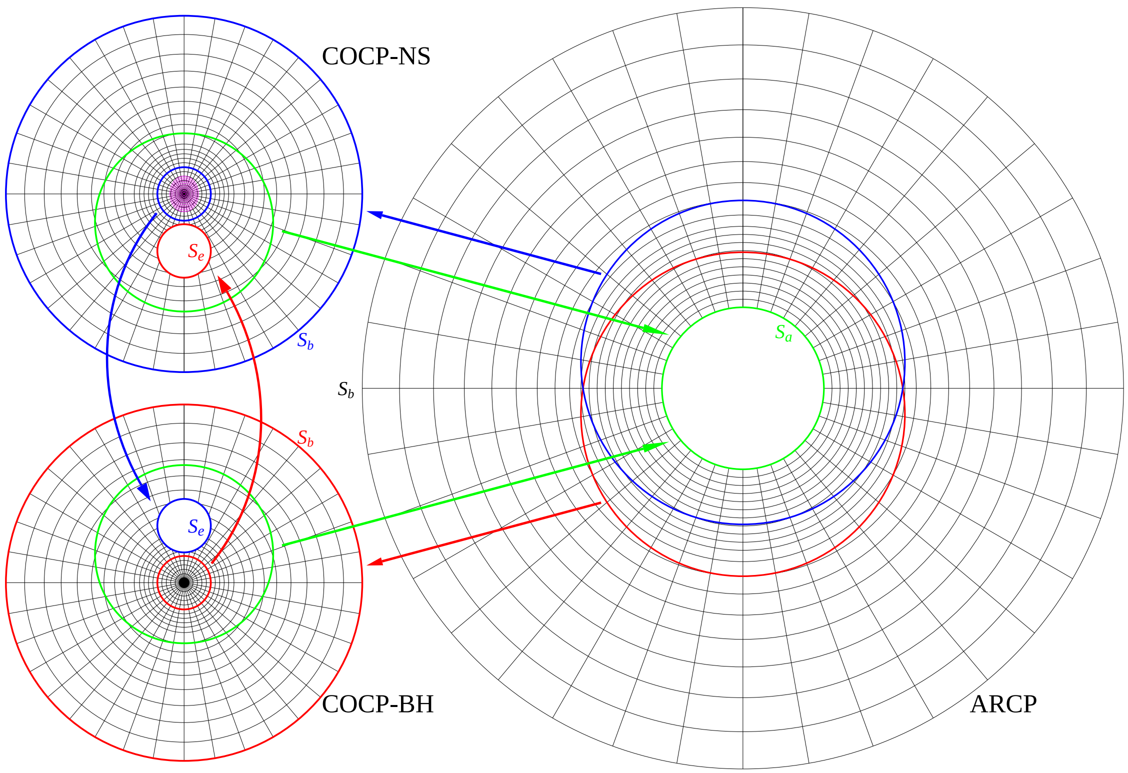
Applicant Slides

Faculty Fellows Idea Acceleration Workshop

January 27, 2017

Numerical Astrophysics: Parallelizing the Compact Object CALculator (COCAL) code.

Antonios Tsokaros, Research Scientist, Physics Department.
tsokaros@illinois.edu



- **About the COCAL code.**

⇒ It is an initial data solver that is designed to compute *any kind* of equilibrium or quasi-equilibrium solution in general relativity.

⇒ It is a *serial* code that in essence solves systems of elliptic equations, on multiple grids.

⇒ It can solve for any star or black hole in isolation or in a binary configuration.

- **What is needed.**

⇒ Efficient parallelization is mandatory in the era of gravitational wave astronomy if we want to explore a large parameter space of complicated star configurations. The expertise of the NCSA staff would be ideal for such an effort.

Minimal Mixing: The least guaranteed entropy of stirring systems

Mark Bell
J. L. Doob Research Associate Professor
mcbell@illinois.edu • math.illinois.edu/~mcbell

Entropy of a mixing system can be determined from the pattern that stirring rods move in. What is the worst possible rate of mixing?

- Stirring patterns correspond to cycles in a directed graph and entropy can be found using linear algebra

Enumerating loops is totally independent

- However eliminating loops gives hints about other loops that can be skipped (an end-game database)

Need to add dependence/message passing so we can use these hints for maximal performance and support with GPU parallelisation for LA

Fast Computational Electromagnetic Tools for Manufacturing Engineered Optical Materials

Kimani C. Toussaint, Jr.

Associate Professor of Mechanical Science and Engineering

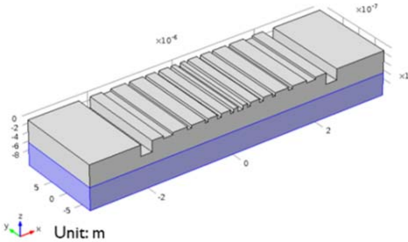
Affiliate, Depts. of Electrical & Computer Engineering, and Bioengineering

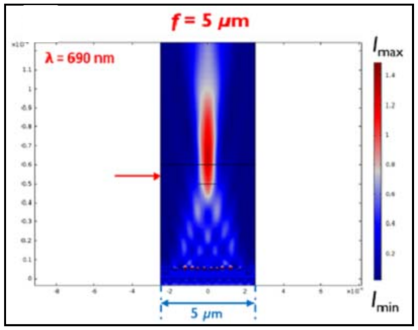
ktoussai@illinois.edu ● 217-244-4088 ● http://probe.mechse.illinois.edu/

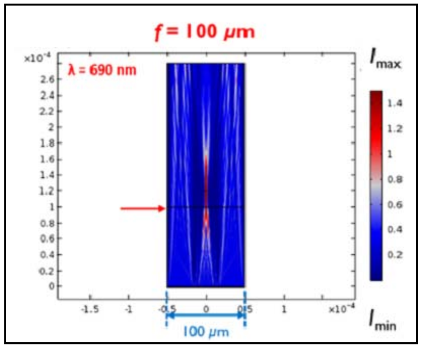
NCSA Theme Area: Materials and Manufacturing

Principal objective: Accurate modeling of optical focusing by a nanostructured planar solar collector

- *Accurately modeling optical field from any engineered optical material will facilitate fabrication strategies with clear path to low-cost nanomanufacturing*







CHALLENGES

- Full electromagnetic (EM) calculations in near-field and far-field
- Fast EM calculations for *x*, *y*, *z*, *t*

PROJECT NEEDS

- HPC resources and expertise for highly computationally intensive full (4D) EM field calculations
- Capability to handle large datasets
- Develop software front end with user-friendly interface

SYNERGIES

- Aligns with existing NCSA activities led by Seid Koric
- Potential interests from industries like 3M and NCSA's Partners
- Would lead to sustained collaboration with NCSA to pursue opportunities from agencies such as DOE, NSF, and DOD

Tying the natural environment to human health with large-scale, longitudinal data from remote sensing and Medicare

Matt Browning
brownin@illinois.edu
Recreation, Sport and Tourism; Affiliate in Natural Resources and Environmental Sciences

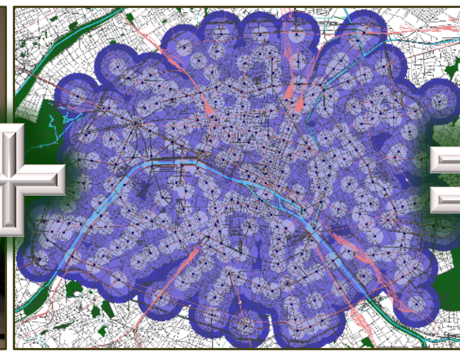
Ming Kuo *Natural Resources and Environmental Sciences*
David Molitor *College of Business*
Marilyn O'Hara Ruiz *College of Veterinary Medicine*
Nolan Miller *College of Business*

Problem Statement and Research Synopsis

Medicare data and remote sensing technology provide new opportunities for research on a scale – and with policy implications – not previously possible.

Nation-wide high resolution remote sensing data

Hospitalization and mortality data from 74M older adults from 1992-2014 *(through current DUA or new grant)*



GIS analyses of 9-digit zip codes around homes

Applications for greenspace development and urban forestry planning, policy

The Need for Supercomputing Power

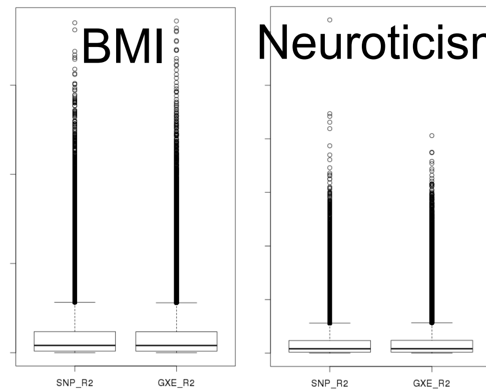
- Spatially analyze high-resolution, nation-wide maps of residential environments over time
- Create prediction models for health and greenspace with 20+ years of data from 74,000,000 people

Gene-by-Environment Interactions in Human Health & Behavior

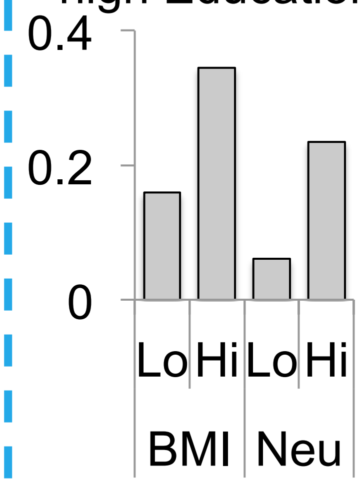
Bioinformatic and Health Sciences

JAIME DERRINGER
Assistant Professor of Psychology
derringerlab.org
jderr@illinois.edu

GxE effect sizes similar to SNP main effects



SNP-based heritability of BMI and Neuroticism increases with high Education



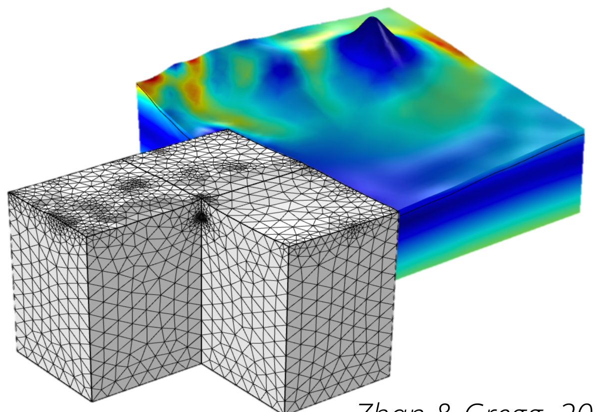
- **Problem:** gene-environment interaction theoretically pervasive in human behavior; no existing successful examples
- **Pilot study** in HRS (N = 8,652) shows promising feasibility for samples with N > 100,000
- **Preliminary R code needs optimization** by expert data scientist for scaling up
- UK Biobank provides existing data for **N > 100,000** with small access fee – **requires substantial storage & computing power**

Statistical data assimilation strategies to forecast volcanic unrest

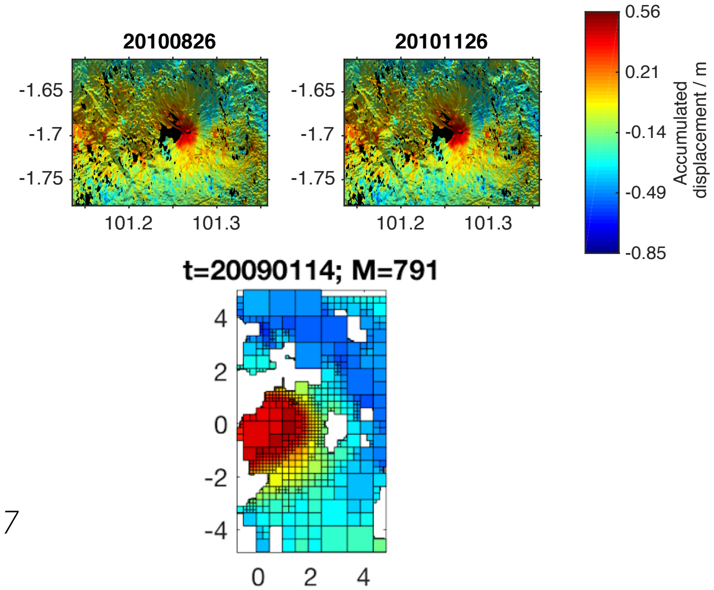
Trish Gregg

Assistant Professor of Geophysics

http://geology.illinois.edu/
pgregg@illinois.edu
@UIUCVolcanoLab



Zhan & Gregg, 2017



Project Goal:

- Create a workflow and computing framework for large ensemble statistical data assimilation of volcanic systems.
- Develop visualizations to communicate model results

Project Challenges:

The current approach is well developed for analytical models, but there are several roadblocks for adapting the ensemble-based approach with our dynamic FEMs

- Computational issues of running large Monte Carlo simulations using our parallelized FEMs
- Visualization of dynamic 4D models to illustrate volcano forecasts

Interdisciplinary effort that links **Geophysics** data, **Geodynamic** modeling, Mechanical/Material **Engineering**, **Supercomputing** Strategies, **Visualization** strategies, **Sociology** of hazard communication, **Educating** vulnerable populations, and many other areas!

Geographical, Environmental, and Social Data Analysis for Health

Diana Grigsby-Toussaint

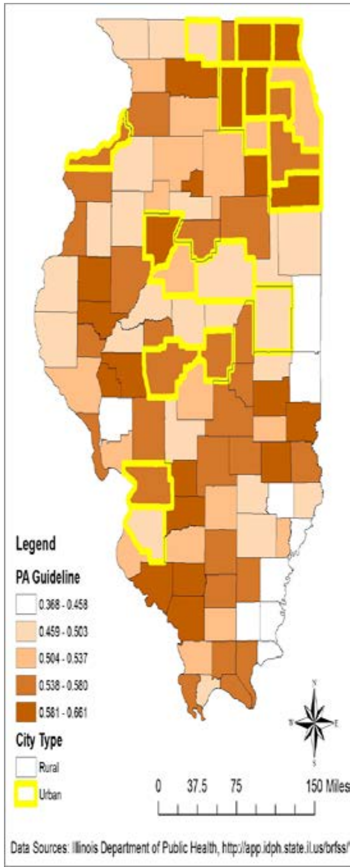
Associate Professor, Kinesiology and Community Health and Division of Nutritional Sciences

dgrigs1@illinois.edu ● 217-333-9207

- **Potential NCSA Theme Areas:**

Earth and Environment **Bioinformatics and Health Sciences**

<u>Principal objective</u>: To explore socio-environmental influences on health outcomes at multiple spatial scales over time.



CHALLENGES

- Limited data integrating multiple socio-environmental factors that influence health over time

PROJECT NEEDS

- Create a web-based environment to facilitate on-demand "what-if" analysis enabled by cyberinfrastructure
- Capability to handle large <u>datasets</u>

SYNERGIES

- May expand efforts from the <u>Earth and Environment</u> and <u>Bioinformatics and Health Sciences</u> thematic areas
- May lead to sustained collaboration with NCSA to pursue opportunities from the NIH, USDA, and Robert Wood Johnson Foundation

http://mechanical.illinois.edu/directory/faculty/ijasiuk
Contact: (217) 333-9259; ijasiuk@illinois.edu

Iwona Jasiuk Professor

Mechanical Science and Engineering

Research Interests

- Mechanics of materials
  - Multiscale characterization
  - Multiscale modeling
- Materials include:
  - Composite Materials
  - Nanomaterials
  - Novel Alloys
  - Biological materials

Applications: Materials for various industrial applications:

aerospace

power transfer

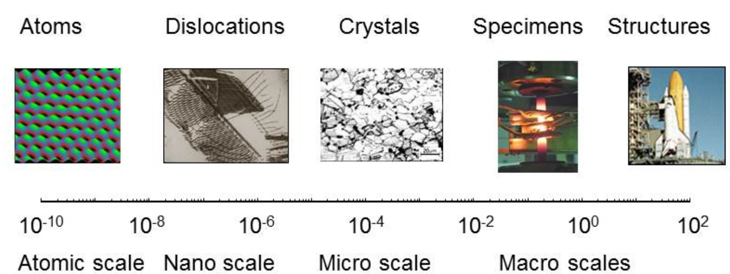
orthopaedics

Current Projects/Areas open for Collaborations

- NSF I/UCRC on Novel High Voltage/Temperature Materials and Structures http://www.hvtcenter.org/
  - Conducting polymer nanocomposites
  - Highly conductive/strong alloys
  - Impact resistant materials
  - Additive manufacturing
- Biomechanics/ Medical applications
  - Multiscale mechanics of bone
  - Bone adaptation, regeneration
  - Bioinspired materials

Ideas for collaboration with NCSA staff and faculty

- Multiscale modeling of highly conductive alloys
- Atomistic models of alloys and nanocomposites



Keywords: novel materials; nanocomposites; mechanical, electrical and thermal properties; orthopaedics; bone fracture; materials characterization; multiscale modeling of materials.

CyberGIS for Air Quality Applications: Emissions of Trace Gases from Fertilized Fields and Air Quality

Dr. Sotiria Koloutsou-Vakakis and Prof. Mark J. Rood, CEE, sotiriak@illinois.edu; mrood@Illinois.edu
http://nitrogen.cee.illinois.edu

Thematic Area: **EARTH AND THE ENVIRONMENT**
Prof. Shaowen Wang

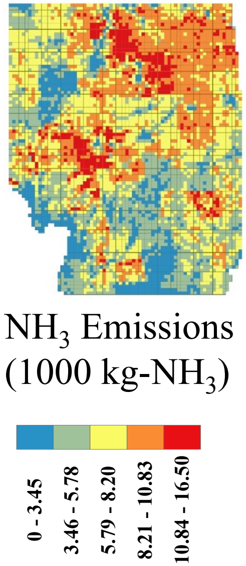
Big picture needs for Atmospheric Chemistry / Air Quality[1]

- Quantify emissions of air pollutants from natural and anthropogenic sources and quantify their removal
- Improve models for predicting formation and spatial distribution of air pollutants

Societal Context

- Human health impacts
- Understanding feedbacks with natural and managed ecosystems

→ Communication for policy, public education and to facilitate interdisciplinary work



Needed Infrastructure

- Spatially referenced databases to allow investigation of emission sources and air quality in relationship to natural and built environments with past and current data sets
- High computational capabilities and expertise for optimizing spatial data analysis, integration with air quality models, visualization and archiving

[1] The Future of Atmospheric Chemistry Research. The National Academies Press. DOI: 10.17226/235730

Efficient Stochastic Analysis with Multifidelity Molecular Models

Hadi Meidani

Assistant Professor of Civil Engineering | meidani@illinois.edu | publish.illinois.edu/meidani

Problem

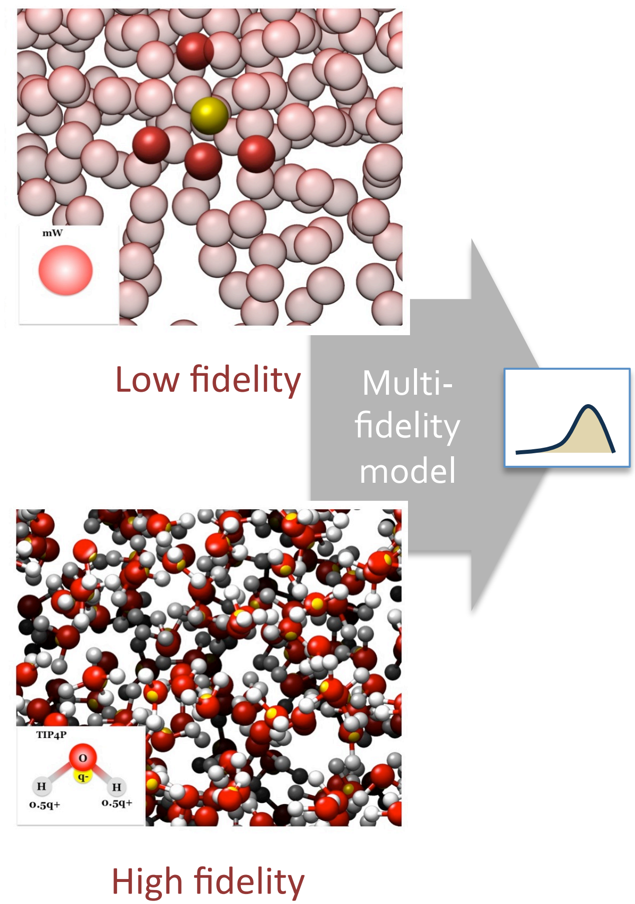
- High fidelity models are expensive to simulate
- Low fidelity models, based on a simpler physics or coarser mesh, suffer from inaccuracy

Key idea

- Combine the results of high fidelity and low fidelity simulations using our near-optimal compressive sampling regression approach

Needs

- Accelerate molecular dynamics simulations
- Accelerate the optimization solvers (a convex and a greedy) in our near-optimal regression method



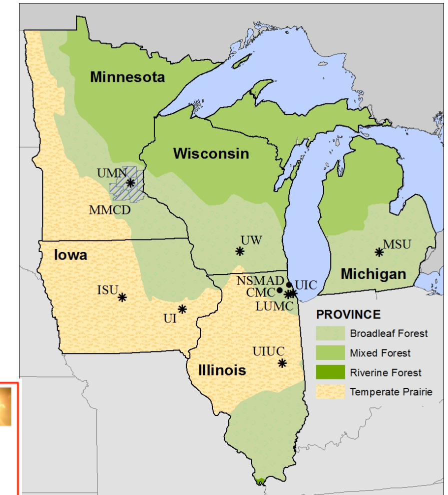
Uncertainty Quantification Group
at the University of Illinois at Urbana-Champaign

Problem: Vector-borne disease cases increasing

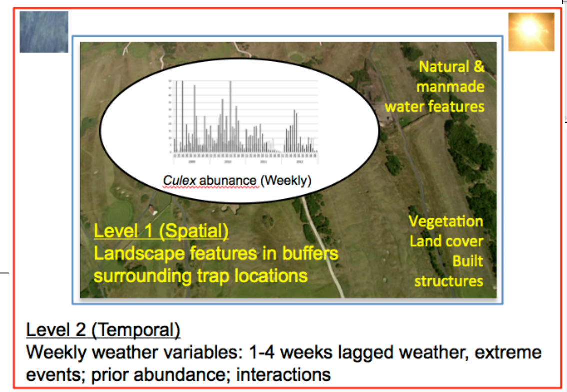
Background: Require better data and models that include weather and landscape features to predict : 1) mosquito and tick abundance, 2) variable exposure risk, and 3) disease outbreaks

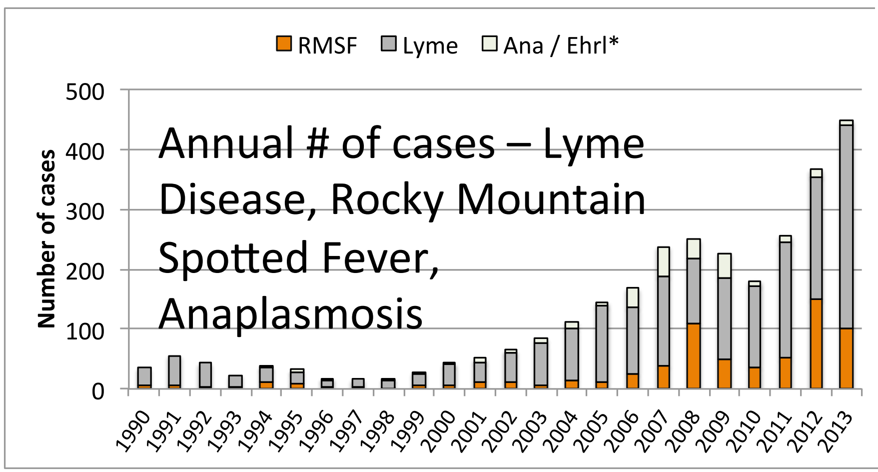
Proposed Solution: Need Advanced CyberInfrastructure, Data integration, dynamic daily weather data, efficient incorporation of new data for systems-appropriate models.

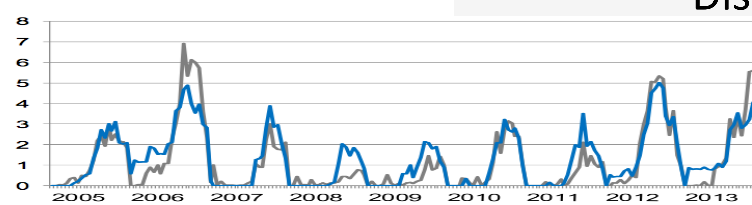
Future: Incorporate genetic data on vectors and pathogens, Delivery to public health agencies, Measure effect of control



The Upper Midwestern Center of Excellence for Vector-borne Diseases





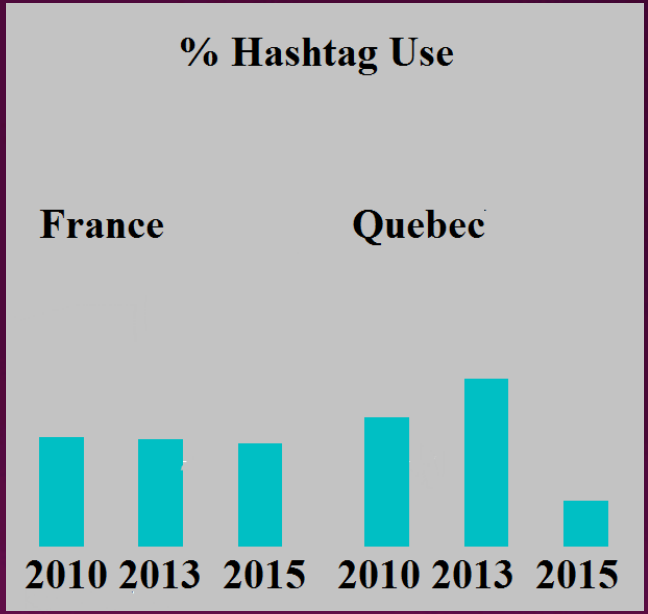


Data Driven Science for Better Control of Ticks and Mosquitoes and the Diseases they Transmit

Marilyn O'Hara Ruiz
Division of Epidemiology
Department of Pathobiology
moruiz@illinois.edu

Dialects of Twitter? Tracking local variation in vocabulary and grammar on English and French social media

- Build and store a stratified corpus of tweets
- Time depth of several years
- British & American English and French in Quebec and France
- Tools and front end to process and visualize this data
- Public access to raw and processed data (part of speech, hashtag disambiguated)



Joseph Roy
jroy042@Illinois.edu
Director, Linguistic Data Analytics Lab
School of Literatures, Cultures and Linguistics

Zsuzsanna Fagyal
zsfagyal@illinois.edu
Associate Professor, Director of Graduate Studies of
Department of French and Italian
Faculty affiliate in Linguistics, European Union Center,
LAS Global Studies, and Center for Global Studies

Alexander Schwing

Professor, Department Engineering

http://aschwing.ece.illinois.edu

aschwing@illinois.edu

January 2017

Research Interests

- Machine Learning and Computer Vision
  - Deep Learning
  - Generative Modeling
  - Multivariate and Structured Prediction
  - Optimization in Learning and Vision
  - Reasoning and Intelligence
  - Reinforcement Learning
  - Scene Understanding

Recent project result: given a single image, generate a diverse set of questions:

Is there a candle burning?

How many people are in the room?

What kind of wine is the woman drinking?

What are they celebrating?

Is this a formal occasion?

NCSA Proposal Idea Description:

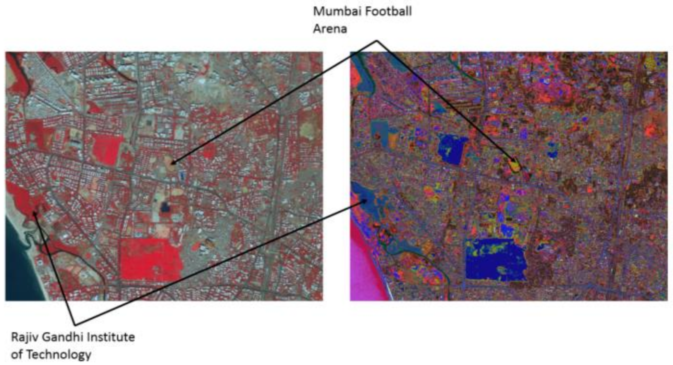
- Many applications involve joint prediction of multiple variables
- Jointly predicting multiple variables involves solving graph structured optimization problems
- Graph structured optimization problems traditionally require synchronization across cluster nodes
- Synchronization is often expensive
- Need to develop asynchronous optimization techniques

Applications:

- Computational Education and Intelligent Tutoring Systems, e.g., student modeling
- Computational Biology, e.g., protein folding
- Structured Generative Modeling Tools

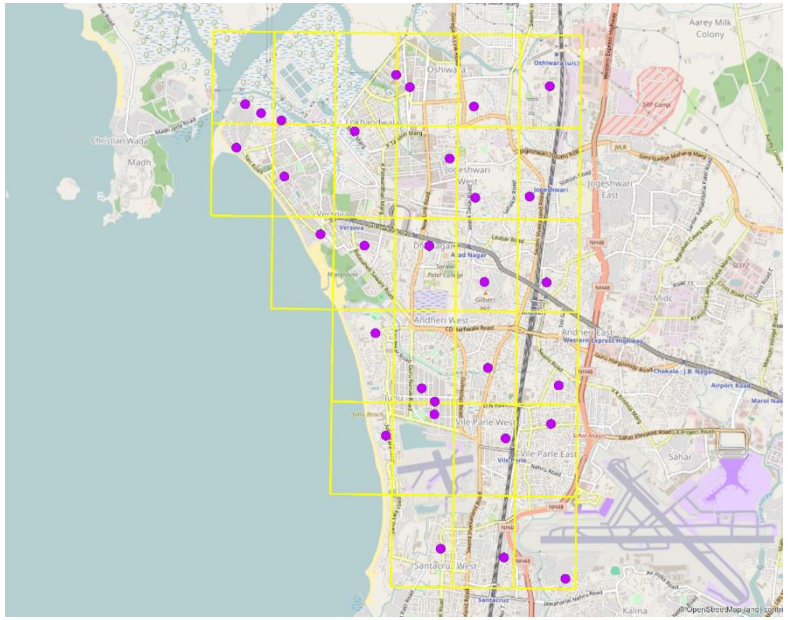
Using Real Time Traffic Data and Satellite Imagery to Understand the Relationship Between Land Use and Congestion in Developing Countries

Bev Wilson
Associate Professor of Urban & Regional Planning
bevwilso@Illinois.edu



Motivation:

- How to manage the environmental impacts of sustained and accelerating urbanization? (congestion, air pollution, etc.)
- Data limitations in developing country contexts → role of technology in bridging these gaps



Approach:

- Travel time as a proxy for congestion
- Satellite imagery and VGI

Needs (scalability):

- Data access (Google Maps API, other sources)
- Data storage and analysis (big data infrastructure and analysis tools)

Zhizhen (Jane) Zhao

http://zhizhenz.ece.illinois.edu/
zhizhenz@illinois.edu

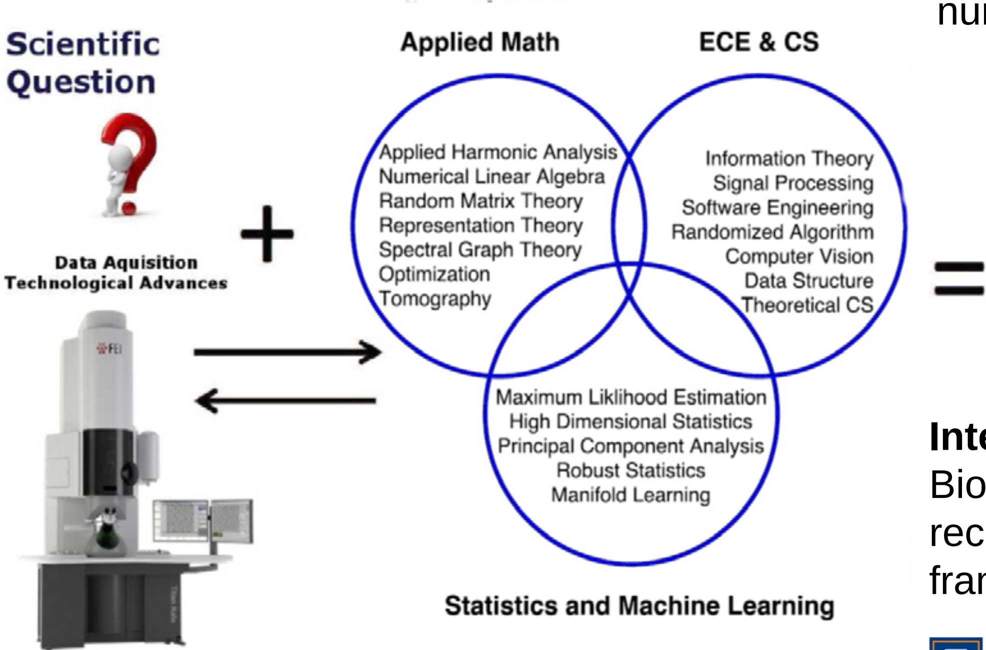
Assistant Professor, Electrical and Computer Engineering

Research Interests

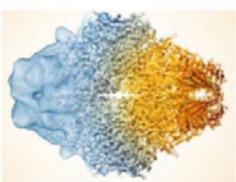
- Machine learning and data science
- Imaging sciences and inverse problems
- Scientific computing
- Structural biology

Current Projects

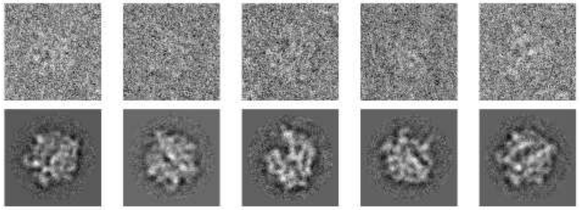
- Cryo-electron microscopy (EM) single particle reconstruction
- X-ray free electron Laser imaging
- Efficient algorithms for analyzing large noisy data with randomized numerical linear algebra



Scientific Discovery



Denoising of cryo-EM images (top row: raw images)

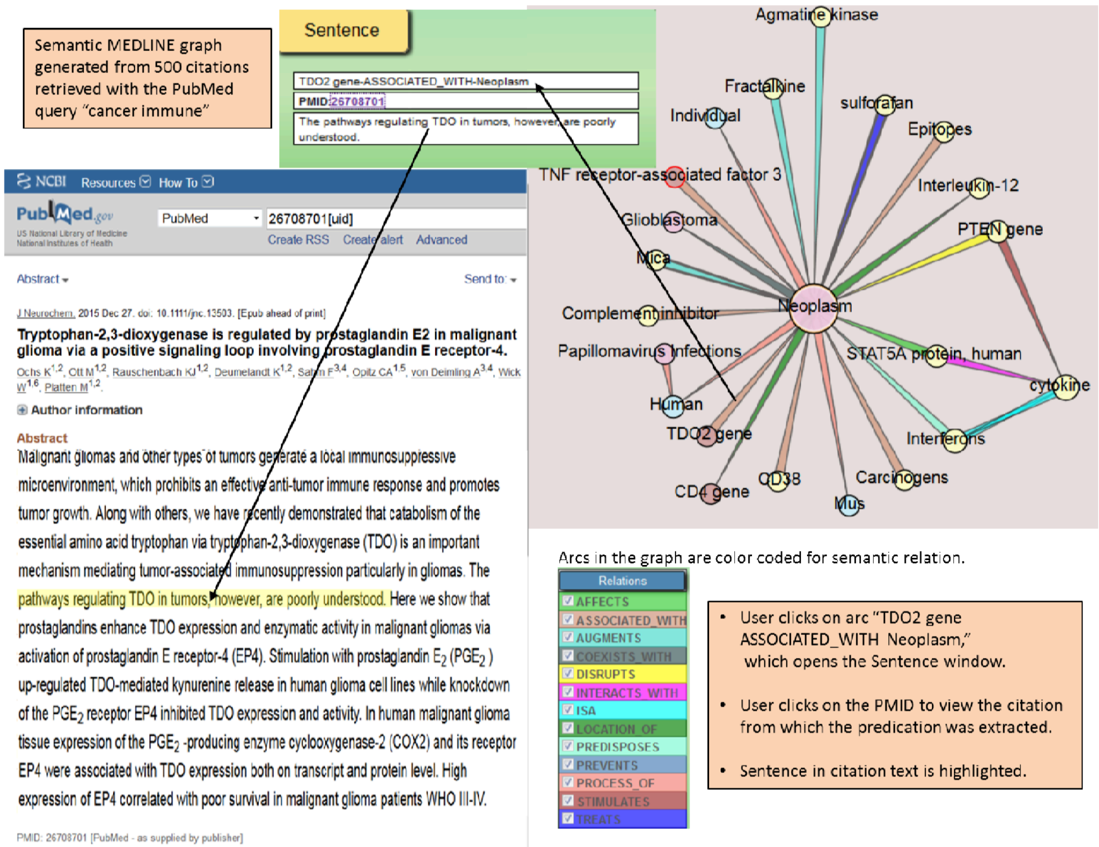


Interest areas for collaboration/future work

Biomedical imaging; machine learning and pattern recognition, signal processing; fast computational framework for single particle reconstruction.

Using Graph Databases to Visualize Connections in Biomedical Literature

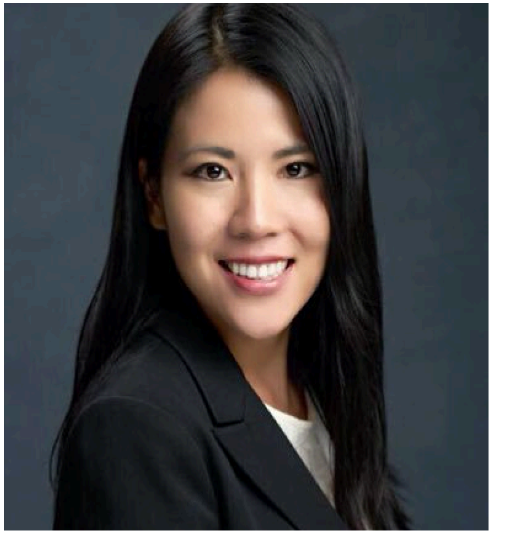
Problem: Information Overload; Silos; Visualization.
Key Idea: Develop scalable, intuitive interfaces.
Needs: Expertise in graph databases, visualization.



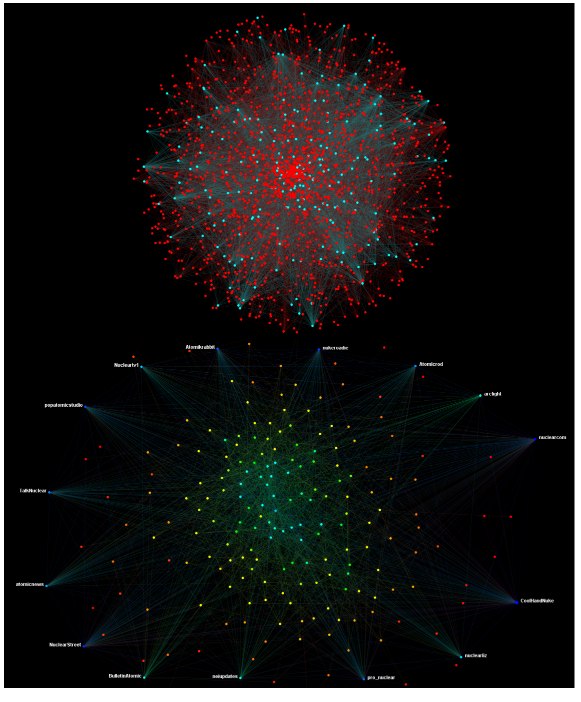
Assistant Professor
jodi@illinois.edu

School of
Information Sciences
The iSchool at Illinois

Understanding Users on Social Media : Sub-Groups in k-Partite Networks, Network Position and Topic Modeling



Wencui Han
Assistant Professor of Business Administration
wenhan@illinois.edu



Questions:

- How can we leverage the vast amount of data (mostly unstructured) generated on social media every second to provide insights for business and society ?
- How is information diffused on social networks? How do we identify the key nodes, groups, and topic trends ?

Methods:

- Identify subgroups in k-partite networks
- Understanding the structure of networks and the positioning of nodes. Information pathway simulation.
- Semantic analysis and topic modeling

Needs: machine learning, visualization of large networks, programing in Python, Wolfram language etc.

Resiliency of the order book for CME Group Futures and Options markets.

ObjectiveThe project will **develop methods to quantify and visualize** the quality and quantity of liquidity.

Mindy Mallory – Agricultural and Consumer Economics

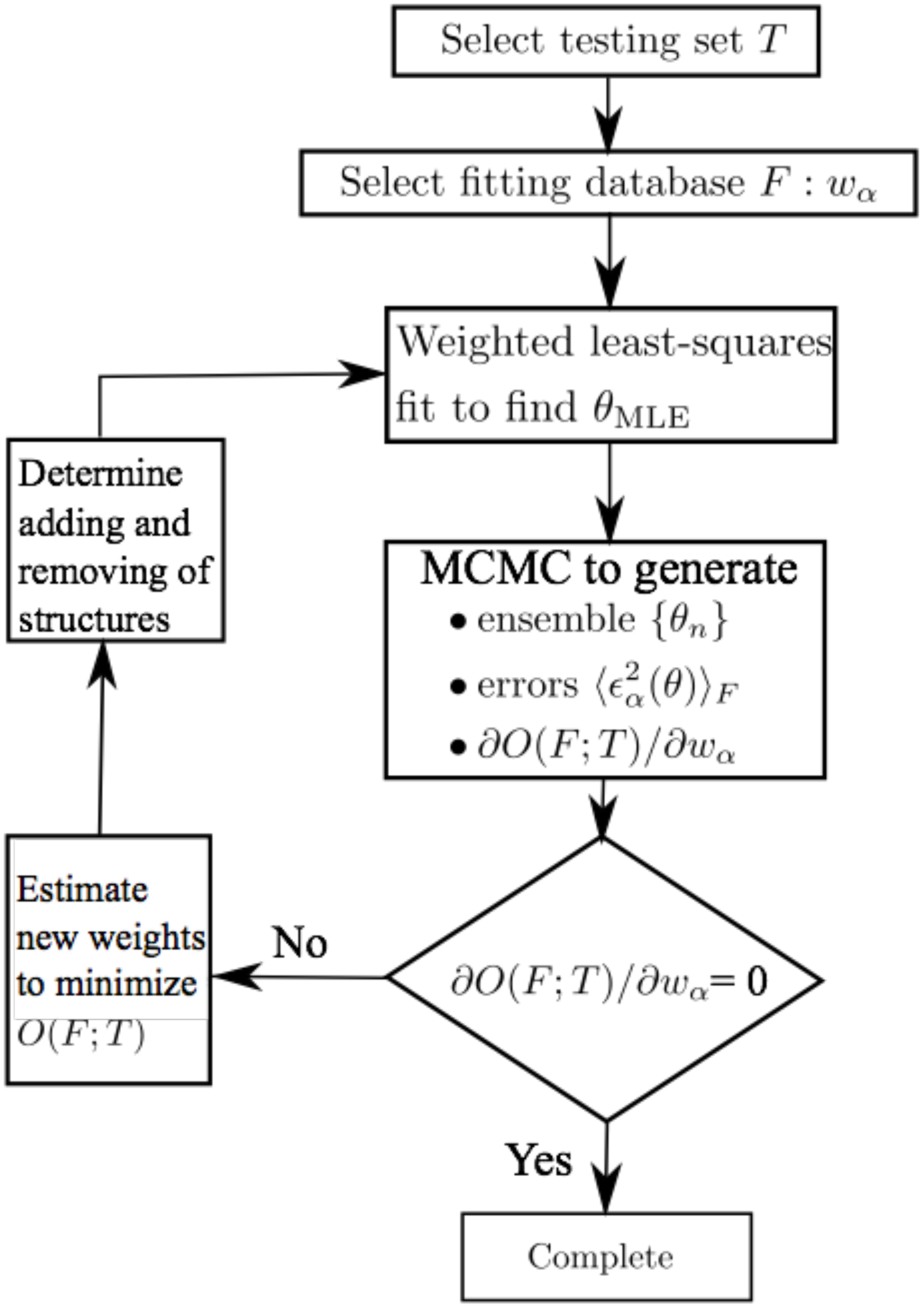
mallorym@Illinois.edu

Data used in project: All messages (orders and cancellation of orders) from CME Group futures and options contracts.

NCSA Resources Needed: Handling of large datasets, visualization of the Bid-Ask spread, computing metrics of resiliency of the book.

Dallas Trinkle: Materials Modeling Optimization

Associate Professor, Materials Science and Engineering
308 MSEB || http://dtrinkle.matse.illinois.edu
dtrinkle@illinois.edu



- optimal parameter θ_{MLE}
- estimate of prediction errors from $\{\theta_n\}$
- transferability measure as testing set T

Goal: *accurate, automated, robust* empirical models of atomic interaction for materials simulation

- *New algorithm* combines **genetic optimization** of fitting database with **stochastic sampling** of parameter space for a Bayesian error estimates.
- **Needs from NCSA:**
  - proof-of-principle algorithm taken from parallel code up to large scale massive parallelization
  - improved parallel optimization and algorithms
- **Impact:**
  - basic element of **predictive multiscale modeling** from electronic structure up through atomic scale (and even mesoscale)
  - novel use of **Materials Data Facility** data
  - strengthen connections between **materials modeling across campus** and NCSA for **future center proposals**

Multiscale modeling for the structure-property relations of polymeric gels

Yuhang Hu
Assistant Professor
Mechanical Science and Engineering
Email: yuhanghu@Illinois.edu
https://publish.illinois.edu/yuhanghu/

Problem:

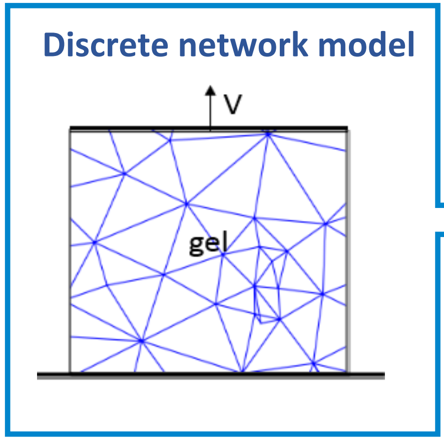
- Polymeric gels are widely used as smart materials in drug delivery, soft robots, microfluidic device, tissue culture scaffold etc.
- Fundamental understanding of the structure-property relation and predictive design in applications requires a multiscale physics based model.

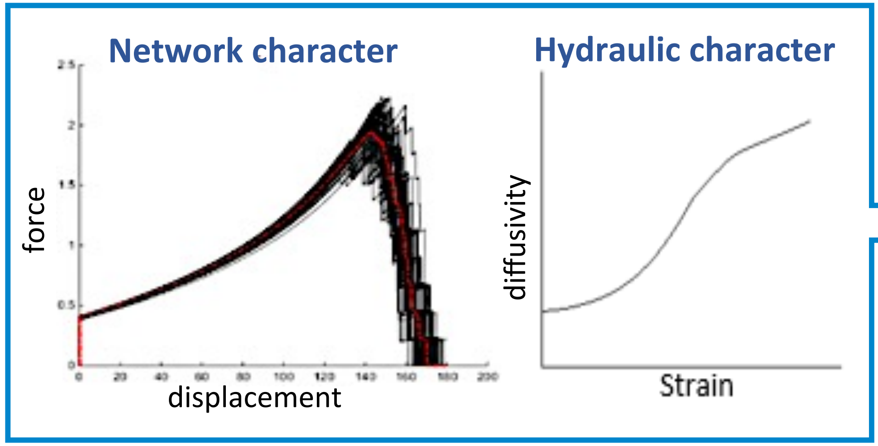
Key idea:

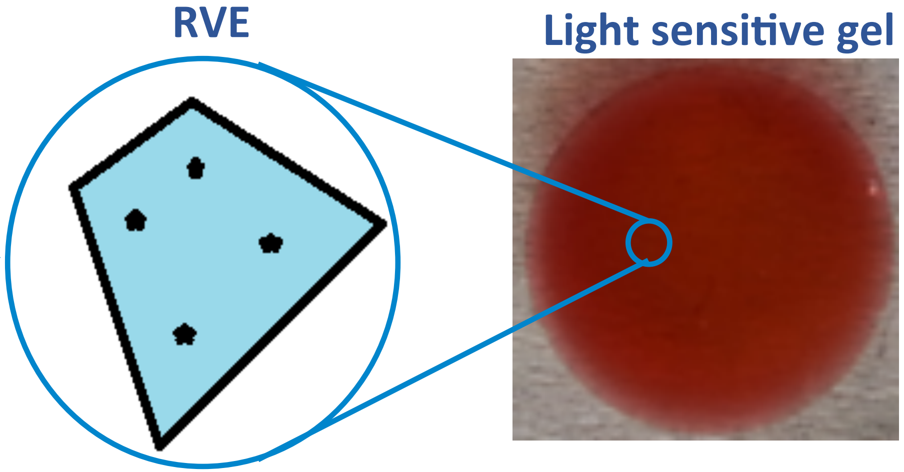
- A new multiscale computational model that is informed by statistical thermodynamics and coupling discrete microstructural features and effective continuum properties

Needs:

- Parallel computing to simulate 3D and big networks for representative RVE characters.
- Super computing infrastructure to compute structures with severe nonlinear deformation, stress concentration and multiphysics driving force.



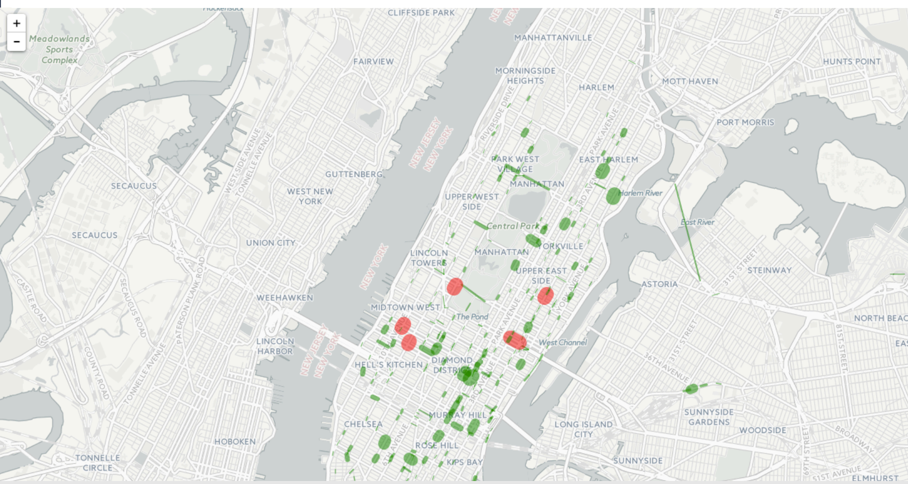




Big data and mobility: insights and visualizations

Richard Sowers r-sowers@Illinois.edu; Professor, Industrial and Enterprise Engineering & Mathematics

Dan Work dbwork@Illinois.edu; Assistant Professor, Civil and Environmental Engineering



Using data analytic tools and novel mathematics to understand new mobility challenges.

Large empirical datasets from New York City + Chicago

Thanks to PI4 and IGL (Dept. of Mathematics)

http://www.math.illinois.edu/~r-sowers/ResearchGroup/MTP.html