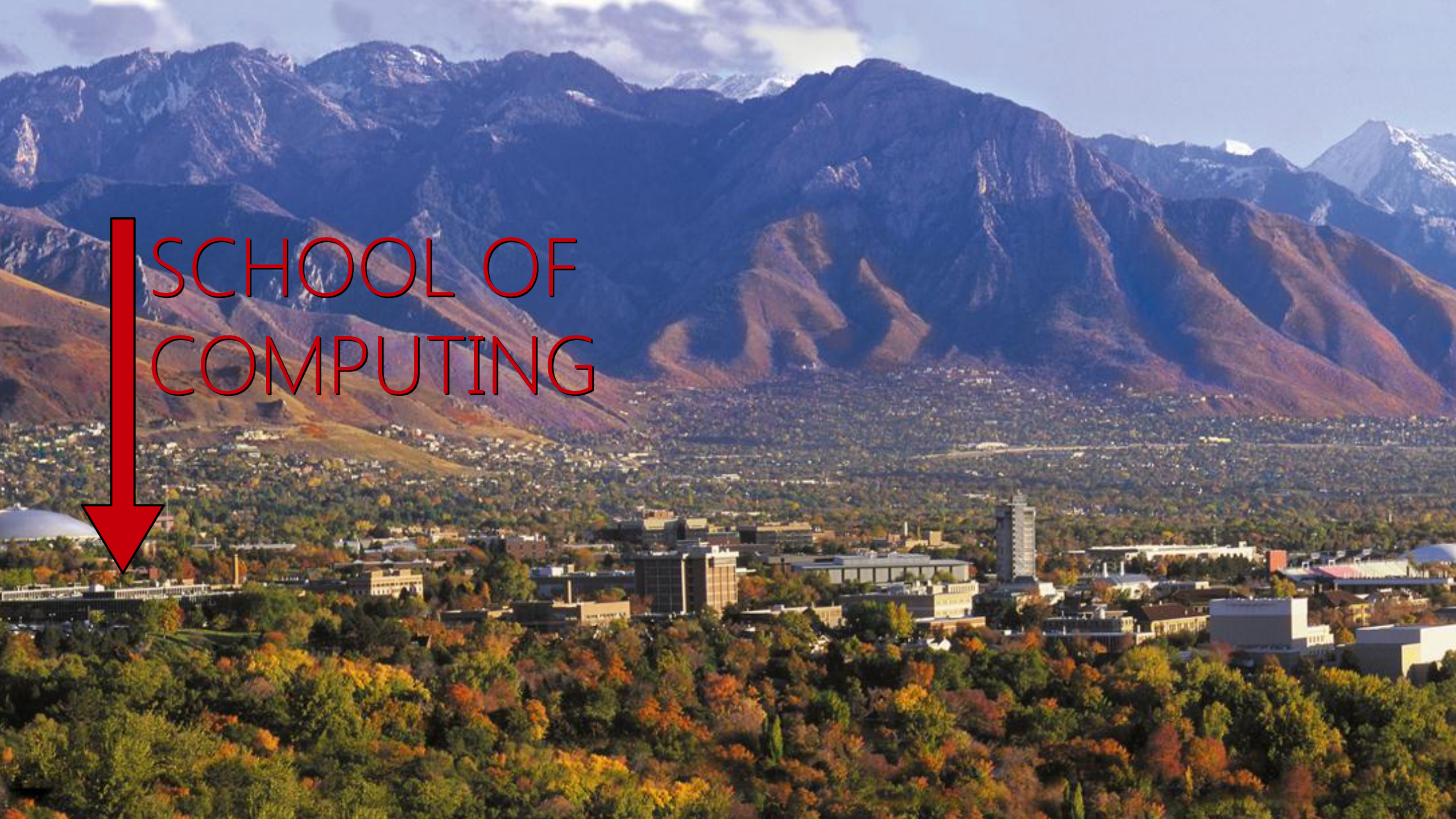
What it Means to be a Faculty Fellow

- Recognition is important to a faculty career
- Ties with industry help ground research
  - Especially database systems research
- Money provides freedom to take risks

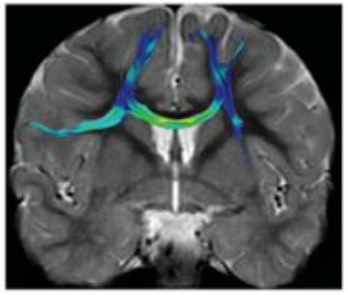
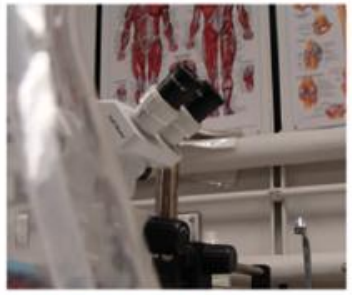
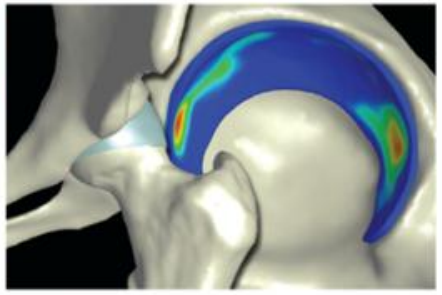
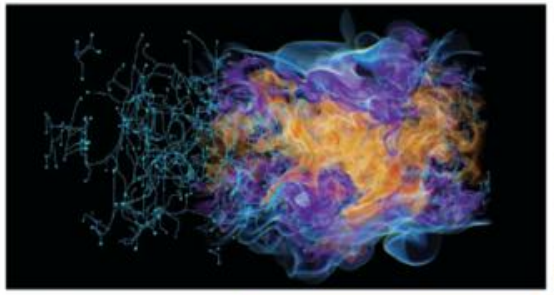
Miriah Meyer

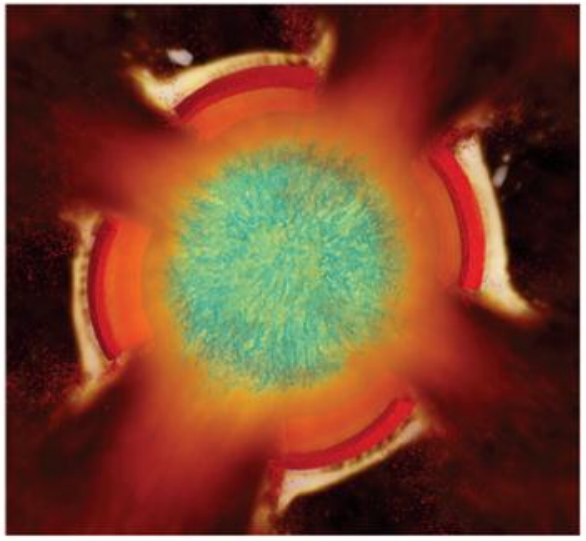
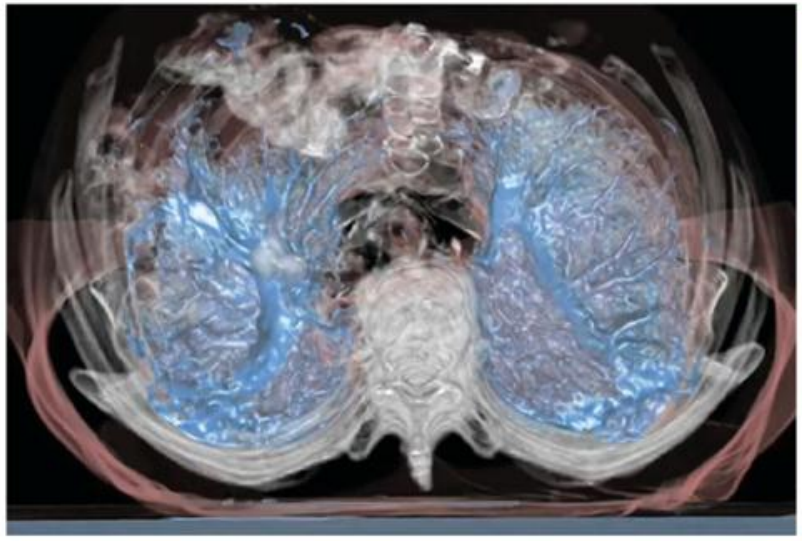
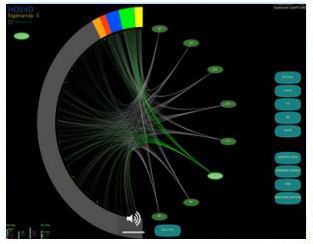
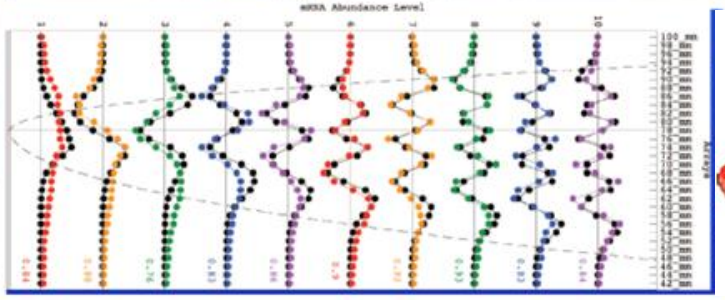
Assistant Professor

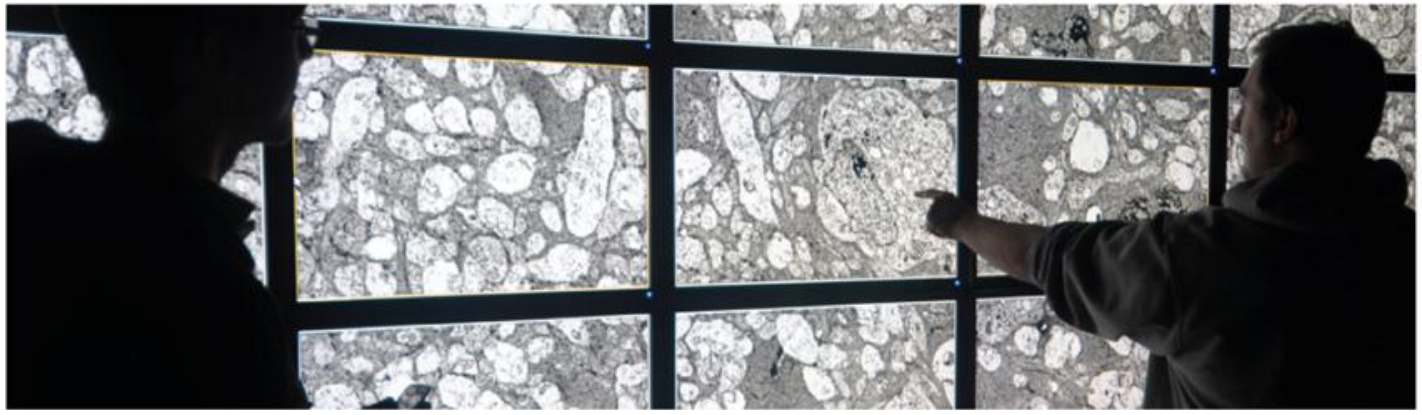
School of Computing
Scientific Computing and Imaging Institute
University of Utah

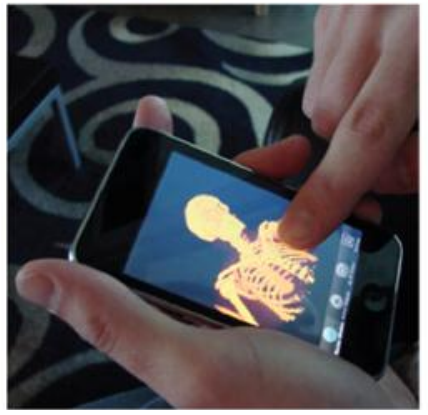

SCHOOL OF
COMPUTING

SCIENTIFIC COMUTING & IMAGING INSTITUTE



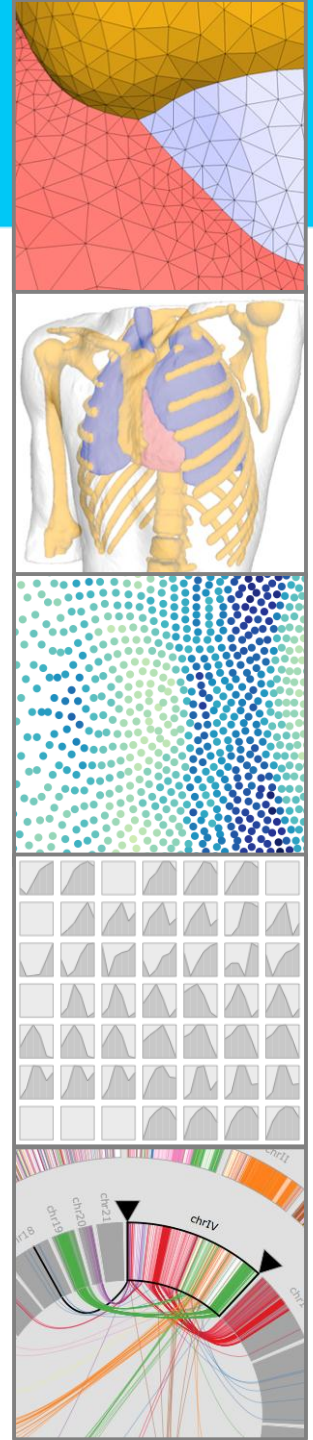
•Education

- BS in Astronomy, Penn State
- PhD in Computer Science, University of Utah
- Postdoc, Harvard University
- Visiting Scientist, Broad Institute of MIT and Harvard

•Research

- PhD: construction of geometry from volumes (medical images)

- Design of interactive visualization tools to explore complex data
- Visual representations and intuitive interactions
- Collaboration with scientists in early-stage analysis
- Methodology for visualization research

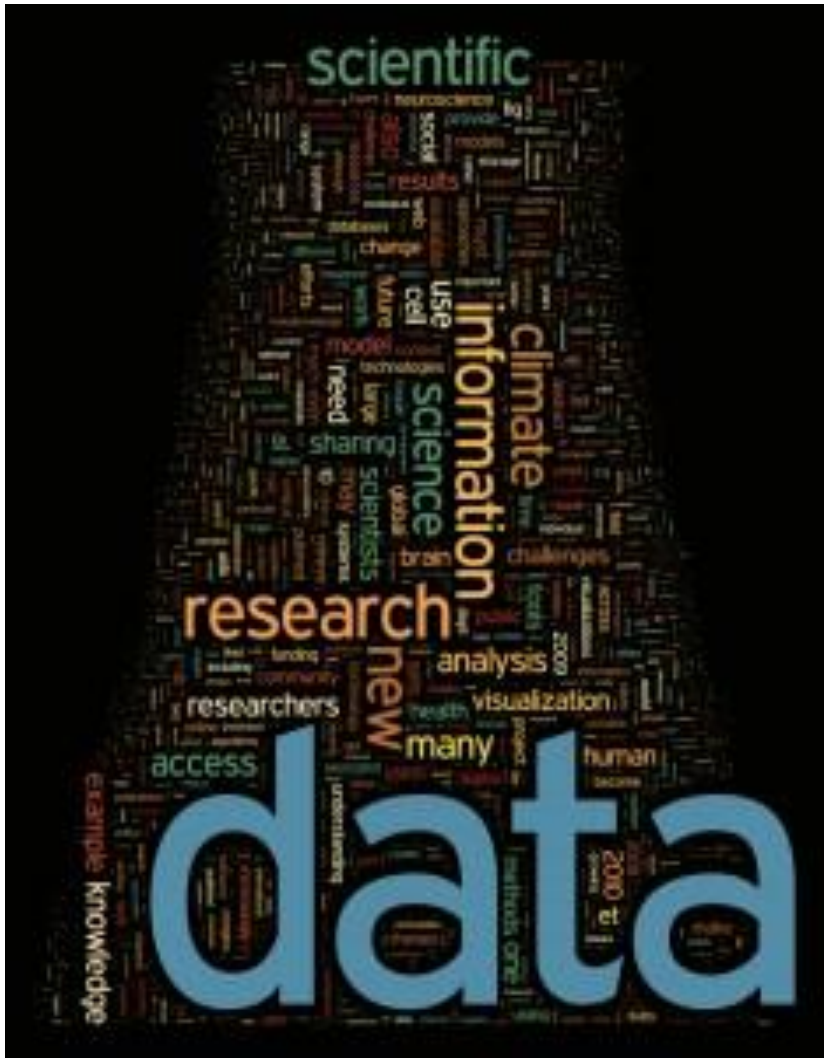


Big Data
Bigger Challenges

Wei Wang
University of California, Los Angeles

July 16 2012

Surfing the Data Flood



Data is Everywhere

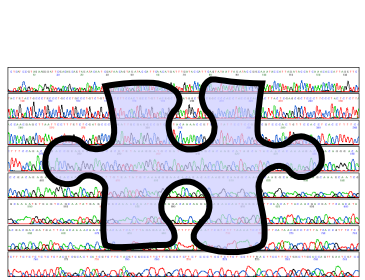
Data Mining is Everywhere

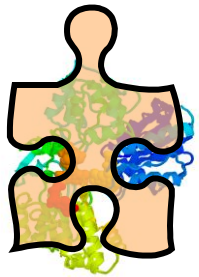
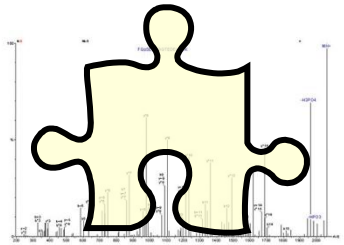
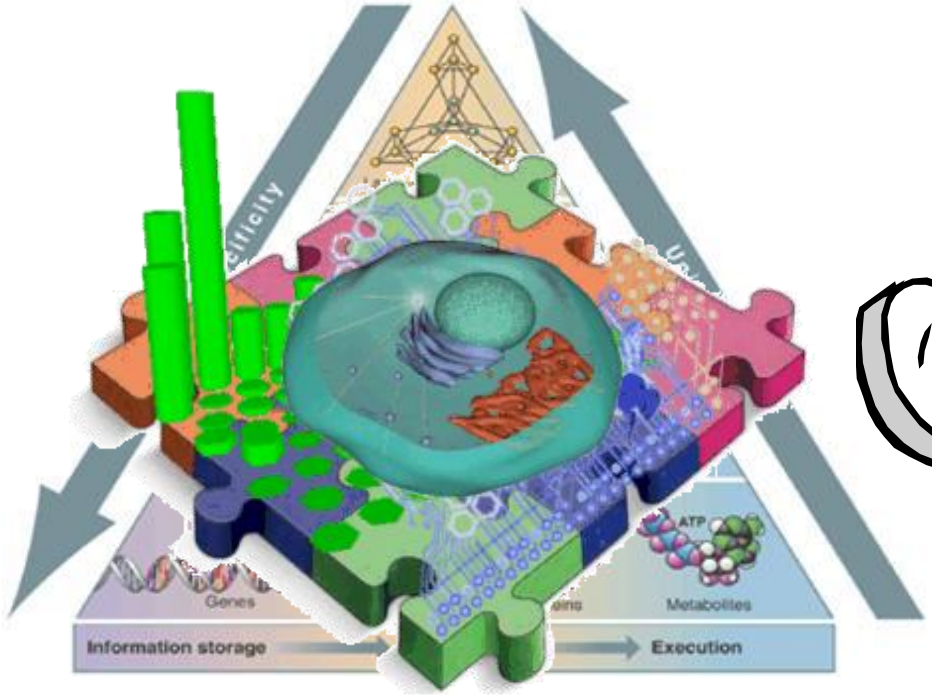
In the past 20 years, over 2 million research papers have been published!

Data Mining is Challenging

In the past 20 years, over 2 million research papers have been published! **BUT …**

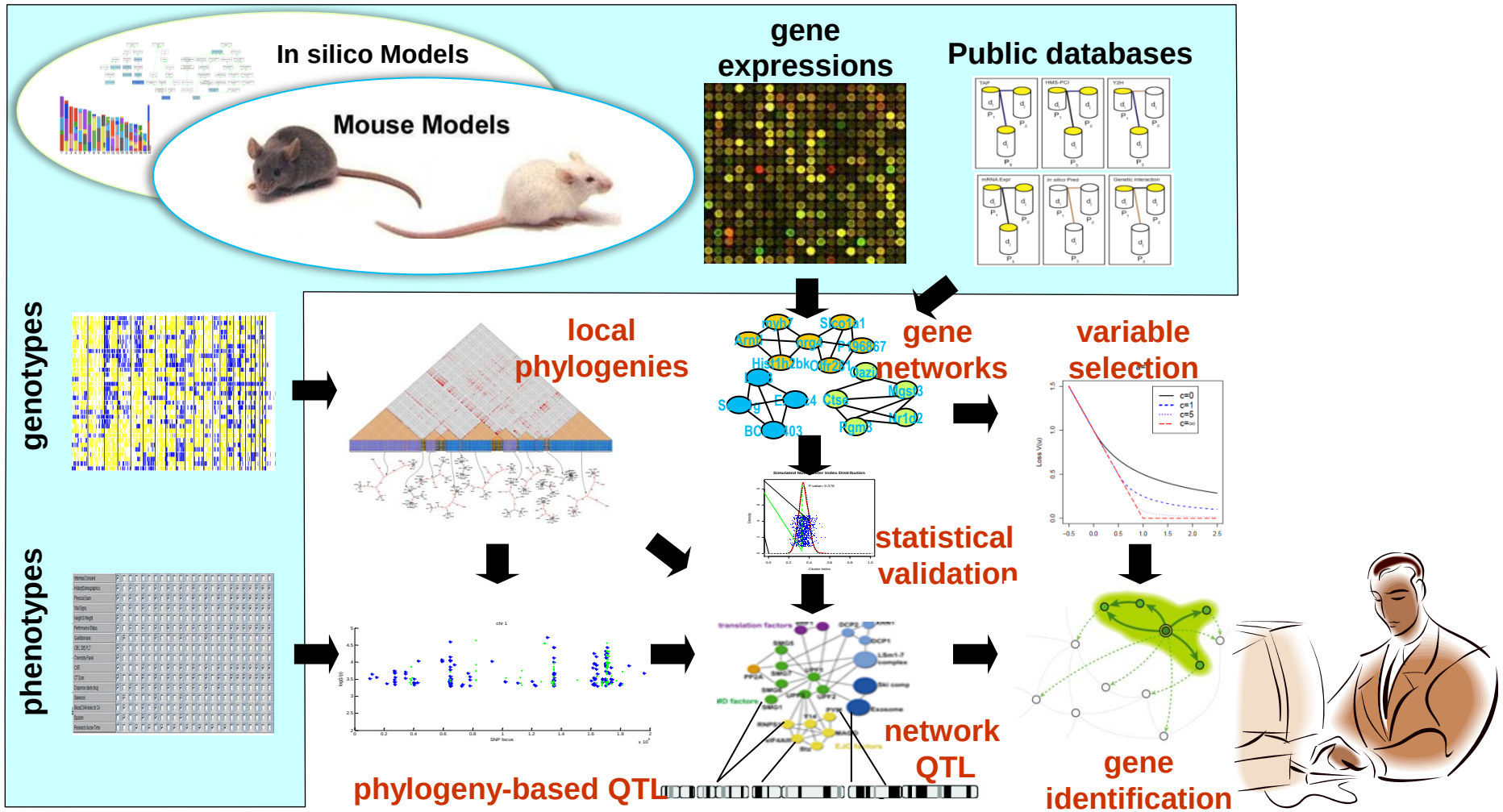
Data Mining in Science



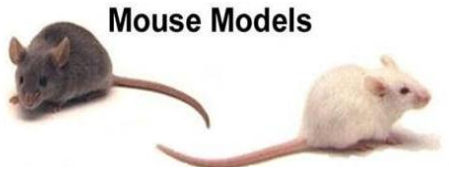




Life's Complexity Pyramid
*Z. N. Oltvai and A.-L. Barabási, Science **298**, 763 (2002)*

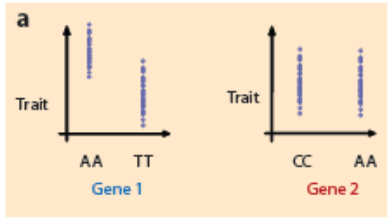
Integrative Computational Analysis for Systems Genetics

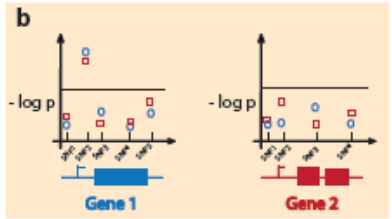


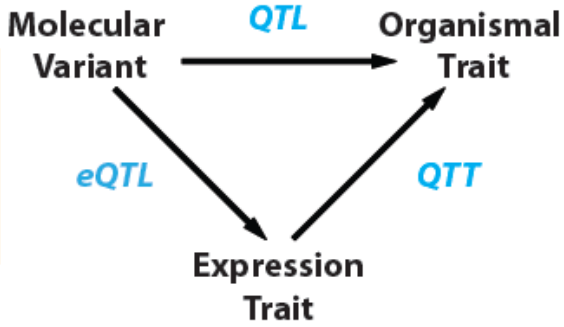
From Mouse to Human

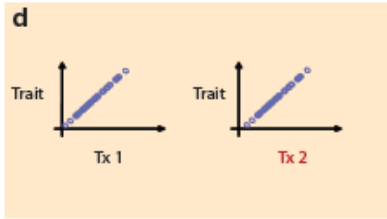


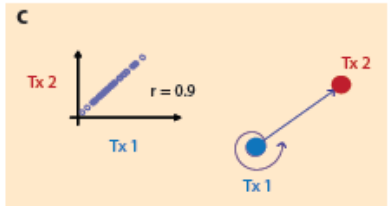
Depression disorder





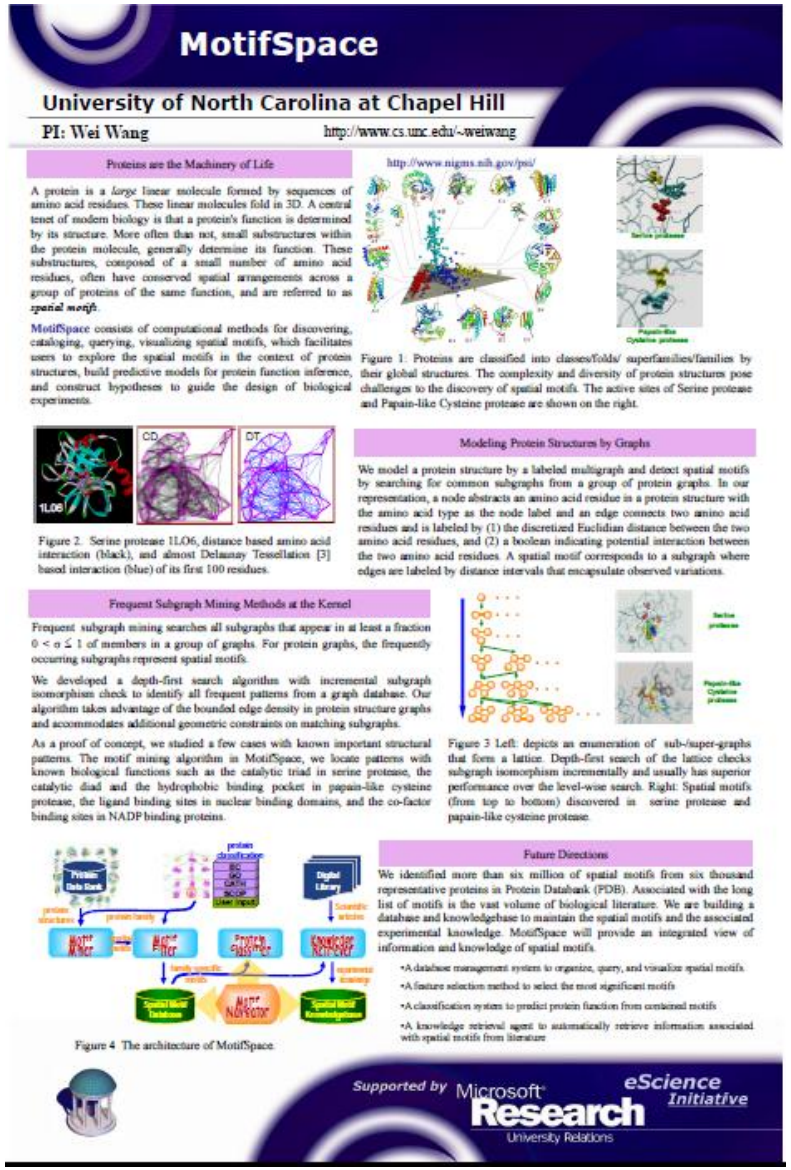






A step towards personalized medicine

Thanks to Microsoft



MotifSpace

University of North Carolina at Chapel Hill

PI: Wei Wang http://www.cs.unc.edu/~weiwang

Proteins are the Machinery of Life

A protein is a *large* linear molecule formed by sequences of amino acid residues. These linear molecules fold in 3D. A central tenet of modern biology is that a protein's function is determined by its structure. More often than not, small substructures within the protein molecule, generally determine its function. These substructures, composed of a small number of amino acid residues, often have conserved spatial arrangements across a group of proteins of the same function, and are referred to as ***spatial motifs***.

MotifSpace consists of computational methods for discovering, cataloging, querying, visualizing spatial motifs, which facilitates users to explore the spatial motifs in the context of protein structures, build predictive models for protein function inference, and construct hypotheses to guide the design of biological experiments.

http://www.nigms.nih.gov/psi/

Figure 1. Proteins are classified into classes/folds/ superfamilies/families by their global structures. The complexity and diversity of protein structures pose challenges to the discovery of spatial motifs. The active sites of Serine protease and Papain-like Cysteine protease are shown on the right.

Figure 2. Serine protease 1LO6, distance based amino acid interaction (black), and almost Delaunay Tessellation [3] based interaction (blue) of its first 100 residues.

Modeling Protein Structures by Graphs

We model a protein structure by a labeled multigraph and detect spatial motifs by searching for common subgraphs from a group of protein graphs. In our representation, a node abstracts an amino acid residue in a protein structure with the amino acid type as the node label and an edge connects two amino acid residues and is labeled by (1) the discretized Euclidian distance between the two amino acid residues, and (2) a boolean indicating potential interaction between the two amino acid residues. A spatial motif corresponds to a subgraph where edges are labeled by distance intervals that encapsulate observed variations.

Frequent Subgraph Mining Methods at the Kernel

Frequent subgraph mining searches all subgraphs that appear in at least a fraction $0 < \alpha \leq 1$ of members in a group of graphs. For protein graphs, the frequently occurring subgraphs represent spatial motifs.

We developed a depth-first search algorithm with incremental subgraph isomorphism check to identify all frequent patterns from a graph database. Our algorithm takes advantage of the bounded edge density in protein structure graphs and accommodates additional geometric constraints on matching subgraphs.

As a proof of concept, we studied a few cases with known important structural patterns. The motif mining algorithm in MotifSpace, we locate patterns with known biological functions such as the catalytic triad in serine protease, the catalytic diad and the hydrophobic binding pocket in papain-like cysteine protease, the ligand binding sites in nuclear binding domains, and the co-factor binding sites in NADP binding proteins.

Figure 3 Left: depicts an enumeration of sub-/super-graphs that form a lattice. Depth-first search of the lattice checks subgraph isomorphism incrementally and usually has superior performance over the level-wise search. Right: Spatial motifs (from top to bottom) discovered in serine protease and papain-like cysteine protease.

Figure 4 The architecture of MotifSpace.

Future Directions

We identified more than six million of spatial motifs from six thousand representative proteins in Protein Databank (PDB). Associated with the long list of motifs is the vast volume of biological literature. We are building a database and knowledgebase to maintain the spatial motifs and the associated experimental knowledge. MotifSpace will provide an integrated view of information and knowledge of spatial motifs.

- A database management system to organize, query, and visualize spatial motifs.
- A feature selection method to select the most significant motifs
- A classification system to predict protein function from contained motifs
- A knowledge retrieval agent to automatically retrieve information associated with spatial motifs from literature

Supported by Microsoft Research eScience Initiative University Relations

Panel Discussion

- **Moderator**: Rane Johnson-Stempson, MSR
- **Lucy Sanders**, CEO NCWIT
- **Emma Brunskill**, Carnegie Mellon
- **Magdalena Balazinska**, University of Washington
- **Miriah Meyer**, University if Utah
- **Wei Wang**, University of California, Los Angeles

CALL TO ACTION- Magdalena

What univ. need to do to grow nb of women faculty?

- Interview and hire more women faculty
- Acknowledge impact of child-rearing on women
- Acknowledge impact of pregnancy
- Help grow pipeline of women faculty & researchers

Universities are already doing a lot in this area!

So I believe it is a matter of continuing and doing more

CALL TO ACTION- Miriah

- Support multidisciplinary curricula
- Cross-cutting CS intro for non-majors
- Help women find each other

- Help dual-career couples
- Formal (and informal) faculty mentorship
- On-campus childcare

Call to Action-Wei

- Women in Computing
  - Fellowships for excellent female students to study computer science
  - Special hiring package that makes relocation easier
  - Selective (Less) committee service
  - Seed fund to enable new collaboration

Become a Change Leader

Join the NCWIT Academic Alliance

Work with over 200 peer institutions on effective practices and utilization of the latest research. Microsoft Research funds the NCWIT Seed Fund for practice implementation.

info@ncwit.org for more information.

ACADEMIC
NCWIT &
ALLIANCE

Opportunities with MSR

Further your research, have your students submit for a

GRADUATE SCHOLARSHIP AWARD

click here

Get involved in changing the world, have your students apply for an

INTERNSHIPS & FELLOWSHIPS

click here

Learn more about our activities by visiting our WEBSITE

click here


Thank you!
Q&A

Microsoft®