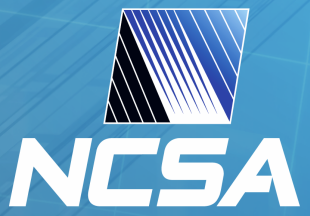
NCSA Faculty Fellows 2018
Idea Acceleration Workshop

Amanda Lombardo
Assistant Director for Research, NCSA
alombar@illinois.edu

Farzaneh Masoud
Interdisciplinary Research Development Officer, NCSA
fmasoud2@illinois.edu



National Center for Supercomputing Applications
University of Illinois at Urbana–Champaign

NCSA Fellow Program

- An opportunity for faculty and researchers at the UIUC to catalyze and develop long-term research collaborations between Illinois departments, research units, and the NCSA.

- Proposal Preparation and Submission: http://www.ncsa.illinois.edu/about/org/fellowships/submissions

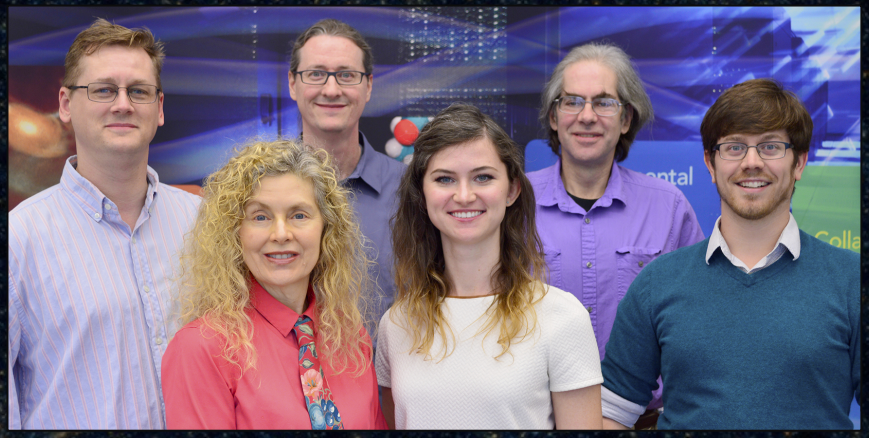
- Key Dates:
  - **Feb. 18, 2017:** Deadline to submit to the NCSA Fellowships Program
  - **April 2017:** Target date for decisions

NCSA Programs

Faculty Fellows Idea Acceleration Workshop

January 26, 2018

Advanced Visualization Lab



Jeff Carpenter, Bob Patterson, Stuart Levy, Donna Cox, Kalina Borkiewicz, AJ Christensen
avl@ncsa.illinois.edu

Data/Scientific Visualization

- Visualizations of computational and observational data
  - Astrophysics, geoscience, molecular biology, sustainable agriculture, geospatial systems, etc.
- Interactive visualization
- Data wrangling

Design, Art and Humanities

- Cinematic-quality data visualization
- Virtual Reality
- Visualization for live performances
- Mobile apps
- **Multimedia and video production**

VI-Bio *Visual Intelligence for Biology*

Colleen Bushell
visari.org

Information Designers

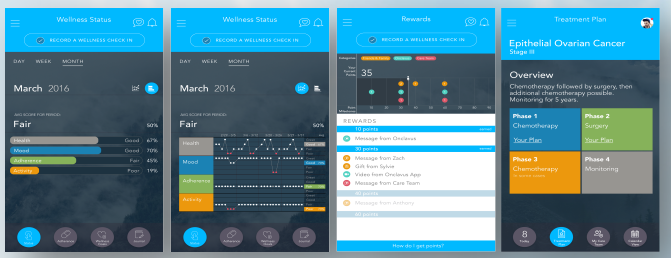
Clarifying complex results for all levels of knowledge

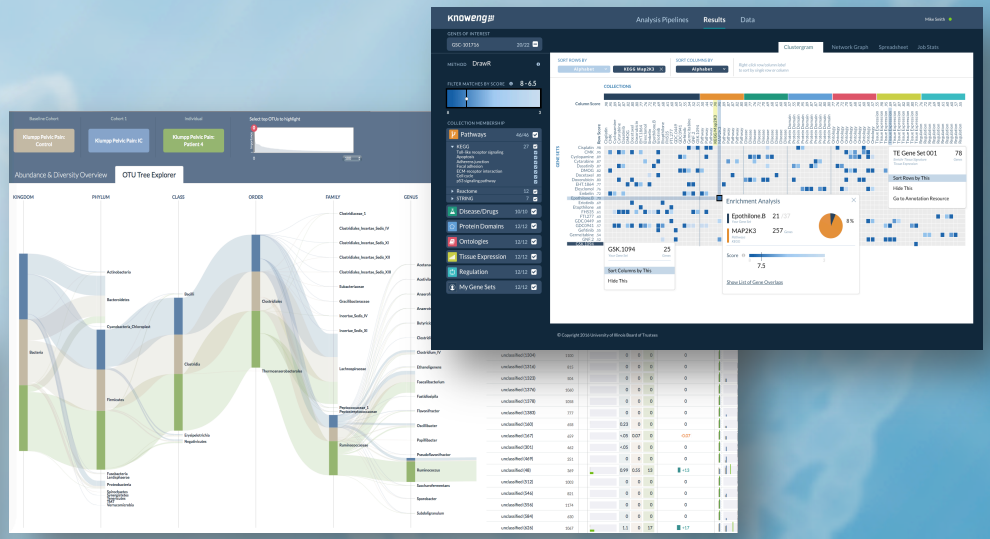
Mathematicians & Data Scientists

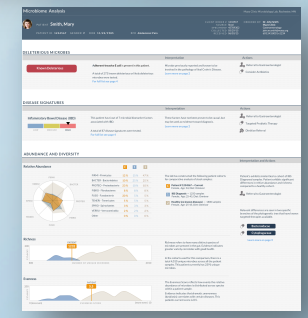
Developing & applying machine/deep learning approaches

Software Engineers

Creating software for executing analysis, studying results

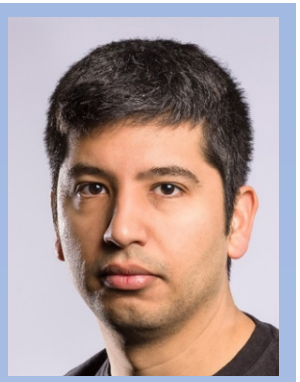






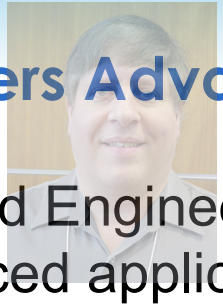
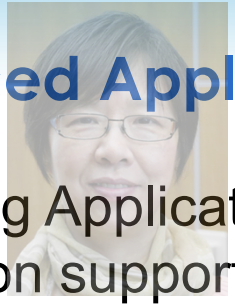
Scientific Software and Applications (SSA) Division

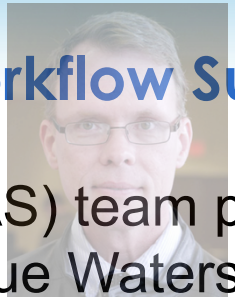
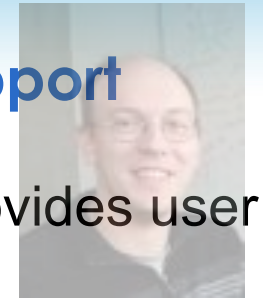
Dan Katz & Kenton McHenry
(dskatz@illinois.edu & mchenry@uiuc.edu)



- SSA Groups
  - Advanced Application Support (AAS) – Jay Alameda
  - Innovative Software & Data Analysis (ISDA) – Kenton McHenry & Jong Lee
  - Scientific and Engineering Applications Support (SEAS) – Greg Bauer
- SSA Expertise
  - Software Development
    - Build scientific software to satisfy current needs & create new opportunities, such as frameworks, science gateways, data access, data storage, machine learning, workflow systems
    - Use of external software development technologies
    - Expertise in software engineering techniques
  - Applications Support
    - Provide consulting & advanced application support for Blue Waters, campus cluster, industry partners, e.g. porting, tuning, benchmarking, debugging, workflows, frameworks, performance tools integration, code refactoring, enabling new programming models, new algorithms
    - Areas: computational chemistry, fluid dynamics (DNS, FEM), computational astrophysics, numerical methods, bioinformatics, multiphysics, molecular dynamics, manufacturing, nuclear engineering, statistics, data analysis and analytics, code profiling and optimization, sparse linear solvers, runtime systems
  - Dan's individual expertise
    - Workflows – programming a series of runs of programs or functions, and running this efficiently on parallel and distributed (cloud) computing
    - Software as infrastructure – making software more sustainable, giving researchers credit for the software they develop, software funding models

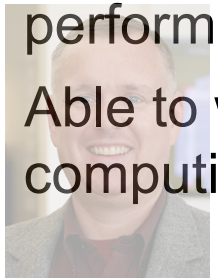
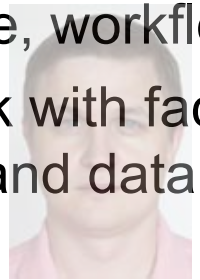
NCSA

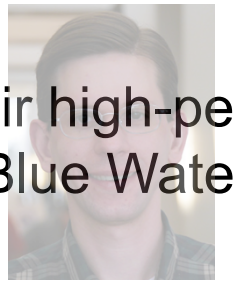




Blue Waters Advanced Application and Workflow Support

- Science and Engineering Application Support (SEAS) team provides user and advanced application support to projects on Blue Waters.
- Range of domain and technical expertise: chemistry, fluid dynamics, runtime systems, numerical methods, astrophysics, application performance, workflows.
- Able to work with faculty interested in enhancing their high-performance computing and data workflows and applications on Blue Waters.

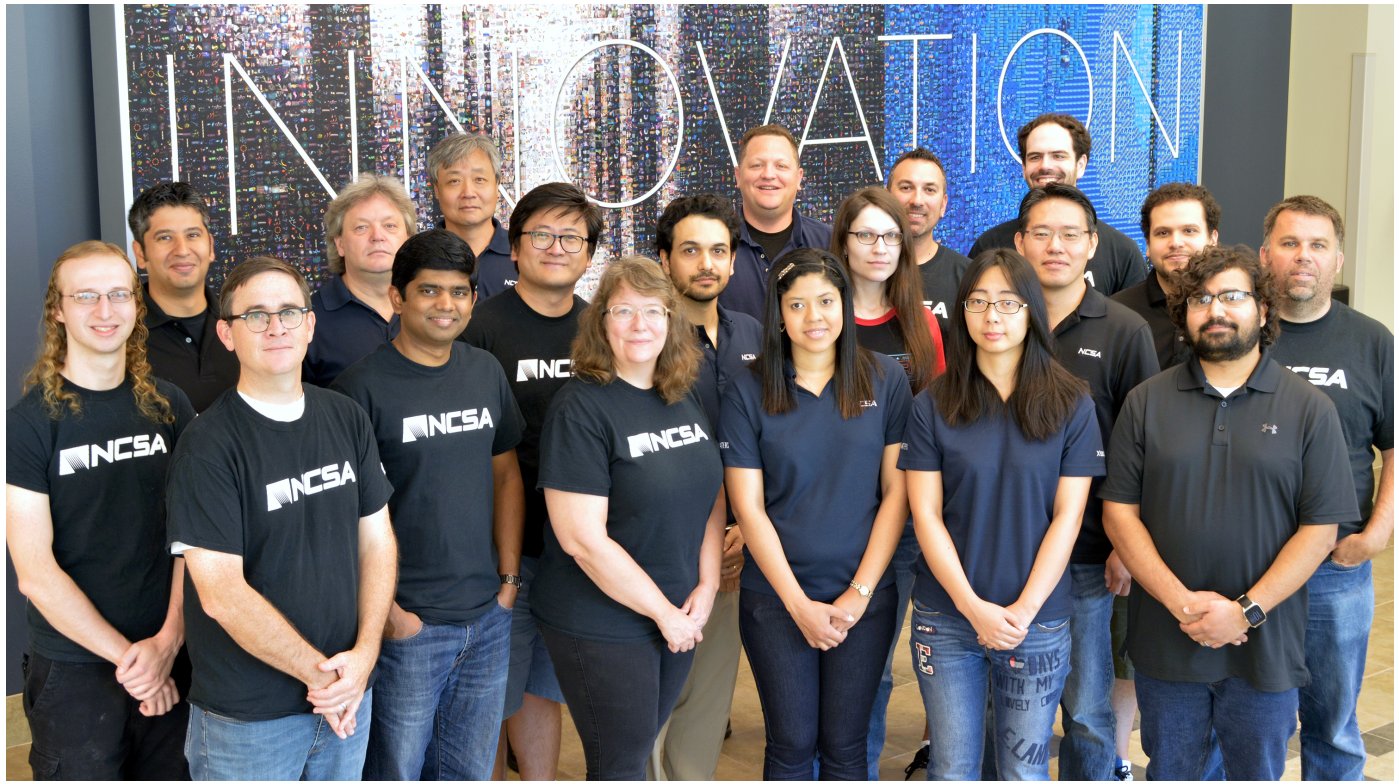



- Gregory Bauer – SEAS lead

NCSA

Innovative Software & Data Analysis

- Research & Development
- Reusable software tools & frameworks for data analysis
- Bridging and amplifying efforts across different projects
- New custom software tools & frameworks



NCSA Industry

Ahmed Taha, Technical Program Manger

Domain Expertise

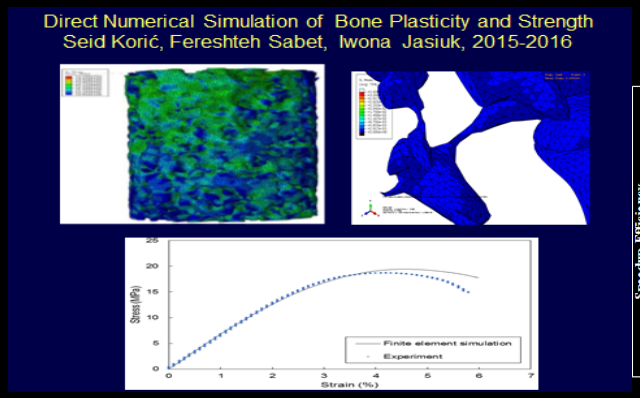
- Multi-physics/scale Modeling in Thermo-Fluids, Manufacturing and Material Processing
- Biomechanics
- Nuclear Engineering
- Combustion
- Molecular Dynamics
- Bioinformatics
- Data Analytics, Big Data, …

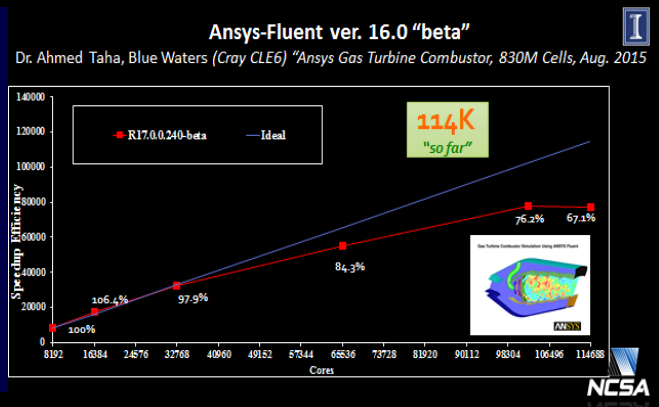
HPC Expertise

- Multiphysics Software (ISV and Open Source)
- Extreme Scalability
- Sparse Solvers
- Code Profiling , Parallelization and Optimization
- Performance Tools & Libraries

Proposals, Projects and Beyond

- Research Collaborations (>40) & Publications (>20 in 2017)
- NCSA Faculty Fellowships
- Industry (Sponsored Projects)
- Institutions (NSF/NIH/NIST)
- Mixed .gov/.com (NDEMC, DMDII, AFFOA)
- Project Management
- Business Development



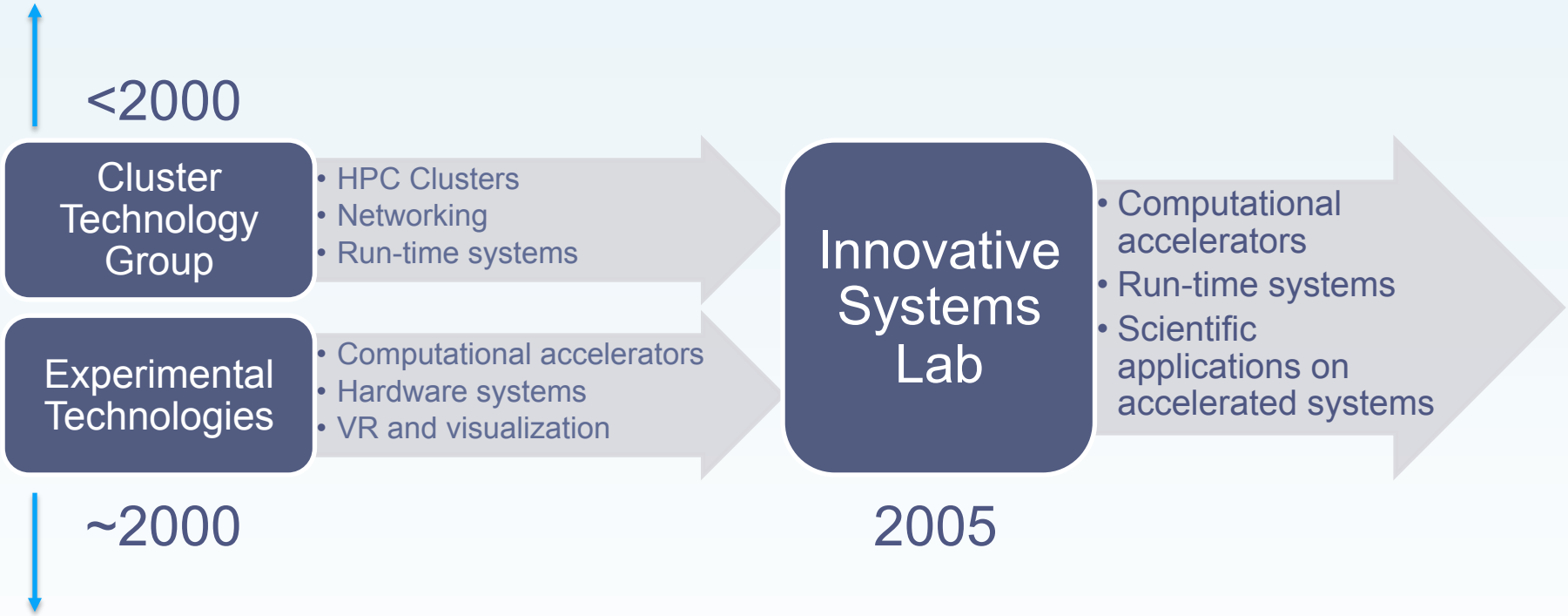


ISL Upcoming Activities

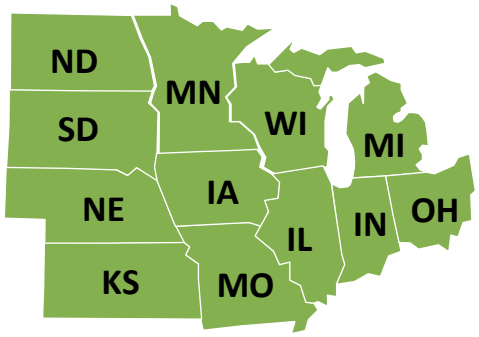
Contact: Volodymyr (Vlad) Kindratenko, D.Sc.
kindrtnk@illinois.edu
Senior Research Scientist, NCSA
Adjunct Associate Professor, ECE

ISL Activities

- **Notable achievements:** Linux and Windows HPC clusters, PlayStation 2 cluster
- **People: Mike Showerman, Jeremy Enos, Craig Steffen**, Guochun Shi



- **Notable achievements:** Distributed VR, IntelliBadge, Software-Defined Radio, Smart Spaces
- **People:** David Pointer, **Volodymyr Kindratenko**



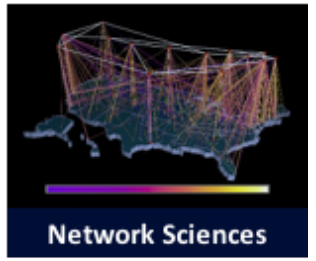
MIDWEST BIG DATA INNOVATION HUB

MIDWEST BIG DATA HUB

Bill Gropp, PI
Director, NCSA

Melissa Cragin
Executive Director

Co-PIs:
Sarah Nusser, VPR @ ISU
Grant McGimpsey, VPR @ UND
Brian Athey, MIDAS co-Chair @ Umich
Inna Kouper, D2I @ IN Univ.



- Grow the national data ecosystem
- Develop effective partnerships across academic-industry-government-NGOs
- Build regional capacity in Data Science and Big Data

http://midwestbigdatahub.org/ Twitter: @MWBigDataHub Email: info@midwestbigdatahub.org

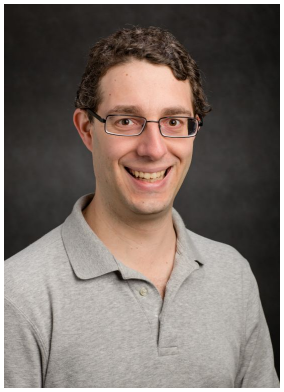
Faculty Applicants

Faculty Fellows Idea Acceleration Workshop

January 26, 2018

Rapid Generation of Phase Diagrams

Bryan Clark
Assistant Professor of Physics
bkclark@illinois.edu
http://clark.physics.illinois.edu/



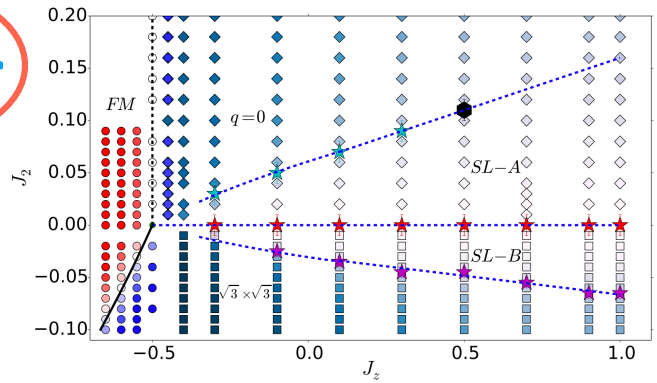
Condensed Matter Physics:

material

$$H[J_z, J_2] = H^{\text{nn}}_{XXZ}[J_z] + J_2 H^{\text{nnn}}_{XXZ}[J_z]$$

$$H_{XXZ}[J_z] = \sum_{\langle i,j \rangle} S^x_i S^x_j + S^y_i S^y_j + J_z \sum_{\langle i,j \rangle} S^z_i S^z_j$$

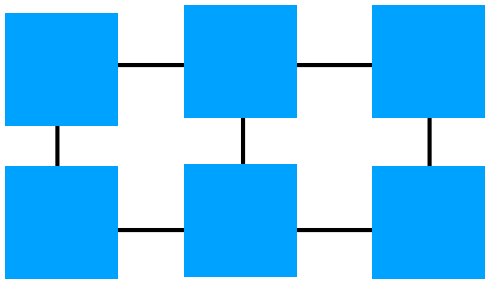
model



Phase Diagram

QMC: Medium accuracy but massively parallel

Tensor Networks: Highly accurate; extremely popular; **3 week runs on a single node.**



Primary Step: Contraction of Tensor Networks

Project/Goal: Massively Parallel Tensor Networks

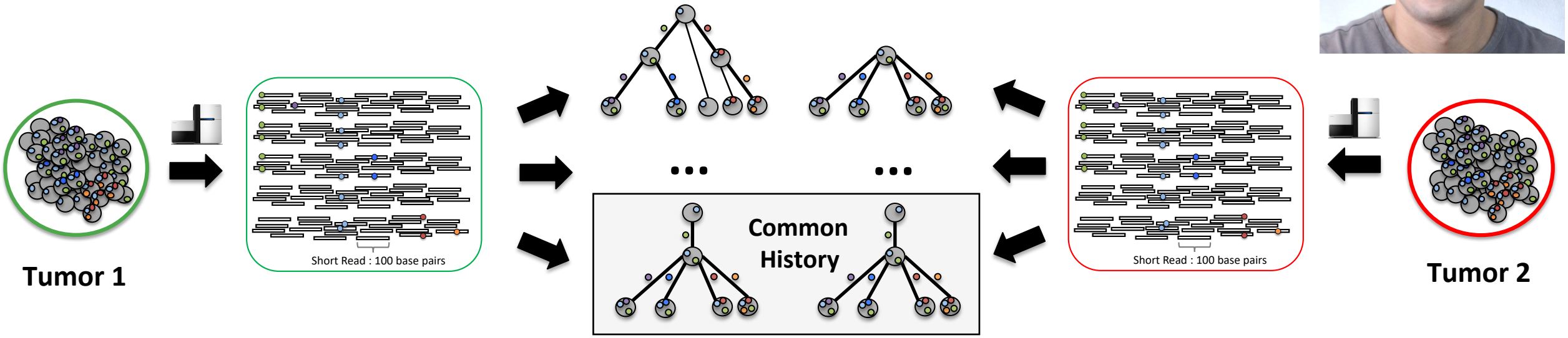
Needs: Professional assistance parallelizing codes

Key Steps: Decisions about approach to parallelization (i.e. how to break up tensors, store information, etc)

Results: 3 weeks → 1 day completely transforms the type of problems which can be solved.

Large-scale Tumor Phylogeny Estimation

Mohammed El-Kebir -- Assistant Professor of Computer Science
melkebir@illinois.edu -- www.el-kebir.net



Problem:

- **Heterogeneity**: not only is each tumor different, but so is each cell within a tumor
- DNA sequencing yields **average measurements**
  => identifiability issue

Key idea:

- **Joint tumor phylogeny estimation** in a large cohort of samples by maximizing common patterns of evolution

Needs:

- Expertise in performance tuning, optimization & parallelization
- NCSA computing resources are vital for analyzing thousands of tumor genomes

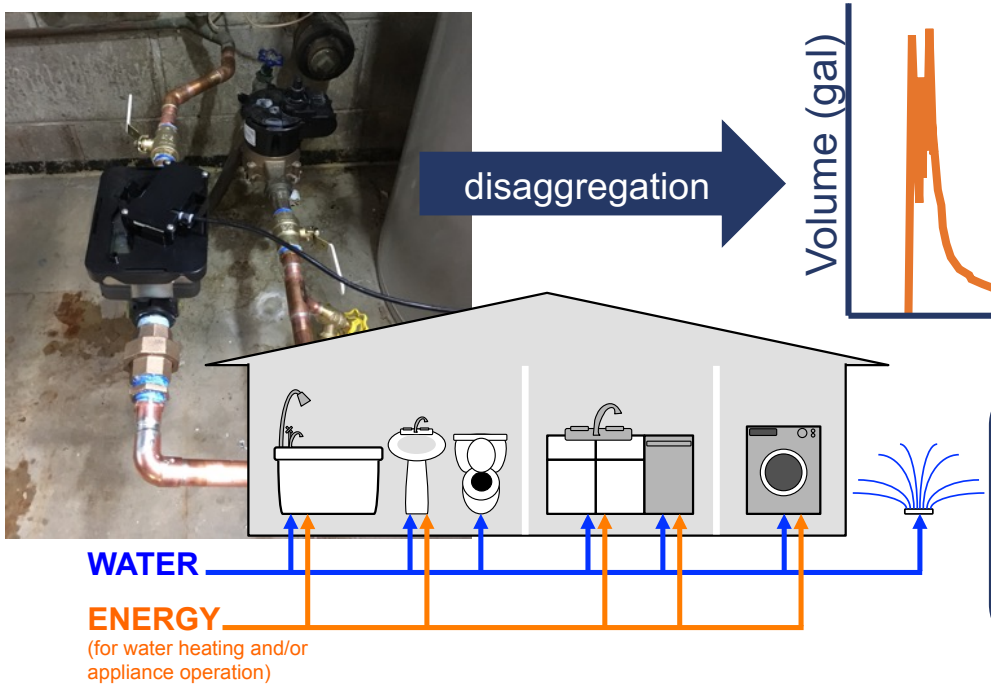
End Use Disaggregation from Residential Smart Water Metering

Ashlynn S. Stillwell

Assistant Professor, Civil and Environmental Engineering

ashlynn@illinois.edu | stillwell.cee.illinois.edu

NCSA Research Focus Area: Food, Energy, and Water



Problem:

- Residential water use is poorly quantified at specific end uses
- Understanding behavior requires accurate disaggregated water flows

Key idea and initial resources:

- Fine-resolution, high-frequency smart water meter data can measure residential end uses
- Smart water meter installed on test home with training dataset of independent end uses

Collaboration needs:

- Expertise in machine learning to disaggregate data from concurrent water uses
- Parsing, visualization, storage of big water data

A Heterogeneous Data Analytics Framework for Predicting the Deterioration of

Nora El-Gohary
Associate Professor, Civil and Environmental Engineering
gohary@illinois.edu • https://cee.illinois.edu/directory/profile/gohary

Problem

- Need for data analytics to better understand infrastructure deterioration & long-term maintenance effectiveness
- The heterogeneity of data & the needed complex analysis makes it impossible to use any existing data analytics tools

Key idea

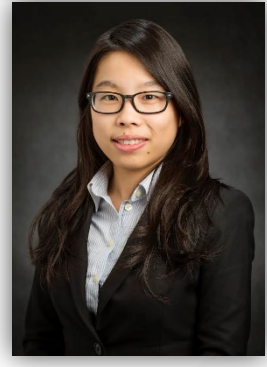
- Novel computational framework for learning from multi-source heterogeneous data

Needs

- Implementation and optimization of our machine learning platform

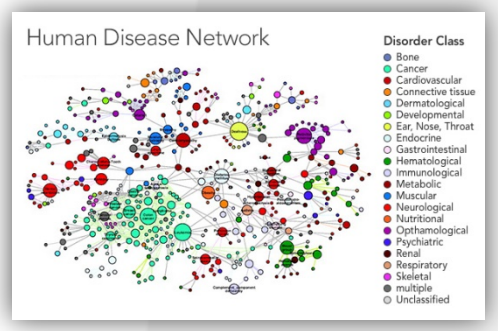
Learning Huge-Scale Diffusion Networks in Real-Time

Niao He
Assistant Professor, Industrial and Enterprise Systems Engineering (ISE)
Affiliated Assistant Professor, Coordinated Science Laboratory (CSL)
niaohe@illinoi.s.edu
http://niaohe.ise.illinois.edu/



Social Networks

Disease Networks



Cyber Crime Networks

Online and Deep Learning of Diffusion Networks

- Huge volumes of event data
- Huge amount of nodes
- Complex temporal dynamics
- Complex causality & reciprocity
- Latent network structure

Stochastic Modeling and Simulation → **Data Analytics & Management**

- Using mutual-exciting point processes to model dynamics
- Representing intensity functions with deep neural nets

Scalable Algorithms for Inference & Prediction → **Software Development**

- Applying efficient online gradient descent algorithm
- Reducing computation overhead via randomization

Asynchronous Parallel Optimization → **CPU & GPU Clusters and Infrastructure**

- Developing asynchronous parallel algorithms to speed up the learning procedure

Applications | Project Mission | Methodology | NCSA Partnership

Data-Driven Virtual Career Mentoring to Recruit and Retain URM Talents in STEM Undergraduate Programs

- **Goal:** To recruit and retain URM talents in STEM undergraduate programs more effectively
- **Solution:** Data-driven mentoring to augment Human-driven mentoring versus
  - Access to multiple mentors, just-in-time
  - Quality of data – theory grounded
  - Sources of data – human, internet
- **Need:** Prototype design and testing (interface, modality testing, data structure)
- **Contributing** to NCSA mission in expanding and retaining URM college student participation in STEM research and careers (NSF INCLUSION, XSEDE)

- Contact:

  W. David Huang
  wdhuang@illinois.edu

Associate Professor of Human Resource Development

Dept. Education Policy, Organization and Leadership

https://education.illinois.edu/faculty/wdhuang

Emotional prosody in normal vs. disordered speech: a machine learning, corpus approach

Heejin Kim
Research Assistant Professor of Linguistics; hkim17@illinois.edu

Aims

- To identify acoustic parameters that humans use to convey emotion in speech
- To discover how speakers with motor speech disorder (e.g., Parkinson's Disease, Cerebral Palsy) code their emotional status in speech
- To build computational framework that is transferrable to real-life applications

Significance

- Both theoretical and clinical significances: empirical evidence for designing treatment plans for patients
- Foundation for various applications in next-phase research: e.g., HCI that utilizes emotion detection and synthesis in health, education, & industry

Limits in previous work

- Unnatural speech with predefined parameters: leads to compromised accuracy and generalization ability for natural conversational speech

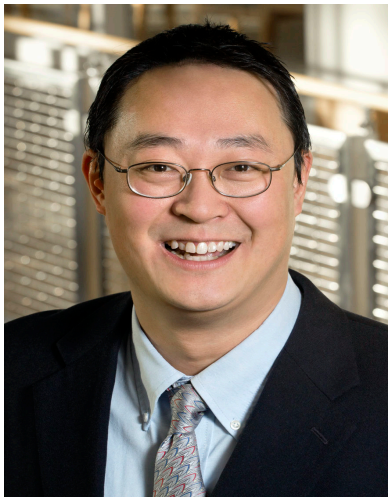
Key approaches

- Data: utilization of diverse speech corpora that is large and natural vs. actors reading sentences
- Semi-supervised machine learning to automatically extract optimal features (invariance in variations)
- Interdisciplinary efforts: Linguistics, ECE/CS, Education, Speech and Hearing Science

Needs

- Computing environment for large-scale audio signal processing and machine learning to develop acoustic parameter extraction, pattern recognition, & emotion classification algorithms

Jumpstart Big Data Research in Finance



Mao Ye
Assistant Professor of Finance, UIUC
National Bureau of Economic Research (NBER) Faculty Research Fellow
Email: maoye@illinois.edu
Personal website: https://yemaofin.com

"If you said that an economist was data mining, that was an insult."
- Susan Athey

Project idea: joint workshops with National Bureau of Economic Research (NBER) to jumpstart big data research in finance

- Stage 1: call for proposals for faculty members and students across the world
  - Finance: evaluated by Toni Whited (JFE editor) and I
  - Supercomputing plan: evaluated by XSEDE
- Stage 2: completed papers

Project needs:

- Staff members from NCSA to give orientations on high-performance computing for the first workshop. (Tentative time: July 14, 2018)
- Basic allocations and staff support for successful proposals

Interactive Map/History of Indian Cinema

Rini B Mehta
Assistant Professor
Comparative & World Literature
rbhttchr@Illinois.edu

PROJECT DESCRIPTION

Gateway to the history and evolution of Indian cinema, the home of Bollywood.

A comprehensive and interactive multimedia archive of material related to Indian cinema, its reception, its interactions with politics, and its global reach.

NCSA Thematic Area: Culture and Society

PLAN & EXECUTION

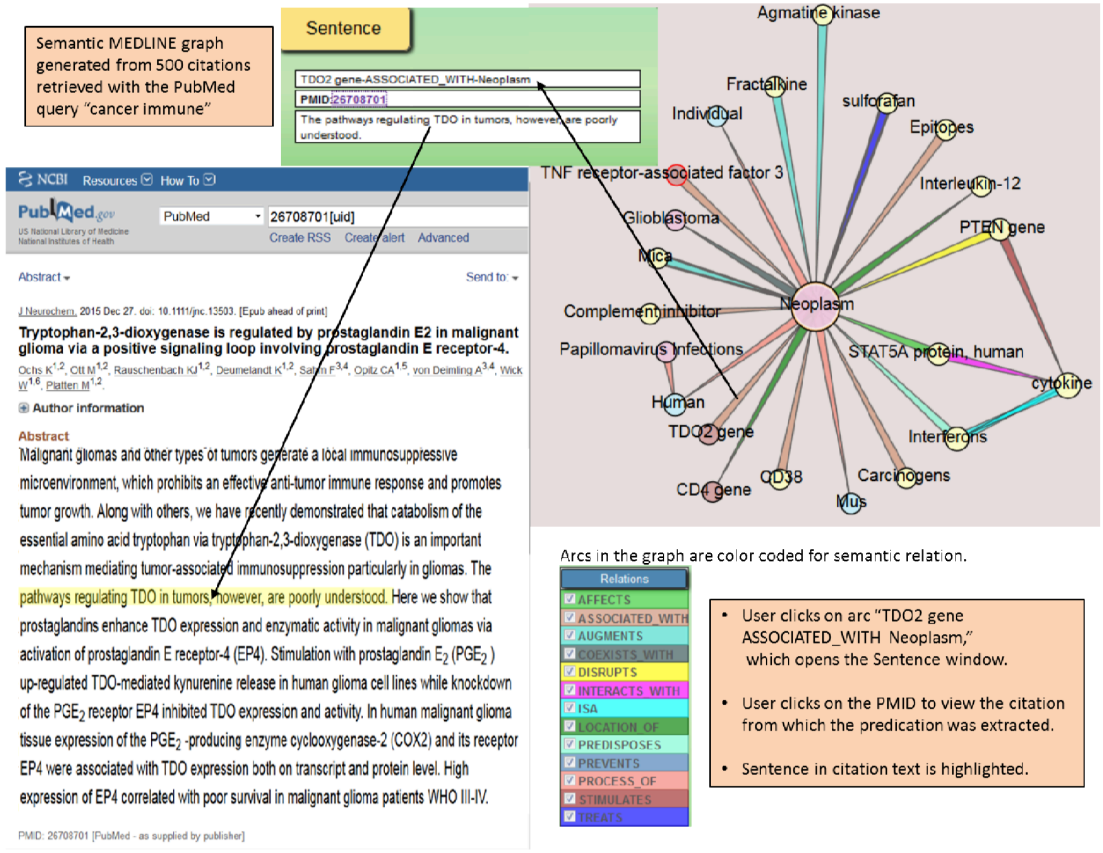
- Develop an interactive web interface for data visualization.
- Collect and curate data related to Indian cinema.
- Link the interface to pre-existing stable databases.
- Create back-end system for managing data flow.
- Explore use of crowdsourcing in gathering more data and annotating videos.

OTHER POSSIBLE FUNDING SOURCES

National Endowment of Humanities, National Science Foundation, Andrew W. Mellon Foundation, U.S. Department of Education, & Google.

Using Graph Databases to Visualize Connections in Biomedical Literature

Problem: Information Overload; Silos; Visualization.
Key Idea: Develop scalable, intuitive interfaces.
Needs: Expertise in graph databases, visualization.



Assistant Professor

jodi@illinois.edu

School of
Information Sciences
The iSchool at Illinois

Diwakar Shukla

Blue Waters Assistant Professor, Chemical & Biomolecular Engineering
Affiliate faculty, Center for Biophysics & Quantitative Biology
Affiliate faculty, Plant Biology
Affiliate faculty, Computational Science & Engineering

203 Roger Adams Laboratory, 600 S. Mathews Avenue,
University of Illinois at Urbana-Champaign, Urbana, IL 61801
http://www.shuklagroup.org

Design of Optimal Probes of Protein Dynamics

NCSA Thematic Areas:

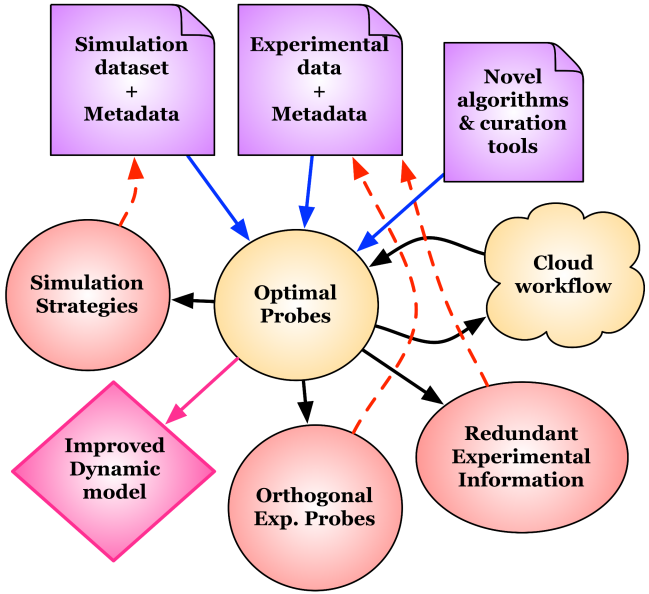
Bioinformatics & Health Sciences
Computing and Data Sciences

Problem

Challenges: (a) **Conflicting and redundant information** reported in literature on protein dynamics. (b) **No standard for reporting** experimental and computational information. (c) **Missing framework for integ**rating different types of computational and/or experimental information. (d) **Design of experiments** is done in an **uninformative** manner.

Approach

Machine learning approaches for (a) **integrating** different types of **experimental & computational datasets** on protein dynamics, (b) **designing** optimal set of **future experiments** that provide maximum information gain and (c) **building** a cloud-based interface for **"living models"** of protein dynamics.



Framework of the cloud platform for deign of optimal probes to study protein dynamics.

Needs

(a) **Data management and analysis tools** to support integration and exploratory analysis of research datasets.
(b) **Web-based access** to datasets, analysis tools and integrated models of protein dynamics.
(c) **Implementation** of the current "Optimal Probes" algorithm with efficient GPU/CPU parallelization.

Promise

The execution of the **"Optimal Probes"** platform will enable researchers to **examine, transform, and seamlessly integrate** disparate research data, **creating "living models"** that will enable new discovery by allowing researchers to **design informed studies of protein dynamics**.

For more information: http://github.com/ShuklaGroup/optimalProbes/wiki; Watch the demo on Youtube.

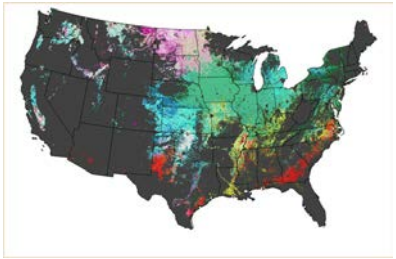
Nicolas F. Martin
Assistant Professor
Quantitative Agronomy
Crop Sciences

NFMARTIN@ILLINOIS.EDU
PROFILE

Enabling digital agriculture with spatial analytics

- Problem: Farmers, agribusiness and researchers aren't making the best use available technology.

- Key Idea: Agronomic research by means of geospatial data capture, integration and analysis.



- Needs: Basic cyber-infrastructure, new algorisms, algorism performance assessment, training a new generation researchers.

ILLINOIS
Crop Sciences
COLLEGE OF AGRICULTURAL, CONSUMER
& ENVIRONMENTAL SCIENCES