

Spatial Data Science: Interdisciplinary Experiences

Shashi Shekhar

McKnight Distinguished University Professor

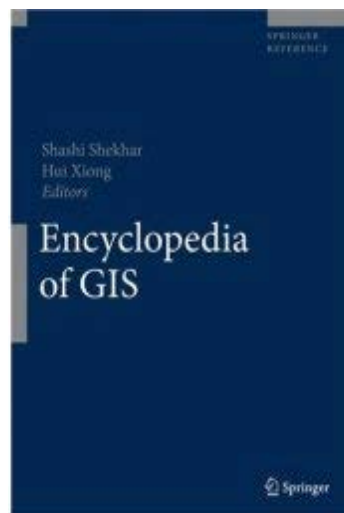
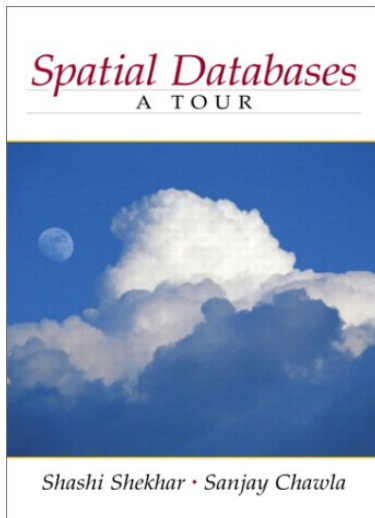
Dept. of Computer Sc.& Eng., University of Minnesota

www.cs.umn.edu/~shekhar

April 10th, 2017

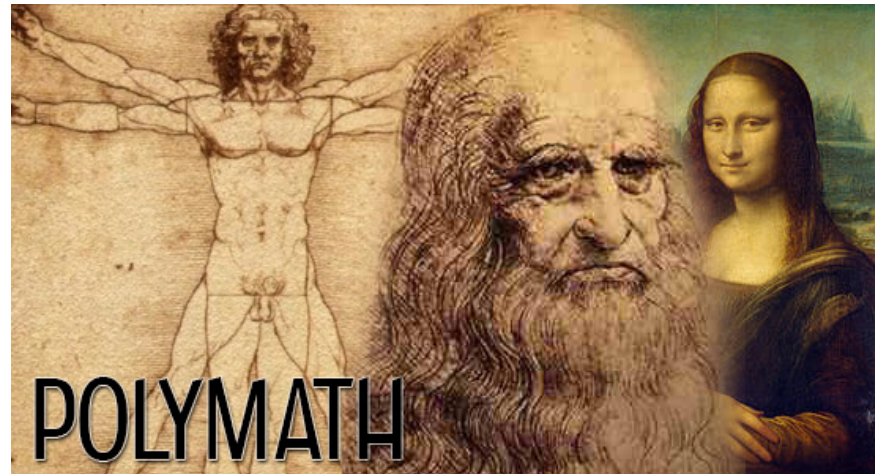
Nebraska Data Analytics Workshop

University of Nebraska, Lincoln, NE



Outline

- **Interdisciplinary Research**
 - Motivation: Intellectual & Societal
 - A definition
- Academic Example with Two disciplines
- Multi-sector Example
- Conclusions



Societal Motivation

- Q? Identify disciplines of societal challenges?
 - Cyber-security, e.g., Fake news, Reduce cancer deaths, Climate change
 - Poverty, Hunger, Health & Well-being, Inequality, ...
 - Interactions: Feed growing population while protecting water
- Q? Identify disciplines of non-academic organizations?
 - Government, e.g., USDOD, USDA, USDOT, NIH, ...
 - Businesses, e.g., Google, Amazon, GE, Berkshire Hathaway, ...



Intellectual Motivation

- Q? Identify disciplines of scholars?
 - Aristotle, Galileo, Da Vinci, Brunini, ...
 - Polymaths, Unquenchable curiosity
- History of Disciplines
 - Only a few hundred years old
 - A way to organize universities
 - Budget, hiring, degrees, courses,...
 - + intra-discipline communication
 - - inter-disciplinary comm., questions, hiring, education, ...
- Analogies
 - Holistic (Systems view) vs. Reductionist
 - Nationalism vs. Globalism



"I'M ON THE VERGE OF A MAJOR BREAKTHROUGH, BUT I'M ALSO AT THAT POINT WHERE CHEMISTRY LEAVES OFF AND PHYSICS BEGINS, SO I'LL HAVE TO DROP THE WHOLE THING."

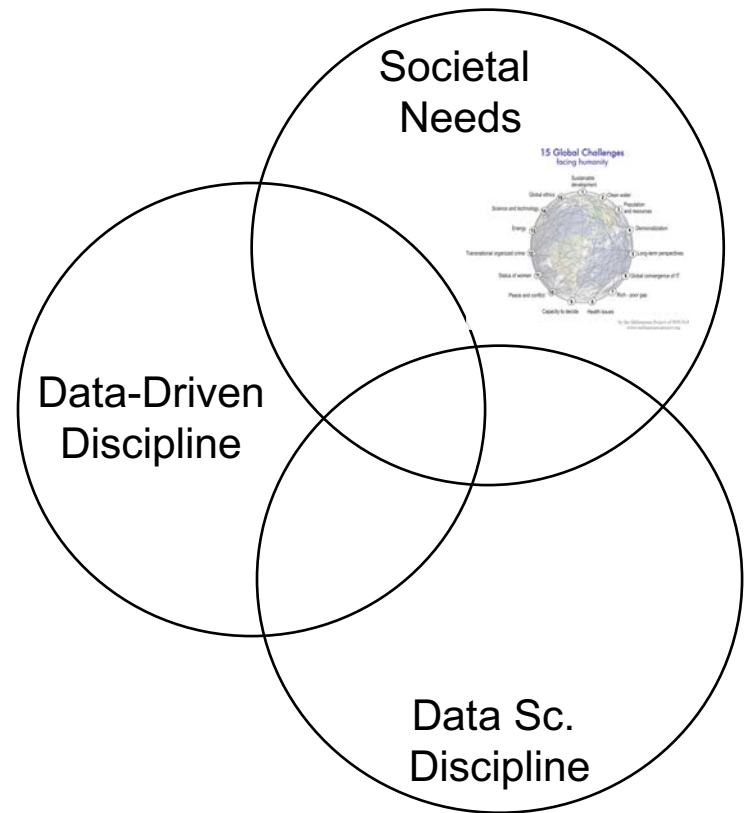
Interdisciplinary Research

- integrates information, data, techniques, tools, perspectives, concepts, and/or theories from two or more disciplines or bodies of specialized knowledge
- to advance fundamental understanding or to solve problems whose solutions are beyond the scope of a single discipline or area of research practice.



Outline

- Why Interdisciplinary ?
- Example with Two disciplines
 - Challenges
 - A Case Study
 - Roles of Computer Science in collaboration
 - Understanding Computer Science
- Multi-sector Example
- Community Building
- Conclusions

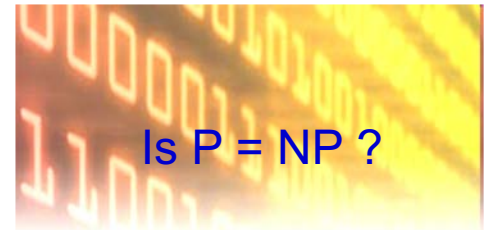


What is Interdisciplinary Research?

- Is it multiple Disciplines working on a single project?
- Is it one discipline helping another?
- My Thoughts:
 - Ideal: Perform research that **enhances all disciplines** involved.
 - Not just a subset!
 - Very Hard To Do!!!
 - A lot of asking questions back and forth

Challenges: Communication, Trust, Finding Win-Win

- Each Discipline has its own language!
 - We need to learn (foreign) language of other discipline
 - Homonym (Civil Eng., C. S.): infrastructure, network, bridge, ...
 - Synonyms (Civil Eng., C.S.): networks vs. graphs
- Disciplines have different values, cultures and goals
 - Different norms for ordering of authors in a joint paper
 - Different emphasis on grants, citations, service, ...
- Misunderstanding of what each discipline really is
 - e.g., “I thought Civil Eng. was all about building bridges!”
 - e.g., “I thought Computer Sc. was all about programming!”
- Break down barriers
 - Build mutual trust and respect via social activities, sharing core competencies, ...
 - Keep talking to each other to find a win-win
 - Have an open mind when discussing each others' interests



Disciplines

Civil (Env.) Engineering

- Professors
 - William Arnold
 - Ray Holzalski
 - Miki Hondzo
 - Paige Novak
- Students
 - Mike Henjum
 - Christine Wennen

Computer Science

- Professors
 - Shashi Shekhar
- Students
 - Jim Kang

Hydrology

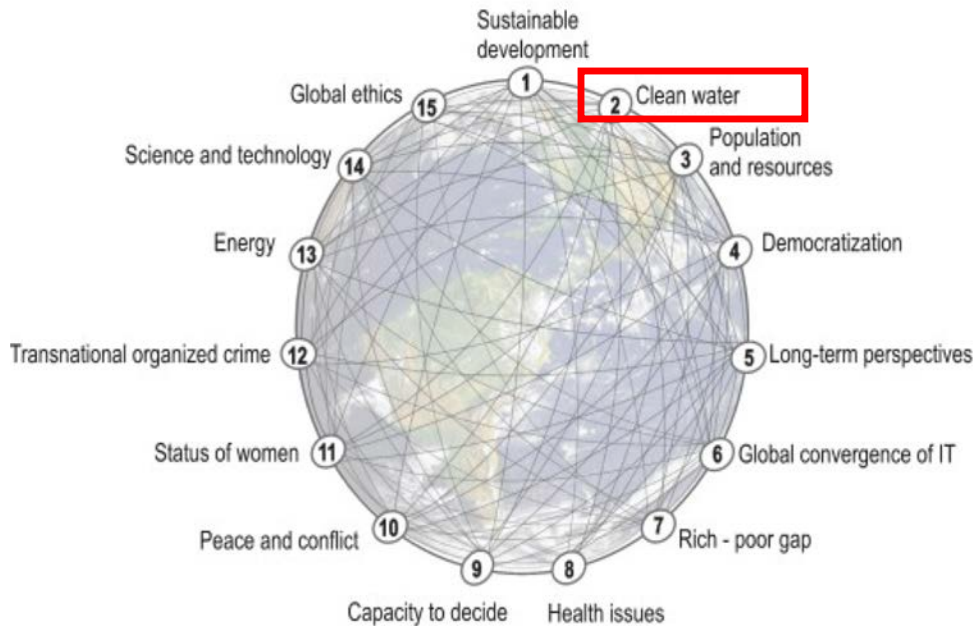
Professors

- David Maidment

Ack: NSF IGERT: Non-equilibrium Dynamics Across Space and Time: A Common Approach for Engineers, Earth Scientists and Ecologists (2008-2012).

Societal Challenge: Clean Water

15 Global Challenges facing humanity



by the Millennium Project of WFUNA
www.millennium-project.org



By 2025, 1.8 billion people could be living in water scarce areas

- Today, 750 million people live below the water-stress threshold of 1.7 K

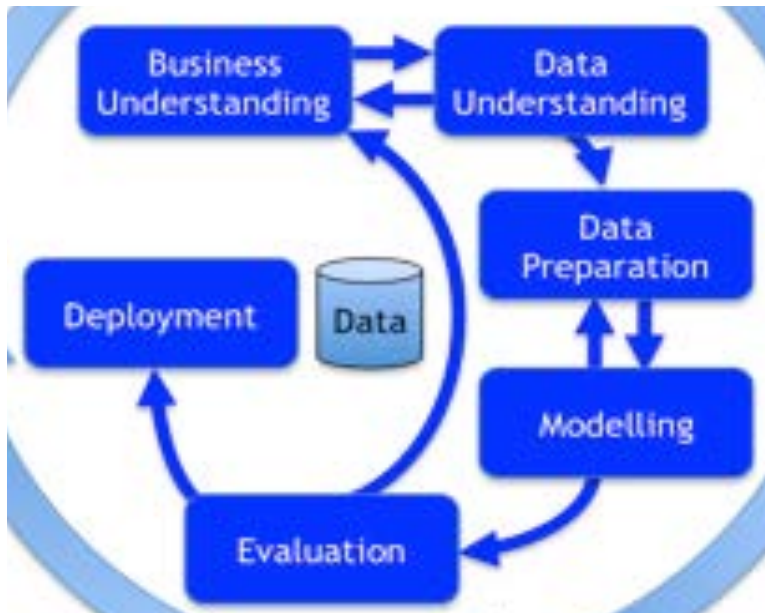
Brainstorming: In the Beginning...

Civil Engineering Questions

- How is Computer Science involved in this work?
- CS: I don't know!
- Need to understand domain questions and dataset first

Computer Science Questions

- How many sensors will there be? Like 1000 or 10,000 or more?
- CE Ans: No Way! A sensor cost \$30,000
- CS: Large data or computations ?
- CE: Time-series length is in 50,000
- CS: time-series similarity query?
- CE: What is similarity?
- CS: time-series correlation
- CE: Not yet!



CRISP-DM Methodology

Brainstorming: A little later...

Civil Eng. Questions

- Can you help publish field-sensor data on national website (HIS) ? dealing with wireless communication, data format conversion, XML, etc.?
- Can you remove errors from the dataset?
- CS Ans:
 - Yes, we will help with these.
 - But these do not advance CS
 - Techniques already exist

Computer Science Questions

- Do you want to know how fast the river is flowing?
- CE Ans:
 - Not really,
 - We can already determine that by the discharge, water depth, and physical characteristics of the river

Brainstorming: Light at the end of the tunnel

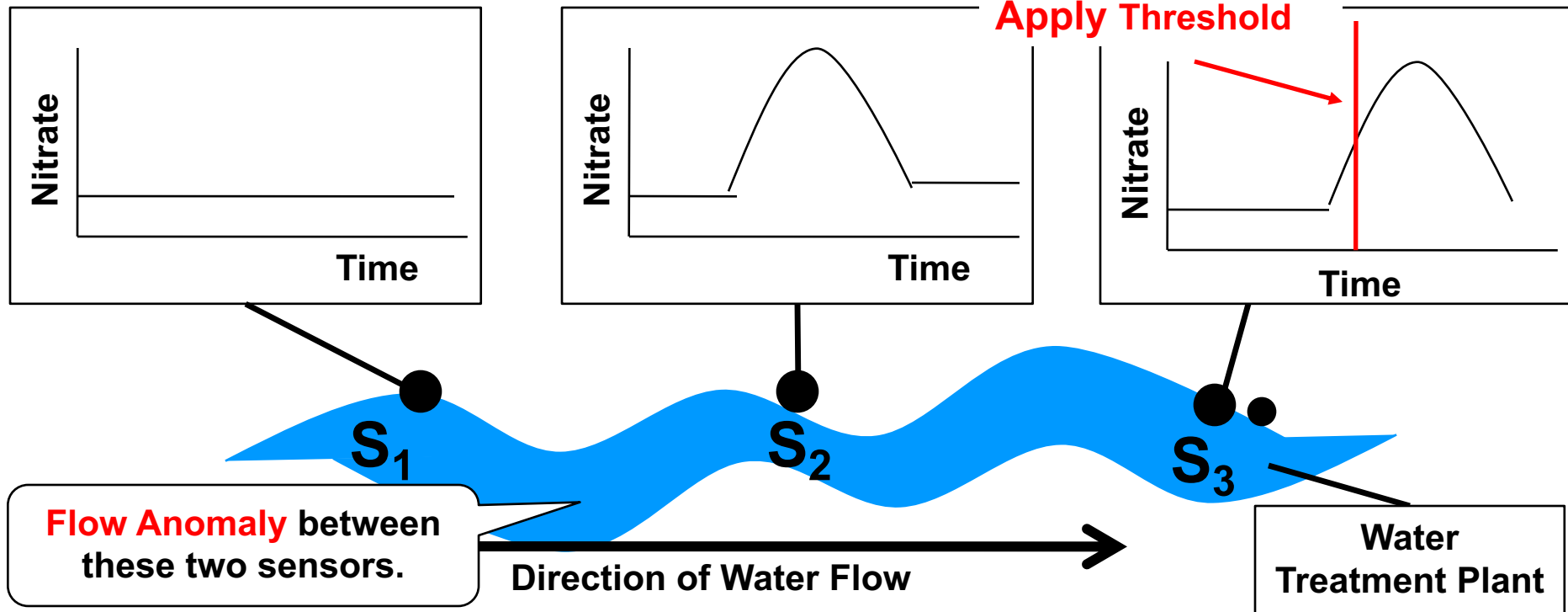
Civil Engineering

- Can you help with Env. Forensics?
- Find **when and where** a contaminant enters a river?
- CS: Is there signature in sensor data?
- CE Ans.
 - Signature across a pair of sensors.
 - Upstream sensor sees clean water.
 - Downstream sensor sees contamination.
- CS.:
 - How hard is it manually? computationally?
- CE Ans.
 - Its too hard to find this manually
 - e.g., hours to sift through the data
 - 50,000 data points per measured variable
 - Quadratic effort, e.g., 2,500,000,000 pairs

Computer Science

- Let me see. Spatial outlier detection won't not catch it!
- A new pattern family: Flow Anomaly: time periods of flow mismatch between up/downstream sensors?
- It may advance Computer Sc.
- Which **Flow Anomalies (F.A.)** ?
 - Transient
 - Persistent
 - Dominant Persistent
- CE Ans:
 - Yes!
 - Dominant Persistent F.A.s please.

Problem Formulation: Flow Anomaly



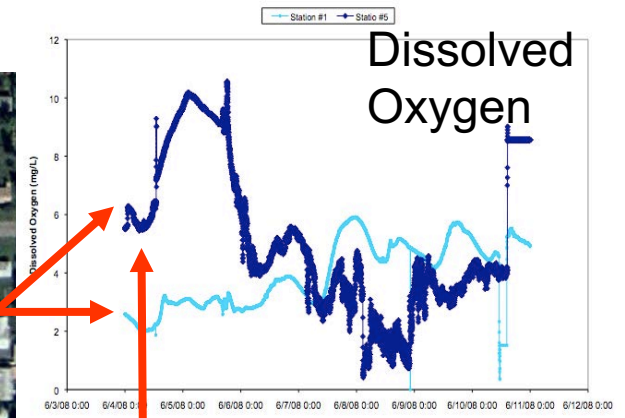
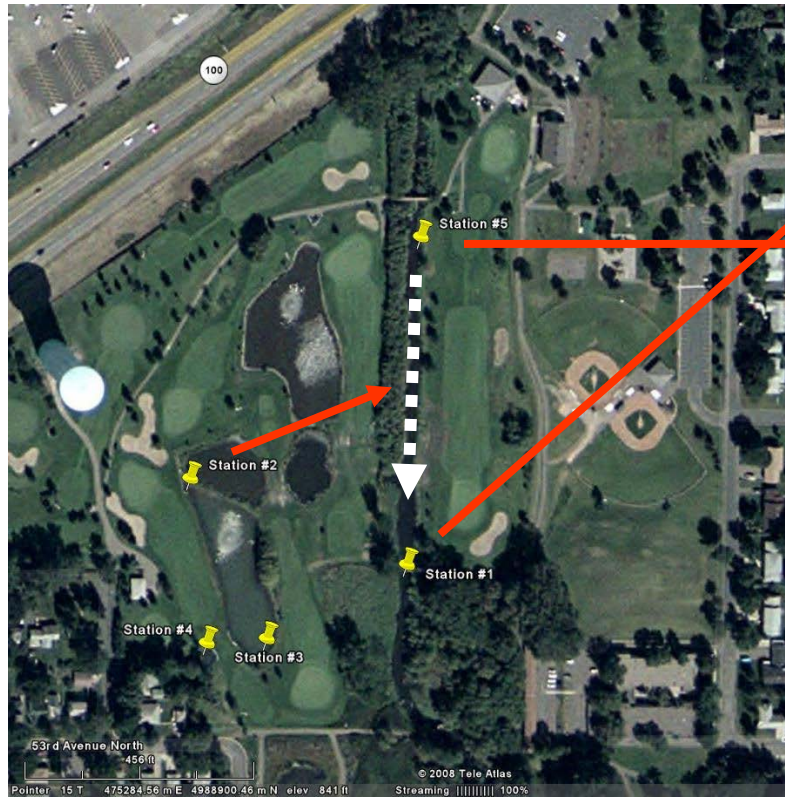
Two Use Cases:

- At the water treatment plant, when should it turn off the water supply from the river?

Where is the source of the contaminant?

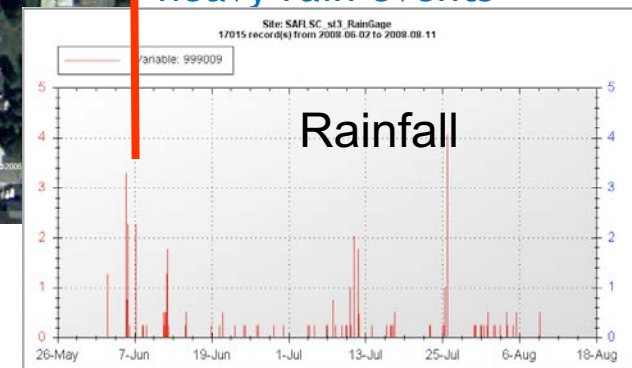
Env. Sc.:Forensics: When and where do contaminants enter Shingle Creek?

Computer Sc.: Scalable detection of flow-anomalies & co-occurrences



Flow anomaly

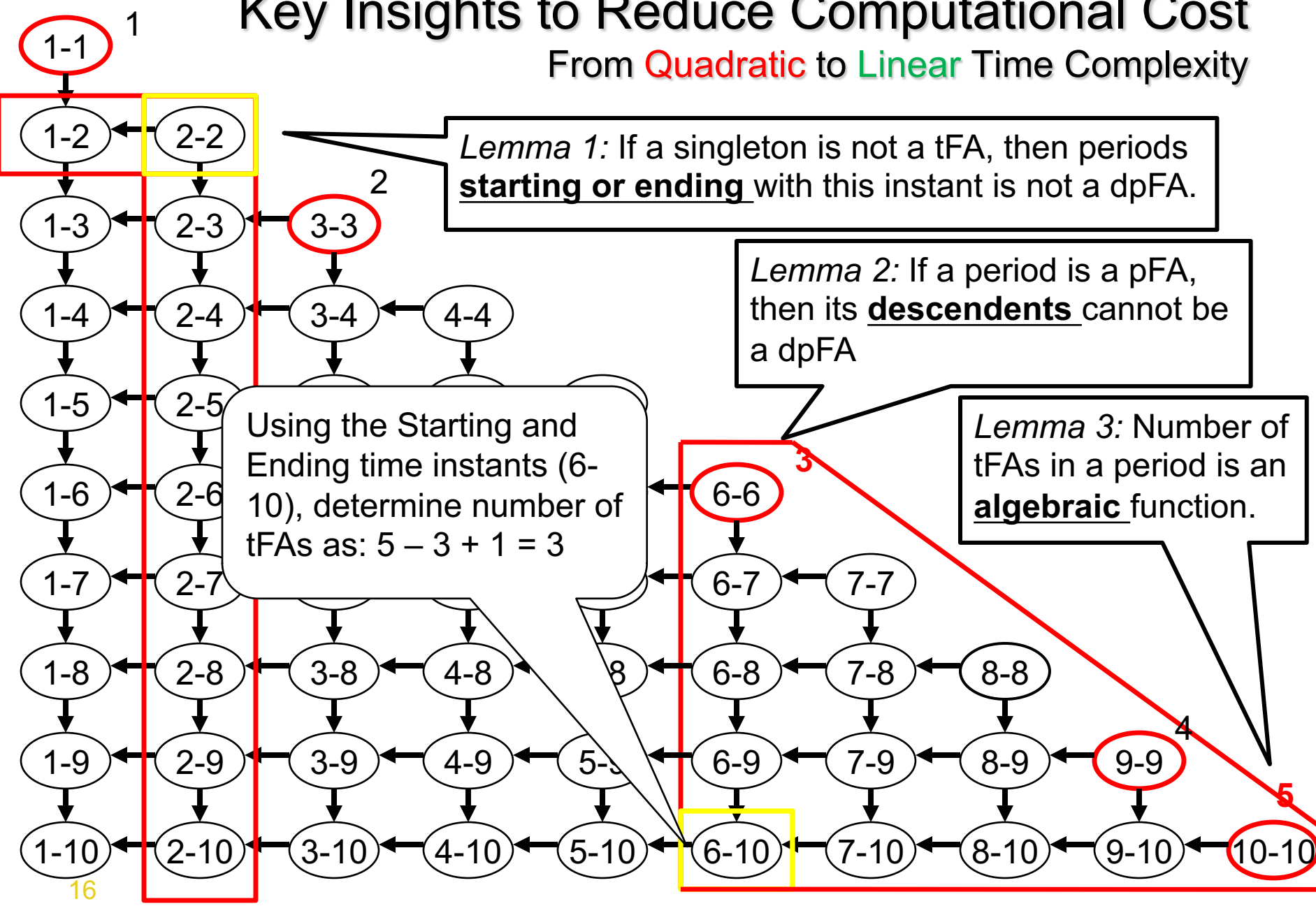
After consecutive heavy rain events



J. M. Kang, S. Shekhar, C. Wennen, and P. Novak,
Discovering Flow Anomalies: A SWEET Approach, IEEE Intl. Conf. on Data Mining, 2008. 15
Ack: NSF IGERT, CISE/IIS/III, USDOD.

Key Insights to Reduce Computational Cost

From **Quadratic** to **Linear** Time Complexity



Contributions & Impact

Civil (Env.) Sc. Contribution:

- State of the Art: Fate, e.g., Where do various contaminants go?
- Contribution: Forensics, e.g., Where did the contamination come from?

Computer Sc. Contribution:

- Fast algorithms for a problem where Dynamic Programming doesn't help.

Broader Impact:

- IGERT Fellow James Kang joined USDOD-NGA Research in 2010
- IGERT helped in adapting to and innovating there
- Received 2016 [Presidential Early Career Award for Scientists and Engineers \(PECASE\)](#).



National Geospatial-Intelligence Agency ✓

January 10 · 🌐

Congrats to NGA's James Kang and other intelligence community early-career scientists at U.S. Army, NSA - National Security Agency, Office of the Director of National Intelligence & more, honored with 🏆 #PECASE 🏆 awards!



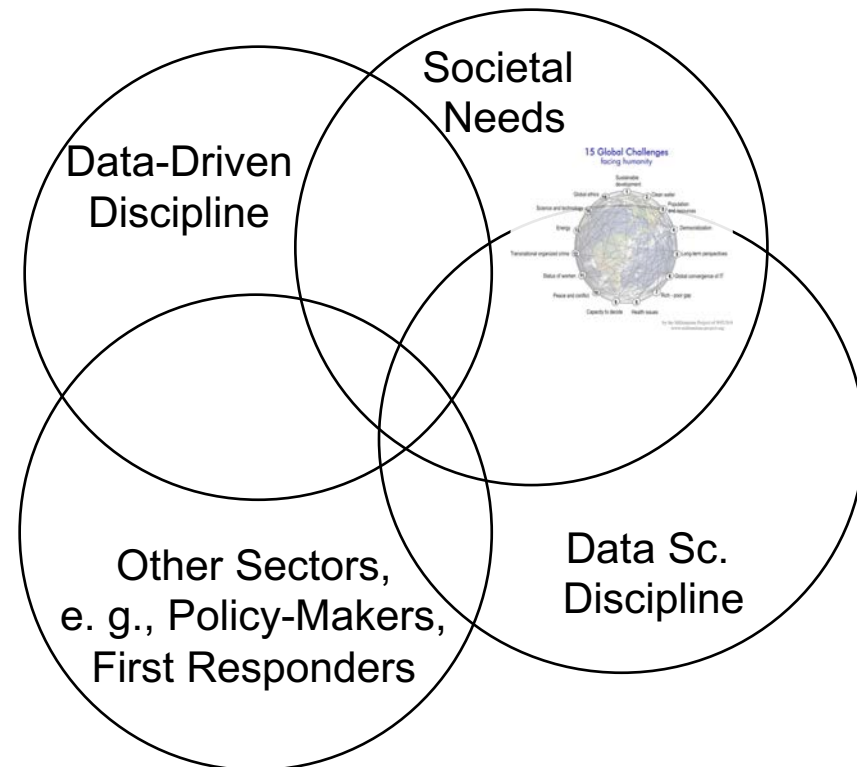
President Obama Honors Federally-Funded Early-Career Scientists

Today, President Obama named 102 scientists and researchers as recipients of the Presidential Early Career Awards for Scientists and Engineers...

WHITEHOUSE.GOV

Outline

- Why Interdisciplinary ?
- Academic Example with Two disciplines
- **Multi-sector Example**
 - Role of Science in Policy/Decision Making
 - A Case Study
- Community Building
- Conclusions



Societal Challenge: Large Scale Evacuation

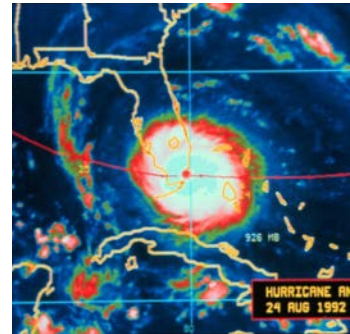
Hurricane: Andrews, Rita

- Traffic congestions on all highways
 - e.g. 100-mile congestion (TX)
- Great confusions and chaos

"We packed up Morgan City residents to evacuate in the a.m. on the day that Andrew hit coastal Louisiana, but in early afternoon the majority came back home. **The traffic was so bad that they couldn't get through Lafayette.**"

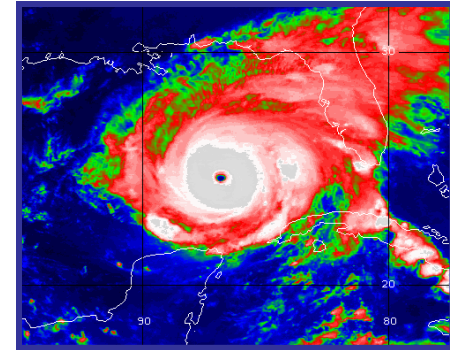
Mayor Tim Mott, Morgan City, Louisiana
(<http://i49south.com/hurricane.htm>)

Florida, Louisiana
(Andrew, 1992)



(National Weather Services)

Houston
(Rita, 2005)



(National Weather Services)



(www.washingtonpost.com)



I-45 out of Houston
(FEMA.gov)

Metropolitan Wide Evacuation Planning - 2

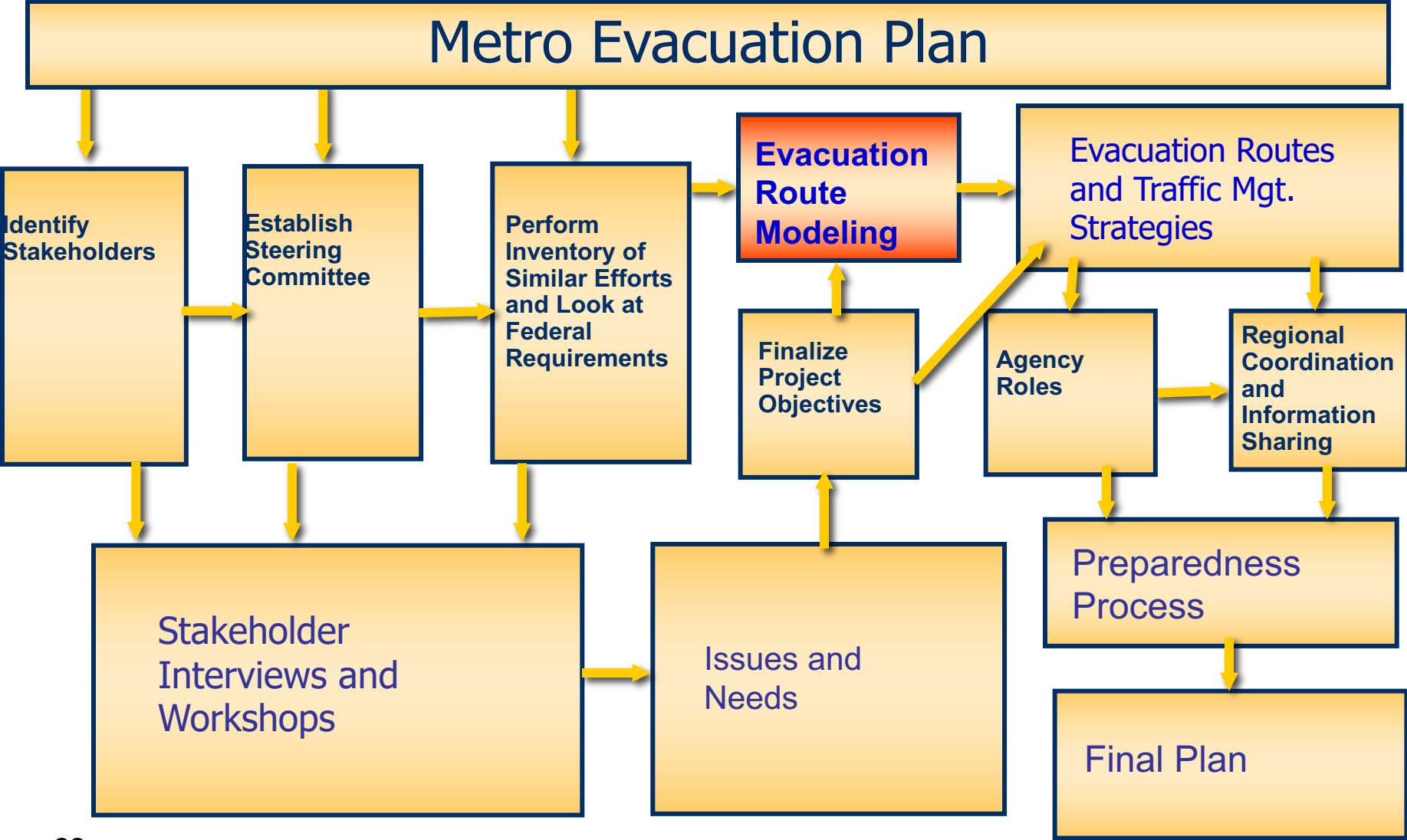
Advisory Board

MEMA/Hennepin Co. -	Tim Turnbull, Judith Rue
Dakota Co. (MEMA) -	David Gisch
Minneapolis Emergency Mgt. -	Rocco Forte, Kristi Rollwagen
St. Paul Emergency Mgt. -	Tim Butler
Minneapolis Fire -	Ulise Seal
DPS HSEM -	Kim Ketterhagen, Terri Smith
DPS Special Operations -	Kent O'Grady
DPS State Patrol -	Mark Peterson

Workshops

Over 100 participants from various local, state and federal govt.

Multi-Sectoral Task-structure



Computational Problem: Evacuation Route Planning

Given

- Number of evacuees and their initial locations
- Evacuation destinations
- A transportation network, a directed graph $G = (Nodes, Edges)$ with
 - Travel time for each edge
 - Capacity constraint for each edge and node

Output

- Evacuation plan = a set of origin-destination routes & schedule

Objective

- Minimize evacuation time

Constraints

- (Spatio-temporal) **Data Feasibility**
- **Computational Scalability** to large population and geographies
- (Societal) Application Domain Interpretability and value
 - Transportation Network Capacity \ll Number of evacuees

Computational Feasibility

Computational Scalability to large population and geographies

Is it computable in reasonable time with current hardware, software ?

Does it require computer science advances?

A. Related Work in Transportation Sc.: Dynamic Traffic Assignment

- Game Theory: Wardrop Equilibrium, e.g. DYNASMART (FHWA), DYNAMIT(MIT)

Limitation: Nice Game Theory, but Extremely high computation time

- Does not scale to medium size networks and populations!

B. Related Work Operations Research: Time-Expanded Graph + Linear Prog.

- Optimal solution, e.g. EVACNET (U. FL), Hoppe and Tardos (Cornell U).

Limitation: - High computational complexity => Does not scale to large problems

- Users need to guess an upper bound on evacuation time

Inaccurate guess => either no solution or increased computation cost!

Number of Nodes	50	500	5,000	50,000
EVACNET Running Time	0.1 min	2.5 min	108 min	> 5 days

C. Computer Science: Capacity Constrained Route Planner

- Extends shortest-path algorithms to honor capacity constraints
- Scales up to Millions of evacuees over hundreds of square miles

Contributions:

- Computer Science: New algorithm to scale up to metropolitan scenarios
- Transportation Science: Walking first mile speeds up evacuation by factor of 3
- Policy: Increase public trust in evacuation plans

Societal Impact highlighted in Fox TV (KMSP) evening news

A 5-minute video-clip is at <http://www.cs.umn.edu/~shekhar/talk/video/fox9 aired mpg.avi>

It highlighted the evacuation route planning work along with its social impact including those on evacuation plan for Minneapolis-St. Paul Twincities.



TWIN CITIES METRO EVACUATION PLAN

TECHNICAL MEMORANDUM #1

UNCLASSIFIED/FOR OFFICIAL USE ONLY (FOUO)
MAY BE DISSEMINATED ON A "NEED TO KNOW" BASIS
NOT FOR MEDIA OR PUBLIC DISTRIBUTION
PROPERTY OF THE MINNESOTA DEPARTMENT OF
TRANSPORTATION

Details: Experiences with evacuation route planning algorithms, International Journal of Geographical Information Science, 26(12): 2253-2265, Taylor and Francis, December 2012.

Outline

- Why Interdisciplinary ?
- Academic Case with Two disciplines
- Multi-sector Case
- Community Building
- **Conclusions**

Lessons Learned

- Interdisciplinary Research is rewarding but HARD
- Reward: Solve complex societal problems
 - Which may not be amenable to a single discipline
- Hardest part is trying to understand the other domain
 - Quote: “If you think Mathematics is hard, try learning Chinese.” - Peter Bol
 - Analogy: “If you think learning Chinese is hard, try
 - Learning language, culture and strengths of another discipline
 - Developing trust across disciplines to find win-win
 - Introspecting to understand when one needs help from another discipline
- Crucial that both sides understand each other before research can begin
- A lot of trial and error between both sides
- Once an “Ah-ha” moment occurs
 - The number of opportunities can be unlimited!