Spatial Computing

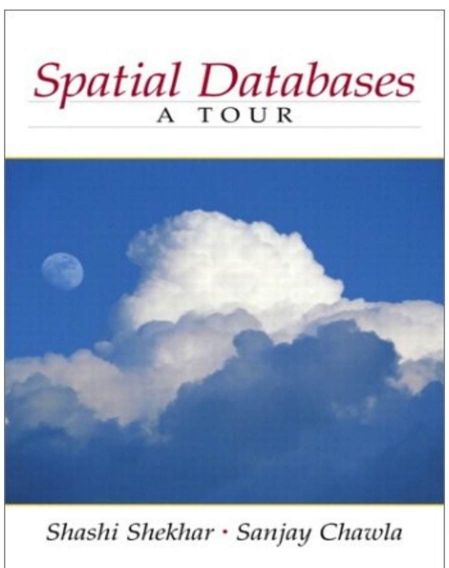
Shashi Shekhar

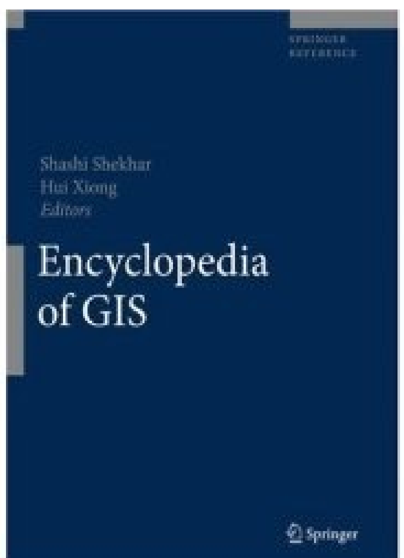
McKnight Distinguished University Professor

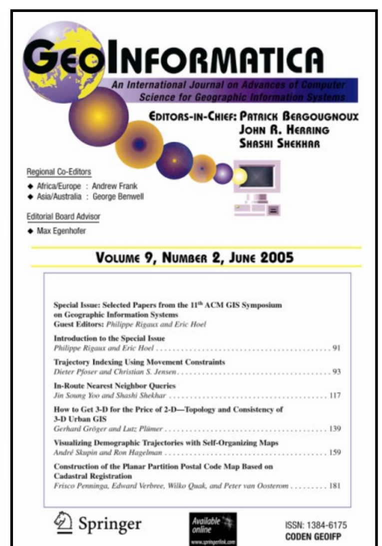
Department of Computer Science and Engineering

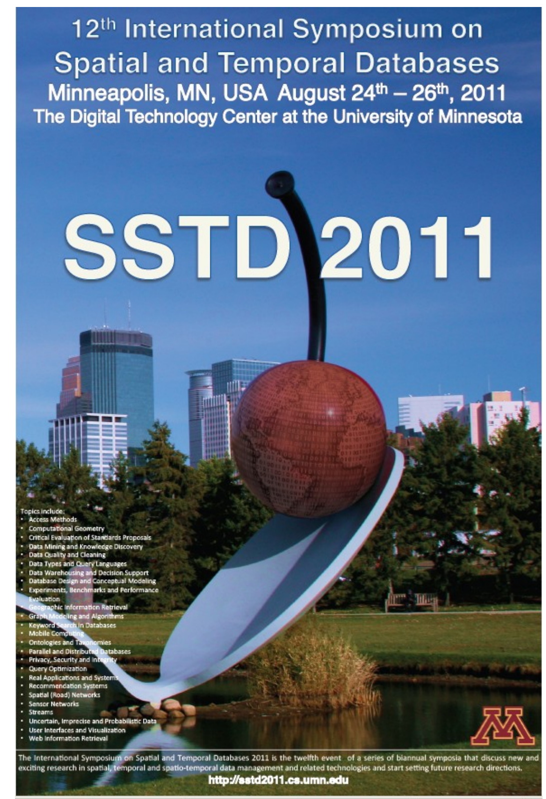
University of Minnesota

www.cs.umn.edu/~shekhar





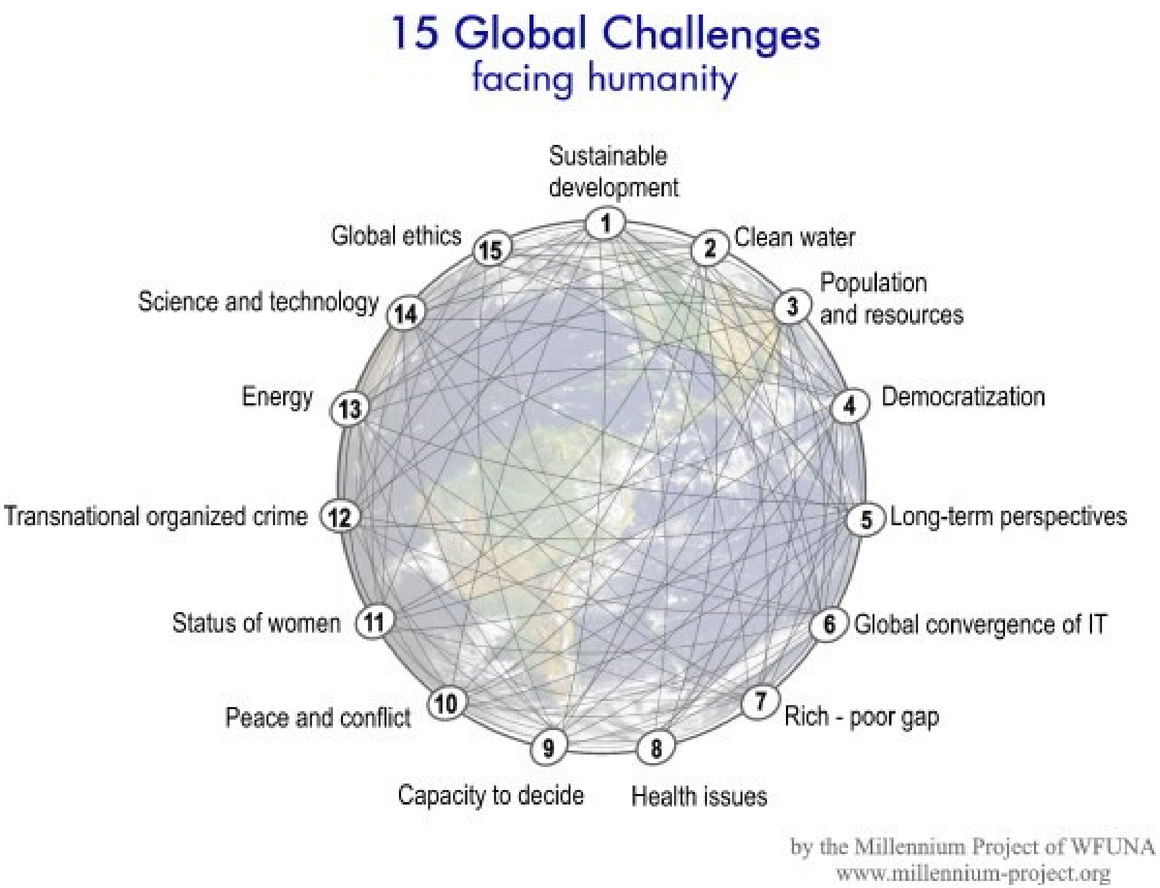




Spatial Computing: Reccent Trends

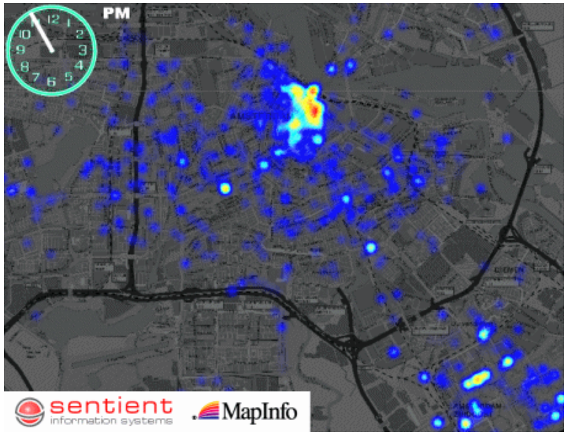
Motivation for Spatial Computing

- Societal:
  - Google Earth, Google Maps, Navigation, location-based service
  - Global Challenges facing humanity – many are geo-spatial!
  - Future of Computer Science (CS) is to address societal challenges!

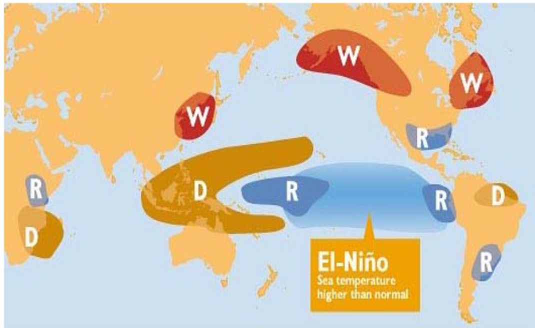


Governement Applications

- Spatial computing
  - NASA Earth Observing System (EOS): Earth science data
  - National Inst. of Justice: crime mapping
  - Census Bureau, Dept. of Commerce: census data
  - Dept. of Transportation (DOT): traffic data
  - National Inst. of Health (NIH): cancer clusters
  - Commerce, e.g. Retail Analysis



- Sample Global Questions from Earth Science
  - How is the global Earth system changing
  - What are the primary forcing of the Earth system
  - How does the Earth system respond to natural and human induced changes
  - What are the consequences of changes in the Earth system for human civilization
  - How well can we predict future changes in the Earth system?



Spatial Computing in Government

Geospatial Information and Geographic Information Systems (GIS): An Overview for Congress

May 18, 2011

Congressional Research Service

Table 1. Members of the Federal Geographic Data Committee (FGDC)

Dept. of Agriculture	Environmental Protection Agency
Dept. of Commerce	Federal Emergency Management Agency
Dept. of Defense	General Services Administration
Dept. of Energy	Library of Congress
Dept. of Health and Human Services	National Aeronautics and Space Administration
Dept. of Housing and Urban Development	National Archives and Records Administration
Dept. of the Interior (Chair)	National Science Foundation
Dept. of Justice	Tennessee Valley Authority
Dept. of State	
Dept. of Transportation	Office of Management and Budget (Co-Chair)

Spatial thinking in Business Analytics

- Experience of an e-commerce pioneer
  - Mid 1990s: Geography is dead in post-internet era.
  - Late 1990s: Logistics and delivery are our biggest challenge!

- Spatial questions are central to many businesses!
  - Where are our customers? Suppliers? Stake-holders? …
  - Where should we do business?
    - Which countries? (Globalization), Which cities? …
  - Where should we locate? – store, warehouse, factories, offices, …
  - Petroleum, Mining – where should one drill / dig ?
  - Where is geographic event, e.g. storm, earth-quake, flood ?
  - What is impact of this event on our organization?
  - …
- Should business analytics address spatial questions?
  - Where is a pattern prevalent ?
  - Context: Patterns involving geographic events, e.g. storm, climate…
  - Ex. Association (diaper, beer)

Economy & Spatial Computing

McKinsey Global Institute

The study estimates that the use of personal location data could save consumers worldwide more than $600 billion annually by 2020. Computers determine users' whereabouts by tracking their mobile devices, like cellphones. The study cites smartphone location services including Foursquare and Loopt, for locating friends, and ones for finding nearby stores and restaurants.

But the biggest single consumer benefit, the study says, is going to come from time and fuel savings from location-based services — tapping into real-time traffic and weather data — that help drivers avoid congestion and suggest alternative routes. The location tracking, McKinsey says, will work either from drivers' mobile phones or GPS systems in cars.

The New York Times

New Ways to Exploit Raw Data May Bring Surge of Innovation, a Study Says

Published: May 13, 2011

Spatial Thinking in Consumer Applications

- Trends:

Consumers account for two-third of US economy
Cell phone outnumber personal computers
Spatial apps dominate the Google android app. contest winner

Research Challenges in Spatial Computing

- Is spatial computing just an application of well-known CSE techniques?
- Are there CSE research challenges and opportunities ?
- Dynamic Programming is a popular algorithm design paradigm
  - Shortest Path Algorithm
  - DBMS Query Optimization
  - Sequence alignment,
  - Viterbi algorithm, …
- However, DP assumes stationary ranking of candidate solutions
  - Is DP appropriate for longitudal problems ?

Spatial Computing Questions

- How do we conceptualize spatio-temporal (ST) worlds?
- How do we measure ST concepts, recognize them in (remotely) sensed information or in the field, and identify their accuracy and quality?
- How do we represent ST concepts with incomplete/ uncertain information, with alternative data models, and possibly with multiple representations for the same data, in digital environments?
- How do we store, access, and transform ST concepts, facilitating data sharing, data transfer, and data archiving, while ensuring minimum information loss?
- How do we explain ST phenomena through the application of appropriate methods of forward or inverse models of physical and human processes?
- How do we visualize ST concepts on a variety of media such as maps on electronic displays or animated displays ?
- How do we use ST concepts to think about spatio-temporal phenomena, and to seek explanations for spatio-temporal patterns and phenomena?
- What ST issues or business organizations? How do we use ST concepts to think about business issues?

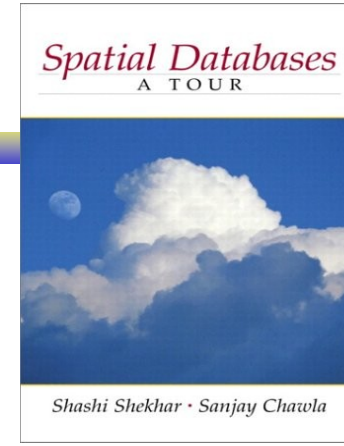
(Source: A daptation from NCGIA proposal to NSF by Goodchild et al.)

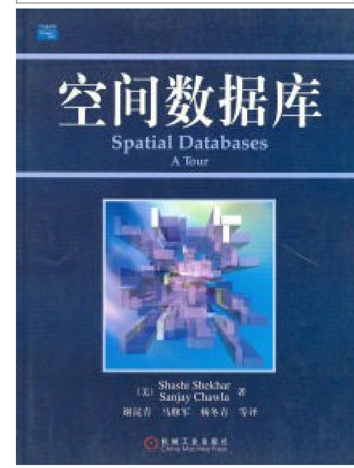
Outline

- Motivation
- Case 1: Infrastructure:
  - Database Management Systems (DBMS)
  - Routing
  - Evacuation route planning
- Case 2: Intelligence: Statistics, Data Mining

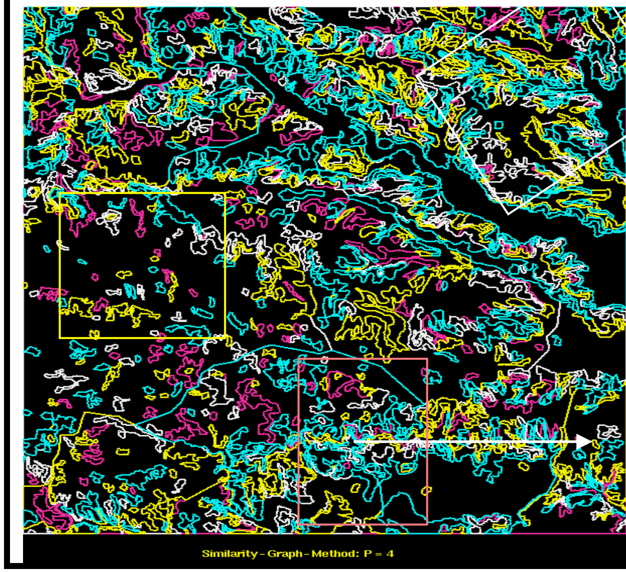
Relational DBMS to Spatial DBMS

- 1980s: Relational DBMS
  - Relational Algebra
  - Query Processing, e.g. sort-merge equi-join algorithm, …
  - B+ Tree index
- Spatial customer (e.g. NASA, USPS) got interested
- But faced challenges
  - Semantic Gap
    - Spatial concepts: distance, direction, overlap, inside, shortest paths, …
    - SQL representation was quite verbose
    - Relational algebra can not represent Transitive closure
  - Performance challenge due to linearity assumption
    - Is B+ tree appropriate for geographic data?
    - Is sorting natural in geographic space?
- New ideas emerged in 1990s
  - Spatial data types and operations (e.g. OGIS Simple Features)
  - R-tree, space partitioning, …





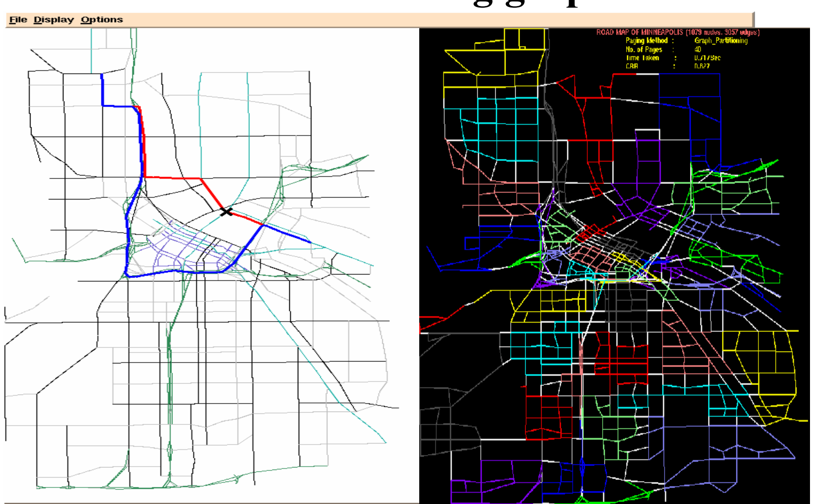
Spatial Databases: Representative Projects



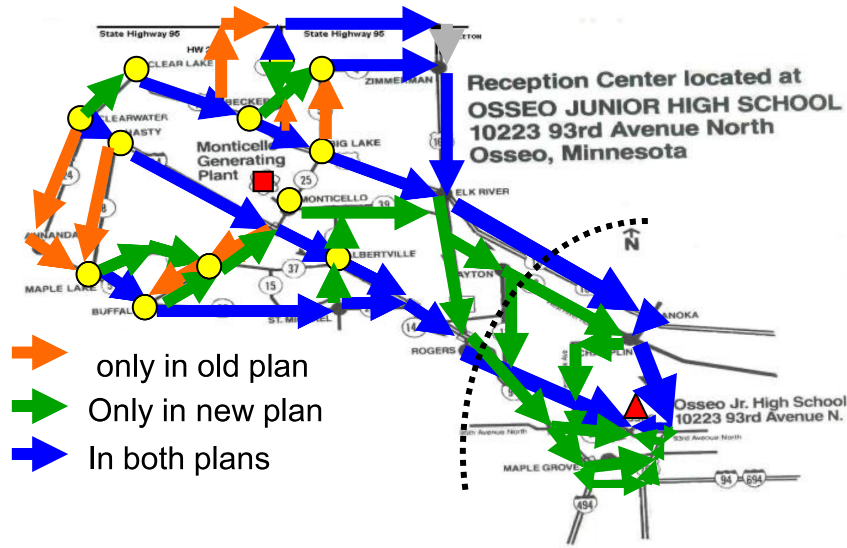
Parallelize Range Queries

Shortest Paths

Storing graphs in disk blocks

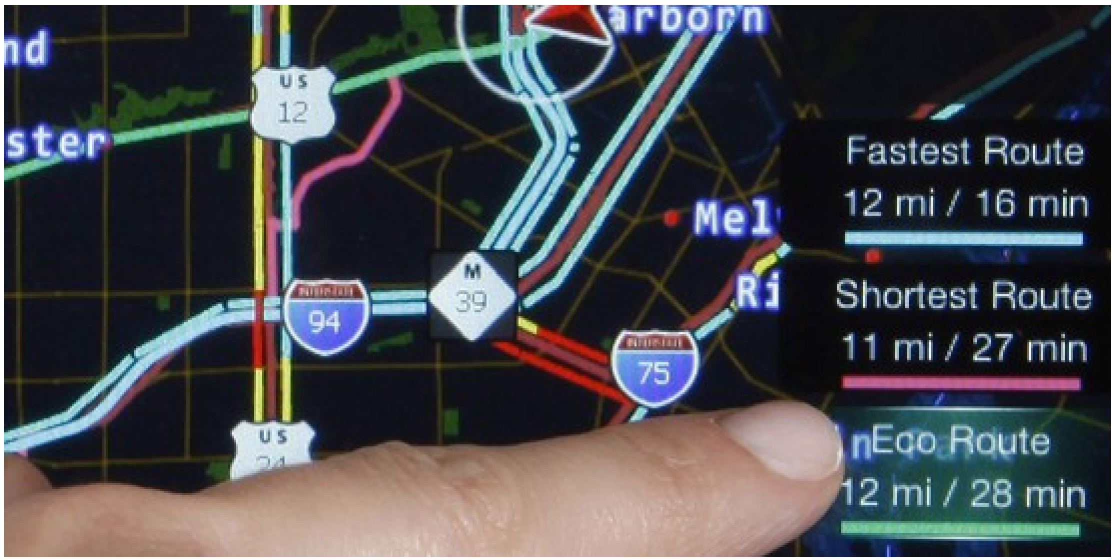


Evacuation Route Planning



Eco-Routing

- Minimize fuel consumption and GPG emission
  - rather than proxies, e.g. distance, travel-time
  - avoid congestion, idling at red-lights, turns and elevation changes, etc.



The New York Times

U.P.S. Embraces High-Tech Delivery Methods (July 12, 2007)

By "The research at U.P.S. is paying off. ....— saving roughly three million gallons of fuel in good part by mapping routes that minimize left turns."

Revisit Shortest Path Problem

The New York Times

U.P.S. Embraces High-Tech Delivery Methods (July 12, 2007)

By "The research at U.P.S. is paying off. ....— saving roughly three million gallons of fuel in good part by mapping routes that minimize left turns."

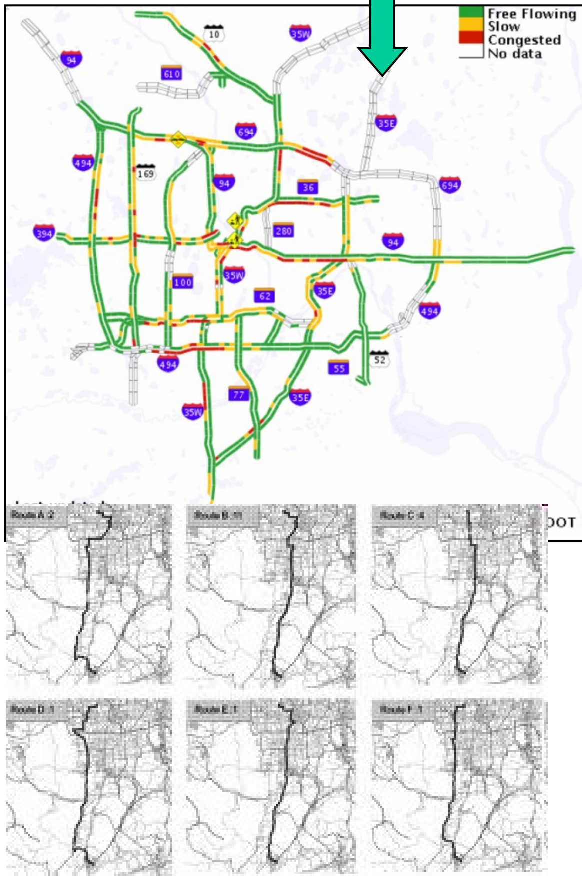
© UPS, 2003

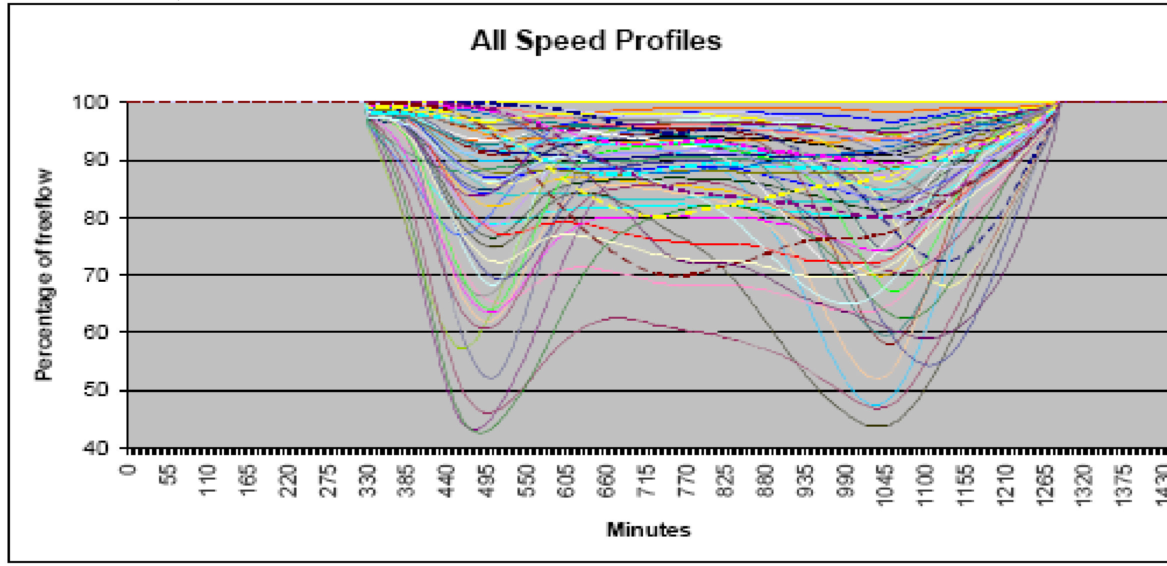
- New Routing Questions
  - Best start time to minimize time spend on network
  - Account for delays at signals, rush hour, etc.

- Time-Variant Flow Network Questions

Static	Time-Variant
Which is the shortest travel time path from downtown Minneapolis to airport?	Which is the shortest travel time path from downtown Minneapolis to airport at different times of a work day?
What is the capacity of Twin-Cities freeway network to evacuate downtown Minneapolis ?	What is the capacity of Twin-Cities freeway network to evacuate downtown Minneapolis at different times in a work day?

Real-time and Historic Travel-time Datasets





FT_DailyHistoricData

EID	Freeflow Speed	Weekday Speed	Weekend Speed	Sun	Mon	Tue	Wed	Thu	Fri	Sat
1	.....	.....	.....	.....	.....	.....	.....	.....	.....	
2	.....	.....	.....	.....	.....	.....	.....	.....	.....	
3	.....	.....	.....	.....	.....	.....	.....	.....	.....	
4	.....	.....	.....	.....	.....	.....	.....	.....	.....	

Historic Daily Speed Profile Table

Speed_0	Speed_1	.....
		.....
		.....

Eco-Routing: Spatial Computing Questions

- What are expected fuel saving from use of GPS devices with static roadmaps?
- What is the value-added by historical traffic and congestion information?
- How much additional value is added by real-time traffic information?
- What are the impacts of following on fuel savings and green house emissions?
  - traffic management systems (e.g. traffic light timing policies),
  - vehicles (e.g. weight, engine size, energy-source),
  - driver behavior (e.g. gentle acceleration/braking)
  - environment (e.g. weather)
- What is computational structure of the Eco-Routing problem?
- Does this problem satisfy the assumptions (e.g. stationary ranking of alternative routes) behind common shortest-path computation algorithms?

Routing in ST Networks

- Predictable Future
  - Stationary — Dijkstra’s, A*….
  - Non-stationary
- Unpredictable Future

Broader Implication of Stationary Assumption

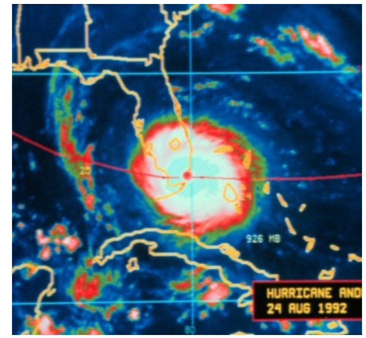
- Dynamic Programming is a popular algorithm design paradigm
  - Shortest Path Algorithm
  - DBMS Query Optimization
  - Sequence alignment,
  - Viterbi algorithm, …
- However, DP assumes stationary ranking of candidate solutions
  - Is DP appropriate for longitudal spatial problems ?

Evacuation Route Planning - Motivation

- No coordination among local plans means
  - Traffic congestions on all highways
  - e.g. 60 mile congestion in Texas (2005)
- Great confusions and chaos

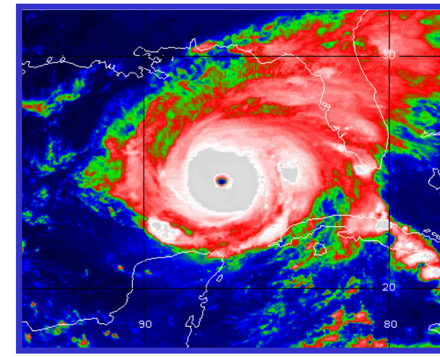
"We packed up Morgan City residents to evacuate in the a.m. on the day that Andrew hit coastal Louisiana, but in early afternoon the majority came back home. **The traffic was so bad that they couldn't get through Lafayette**."
Mayor Tim Mott, Morgan City, Louisiana
(http://i49south.com/hurricane.htm)

Florida, Lousiana (Andrew, 1992)



(National Weather Services)

Houston (Rita, 2005)



(National Weather Services)

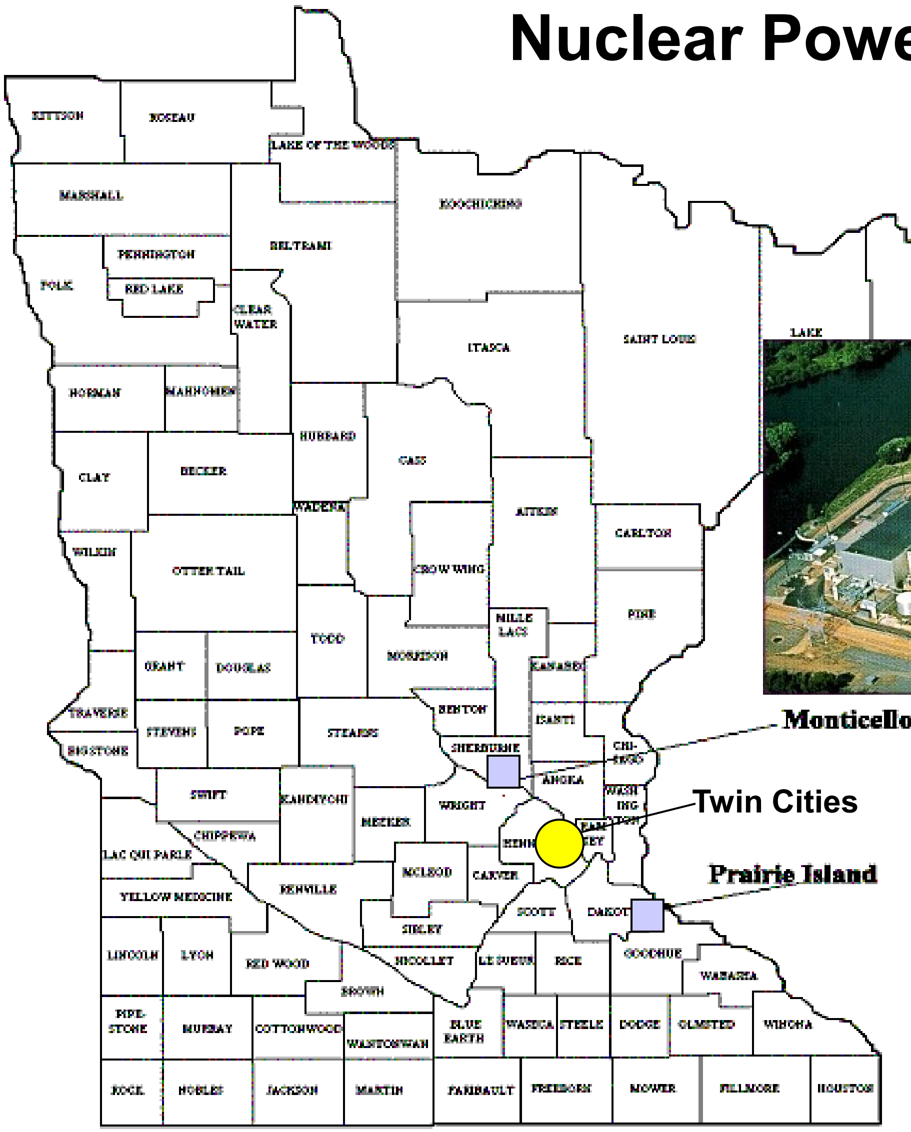
(www.washingtonpost.com)

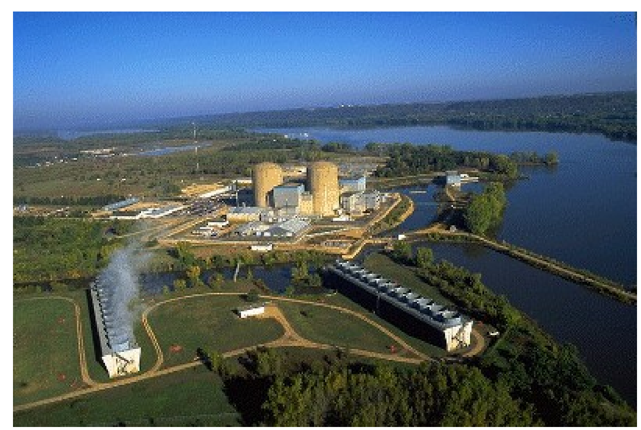
I-45 out of Houston

(FEMA.gov)

A Real Scenario

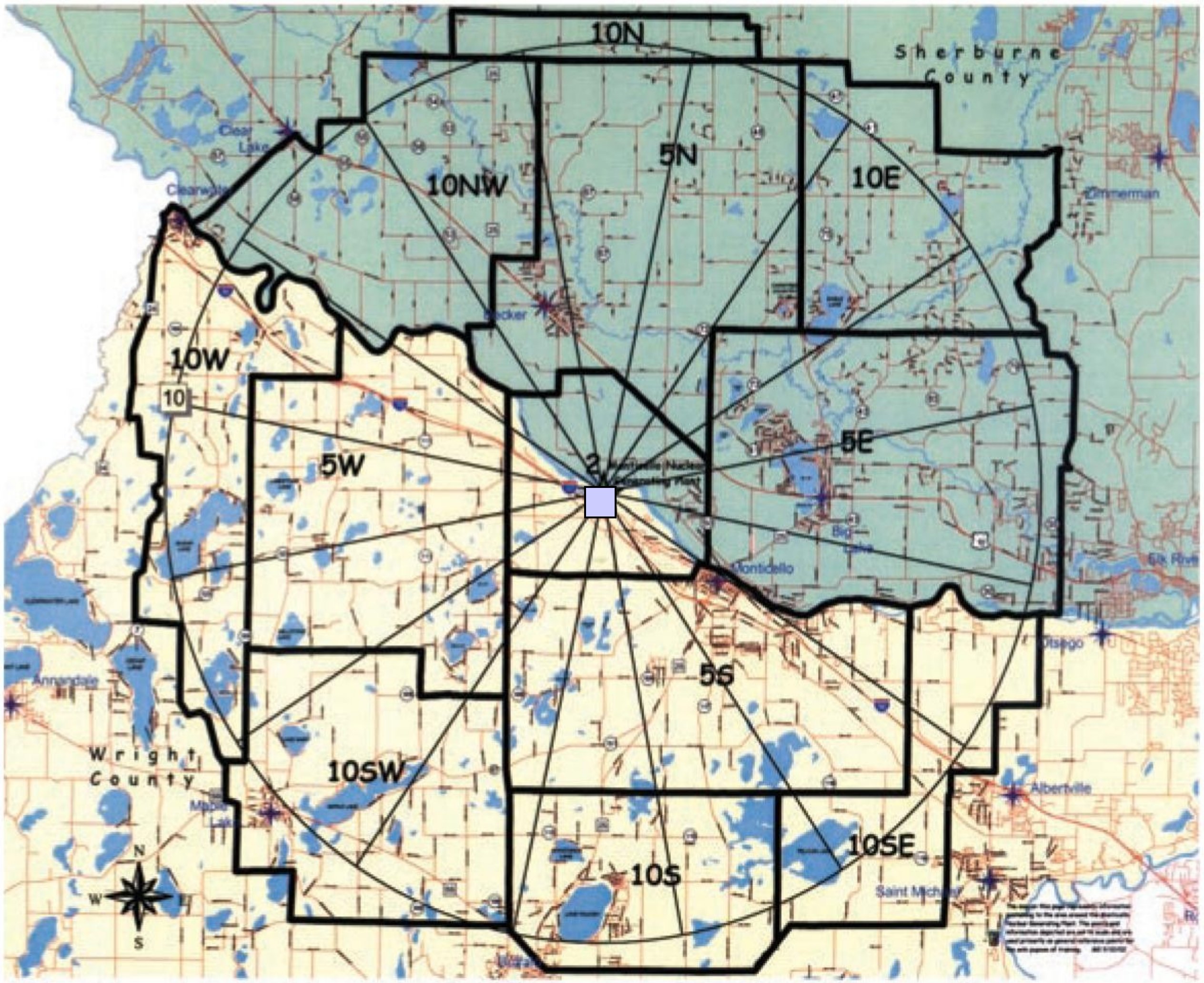
Nuclear Power Plants in Minnesota





Monticello Emergency Planning Zone

Emergency Planning Zone (EPZ) is a 10-mile radius around the plant divided into sub areas.



Monticello EPZ

Subarea	Population
2	4,675
5N	3,994
5E	9,645
5S	6,749
5W	2,236
10N	391
10E	1,785
10SE	1,390
10S	4,616
10SW	3,408
10W	2,354
10NW	707
Total	**41,950**

Estimate EPZ evacuation time:
Summer/Winter (good weather):
3 hours, 30 minutes
Winter (adverse weather):
5 hours, 40 minutes

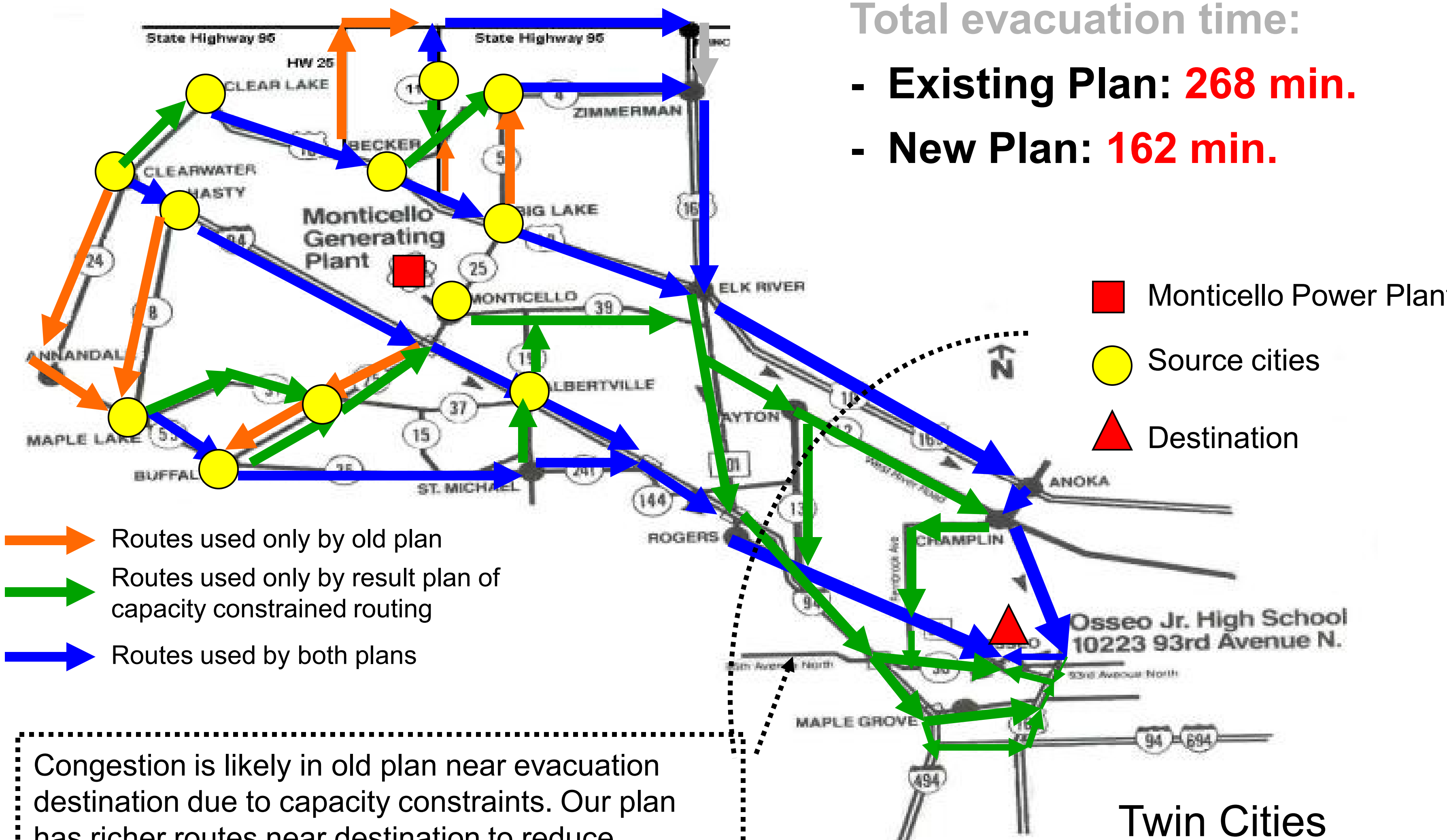
Data source: Minnesota DPS & DHS
Web site: http://www.dps.state.mn.us
http://www.dhs.state.mn.us

A Real World Testcase

Experiment Result

Total evacuation time:

- **Existing Plan: 268 min.**
- **New Plan: 162 min.**



Congestion is likely in old plan near evacuation destination due to capacity constraints. Our plan has richer routes near destination to reduce congestion and total evacuation time.

Problem Statement

Given

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

Output

- Evacuation plan consisting of a set of origin-destination routes
  - and a scheduling of evacuees on each route.

Objective

- Minimize evacuation egress time
  - time from start of evacuation to last evacuee reaching a destination

Constraints

- Route scheduling should observe **capacity constraints** of network
- Reasonable computation time despite limited computer memory
- Capacity constraints and travel times are non-negative integers
- Evacuees start from and end up at nodes

Summary of Related Works & Limitations

A. Capacity-ignorant Approach

- Simple shortest path computation, e.g. A*, Dijktra's, etc.
- e.g. EXIT89 (National Fire Protection Association)

Limitation: Poor solution quality as evacuee population grows

B. Operations Research: Time-Expanded Graph + Linear Programming

- 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. Transportation Science: Dynamic Traffic Assignment

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

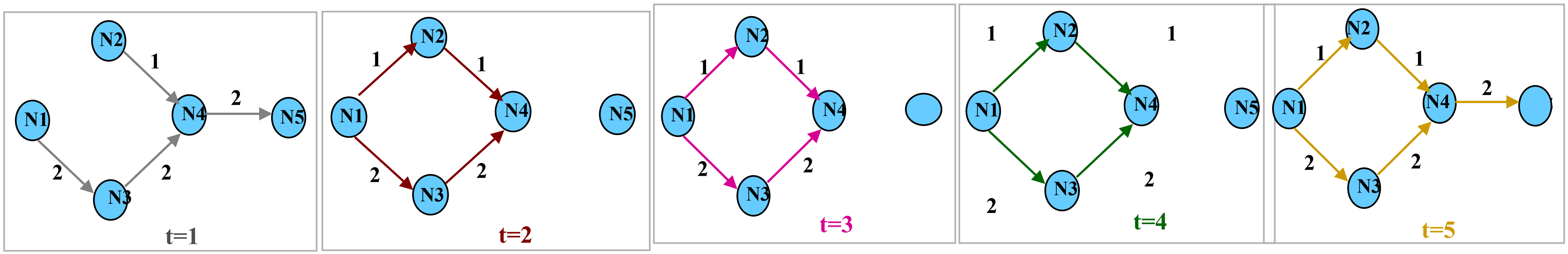
Limitation: Extremely high compute time

- Is Evacuation an equilibrium phenomena?

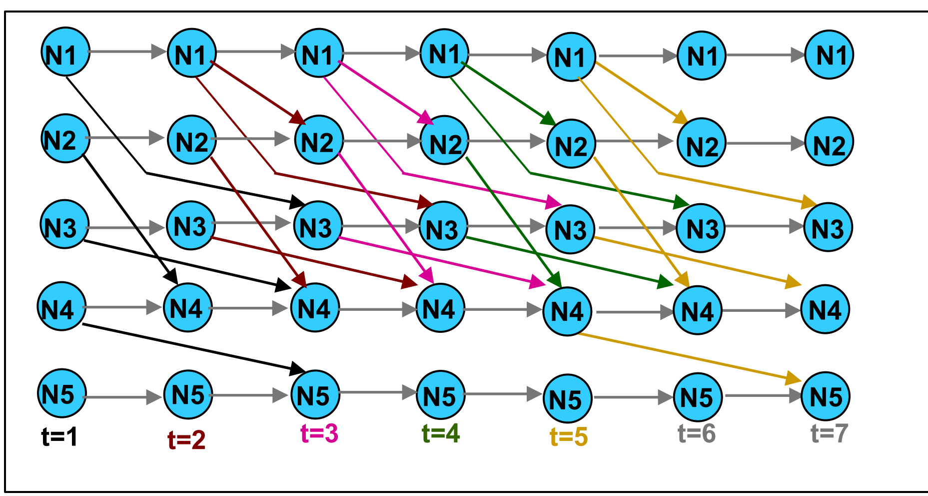
Representations of (Spatio-)temporal Networks

(1) Snapshot Model [Guting 04]

Node: N.. Edge: Travel time →



(2) Time Expanded Graph (TEG) [Ford 65]

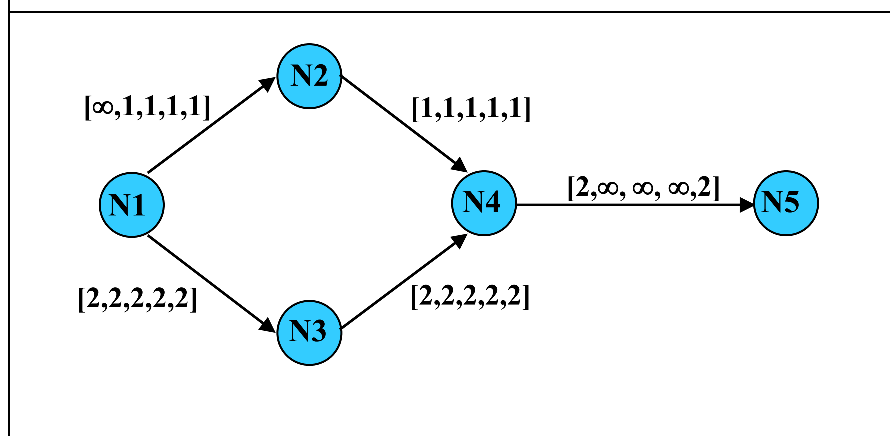


Holdover Edge

Transfer Edges

(3) Time Aggregated Graph (TAG) [Our Approach]

- Attributes aggregated over edges and nodes.



Edge $\xrightarrow{[m_1, \ldots, (m_T]}$ m_i- travel time at t=i

Performance Evaluation

Setup: fixed number of evacuees = 5000, fixed number of source nodes = 10 nodes,
number of nodes from 50 to 50,000.

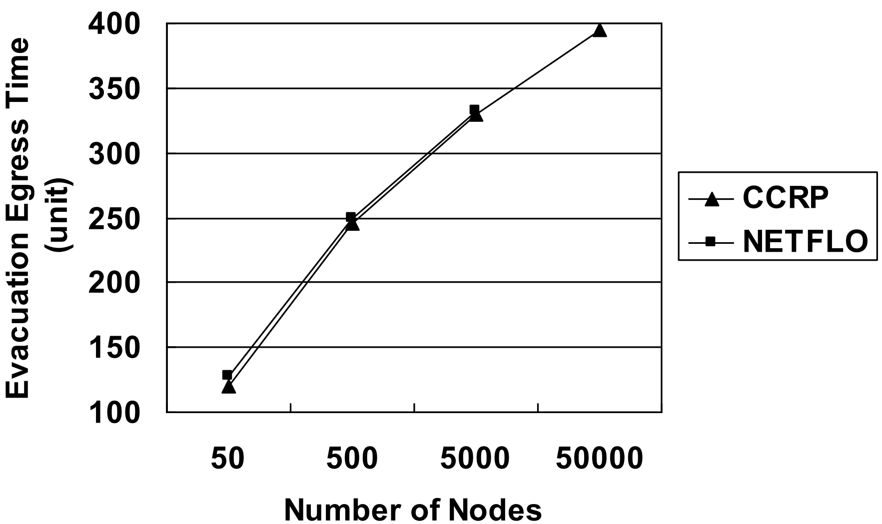
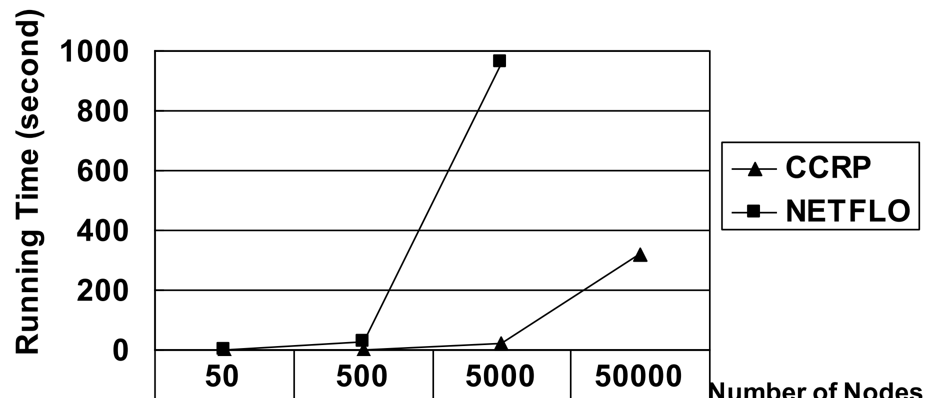


Figure 1 Quality of solution



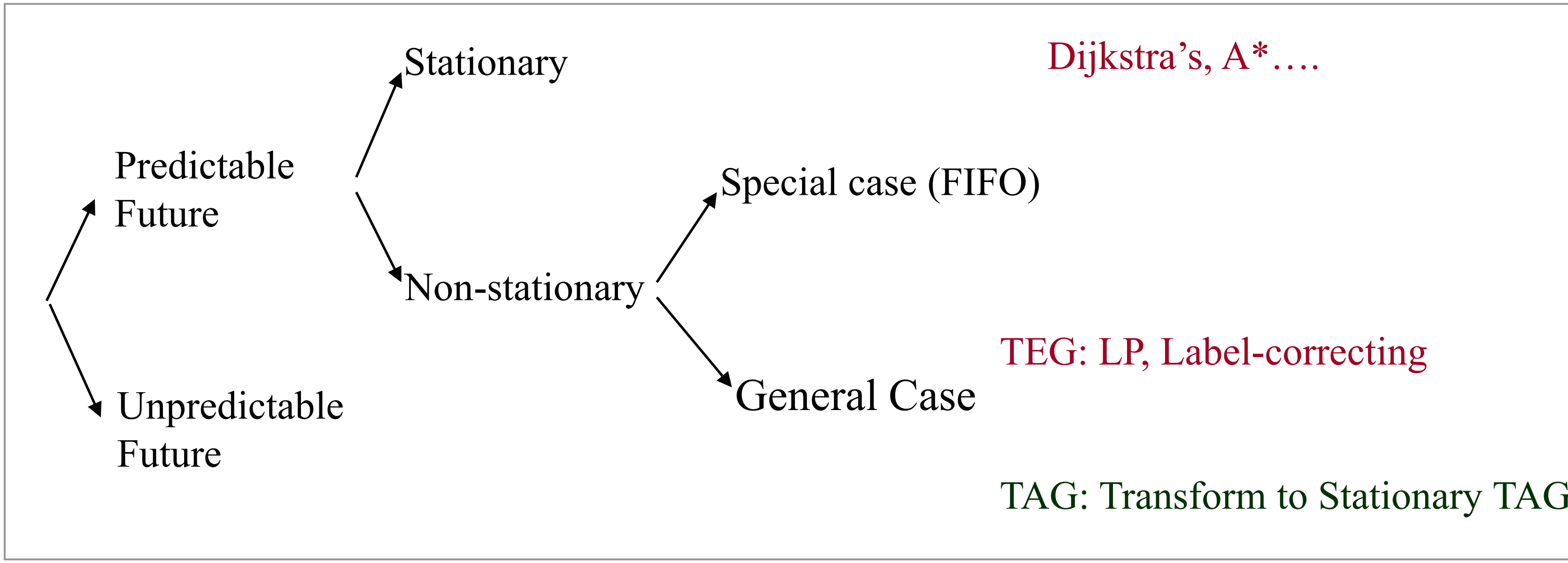
	50	500	5000	50000
CCRP	0.1	1.5	23.1	316.4
NETFLO	0.3	25.5	962.2	

Number of Nodes

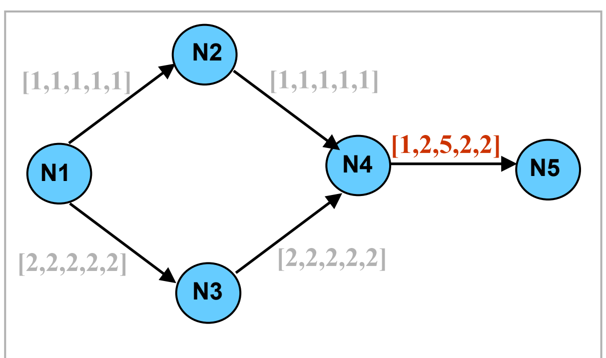
Figure 2 Run-time

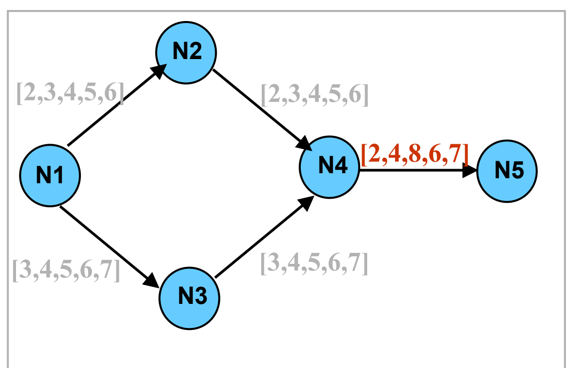
- CCRP produces high quality solution, solution quality increases as network size grows.
- Run-time of CCRP is scalable to network size.

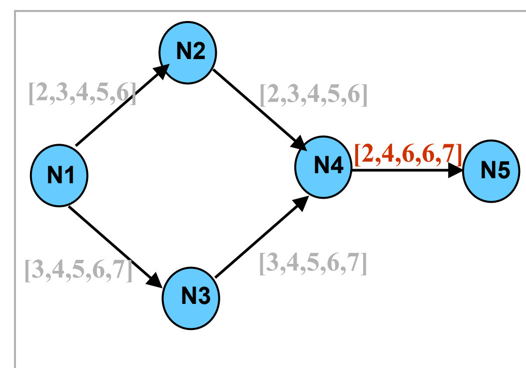
Routing in ST Networks: Scalable Methods



travel times → arrival times at end node → Min. arrival time series







Non-stationary TAG → Stationary TAG

Outline

- Motivation
- Case 1: Infrastructure:
- Case 2: Intelligence
    - Data Mining
    - Statistics

Case 2: Data Mining (DM) to Spatial DM

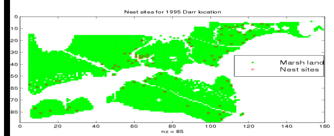
- 1990s: Data Mining
  - Scale up traditional models to large databases
    - Linear regression, Decision Trees, …
  - New pattern families
    - Association rules
    - Which items are bought together? E.g. (Diaper, beer)
- Spatial customers
  - Walmart
    - Which items are bought before/after events, e.g. hurricanes?
    - Where is (diaper-beer) pattern prevalent?
  - Global climate change

- But faced challenges
  - Independence Assumption
  - Transactions,
    - disjoint partitioning of data

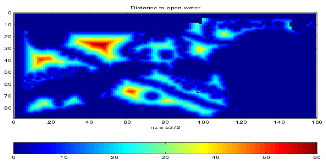
Spatial Data Mining : Representative Projects

Location prediction: nesting sites

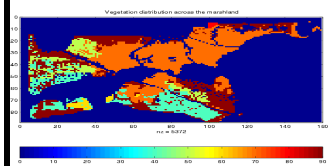
Nest locations



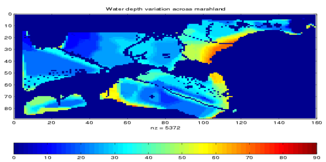
Distance to open water



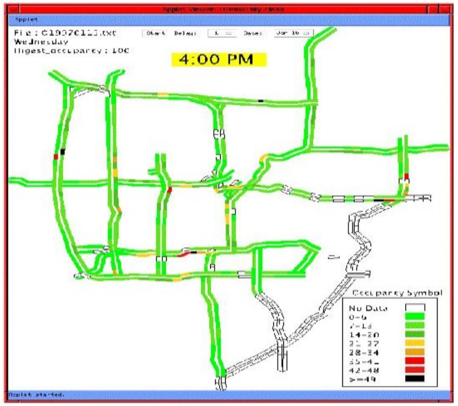
Vegetation durability

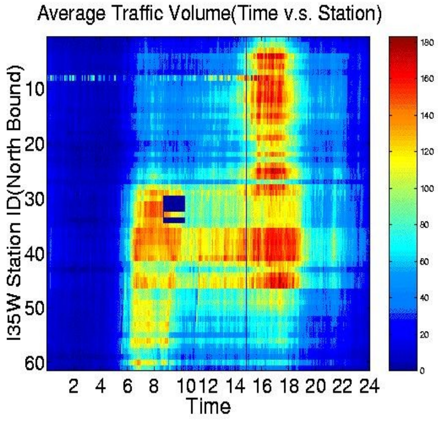


Water depth

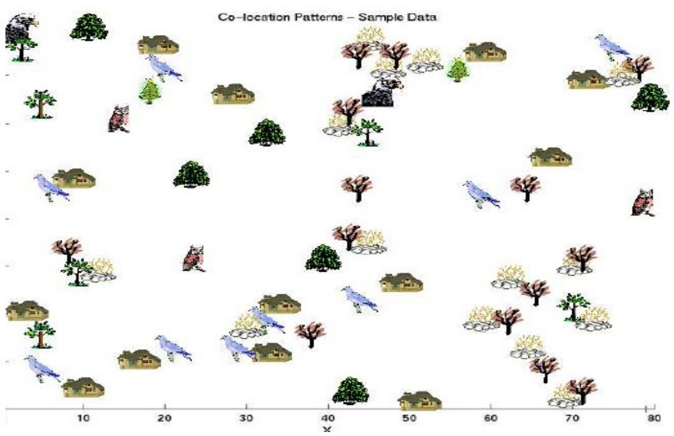


Spatial outliers: sensor (#9) on I-35

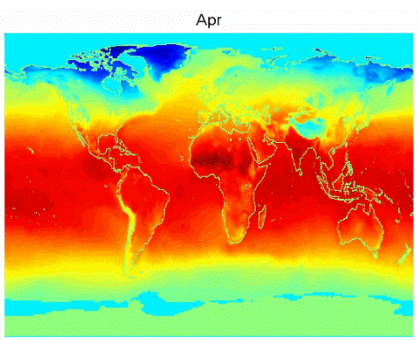


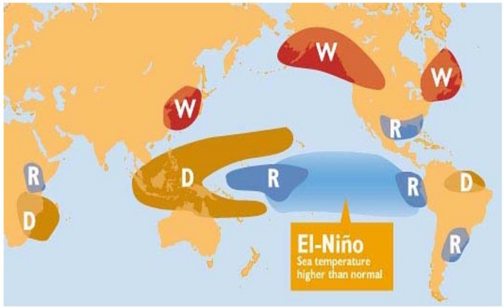


Co-location Patterns



Tele connections





Association Patterns

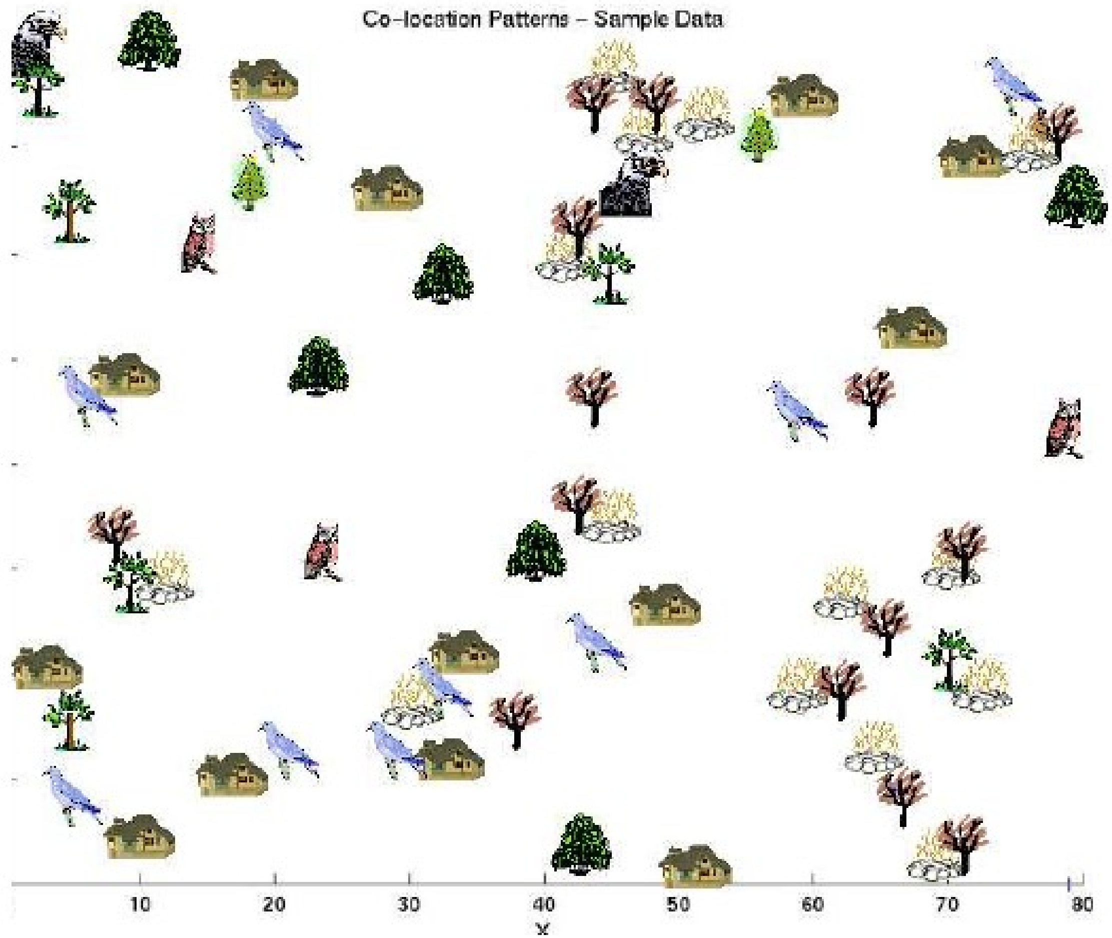
- Association rule e.g. (Diaper in T => Beer in T)

Transaction	Items Bought
1	{socks,  milk,  beef, egg, …}
2	{pillow,  toothbrush, ice-cream, muffin, …}
3	{  , , pacifier, formula, blanket, …}
…	…
n	{battery, juice, beef, egg, chicken, …}

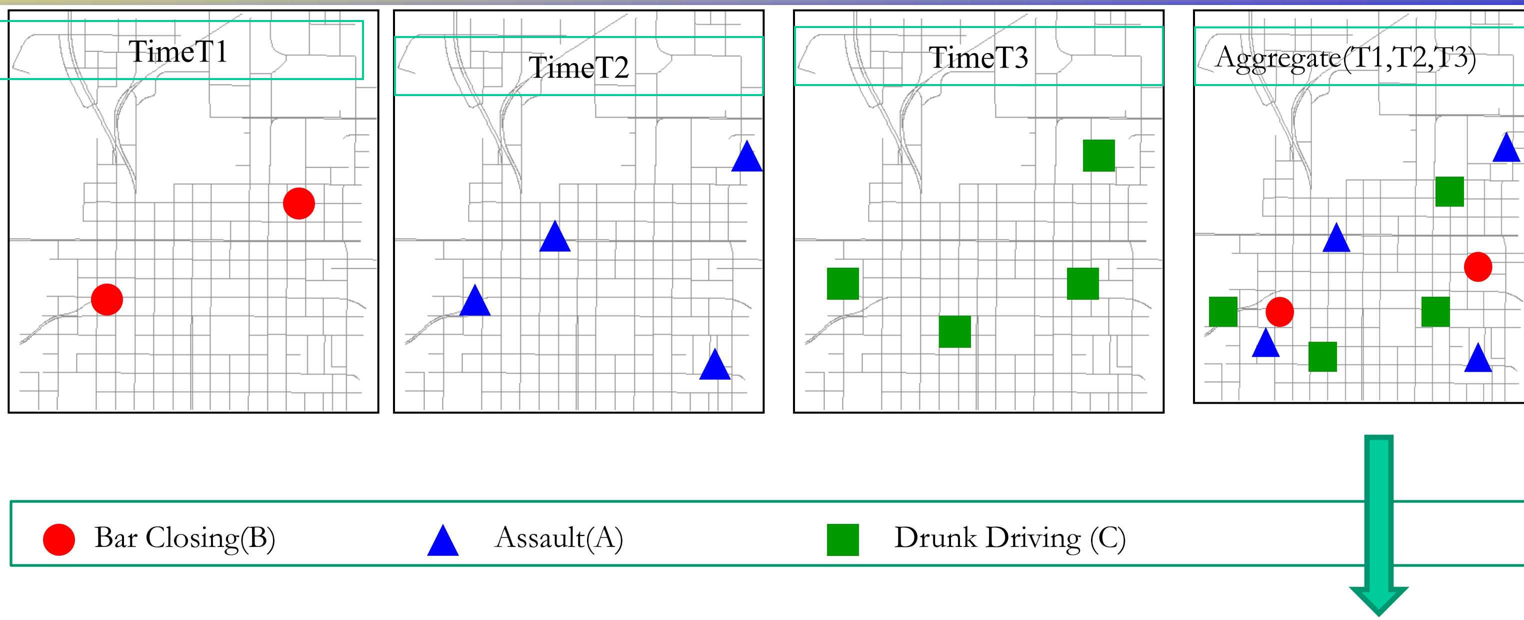
  - Support: probability (Diaper and Beer in T) = 2/5
  - Confidence: probability (Beer in T \| Diaper in T) = 2/2
- Algorithm Apriori [Agarwal, Srikant, VLDB94]
  - Support based pruning using monotonicity
- Note: **Transaction is a core concept!**

Co-locations/Co-occurrence

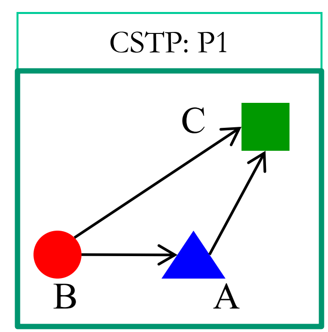
- Given: A collection of different types of spatial events
- Find: Co-located subsets of event types



Cascading spatio-temporal pattern (CSTP)

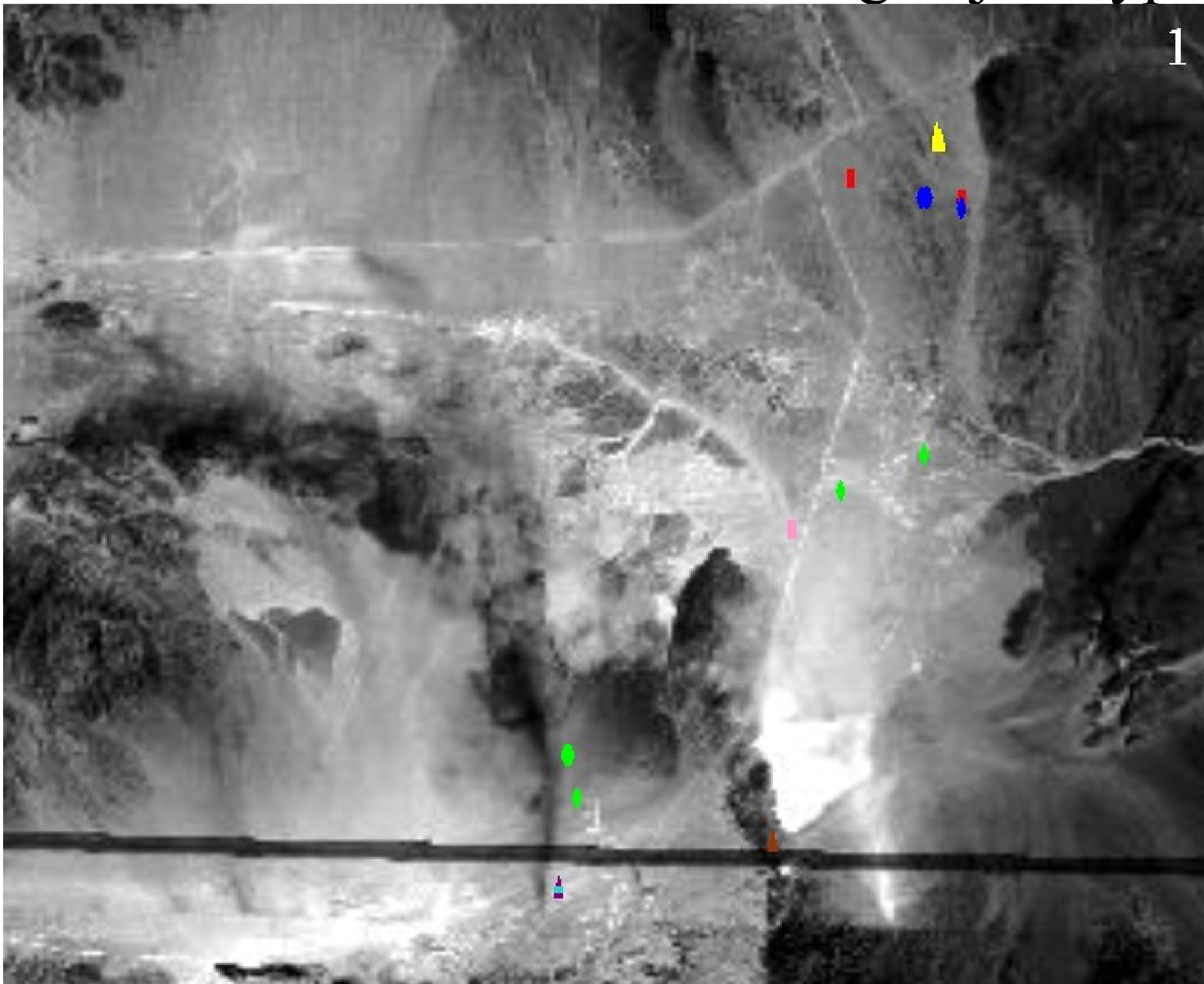


- *Input:* Urban Activity Reports
- *Output: CSTP*
  - *Partially ordered* subsets of ST event types.
  - Located together in space.
  - Occur in *stages* over time.



- Applications: Epidemiology, Disaster Response, …

Co-occurrence of moving object-types!



Manpack stinger
(2 Objects)

M1A1_tank
(3 Objects)

M2_IFV
(3 Objects)

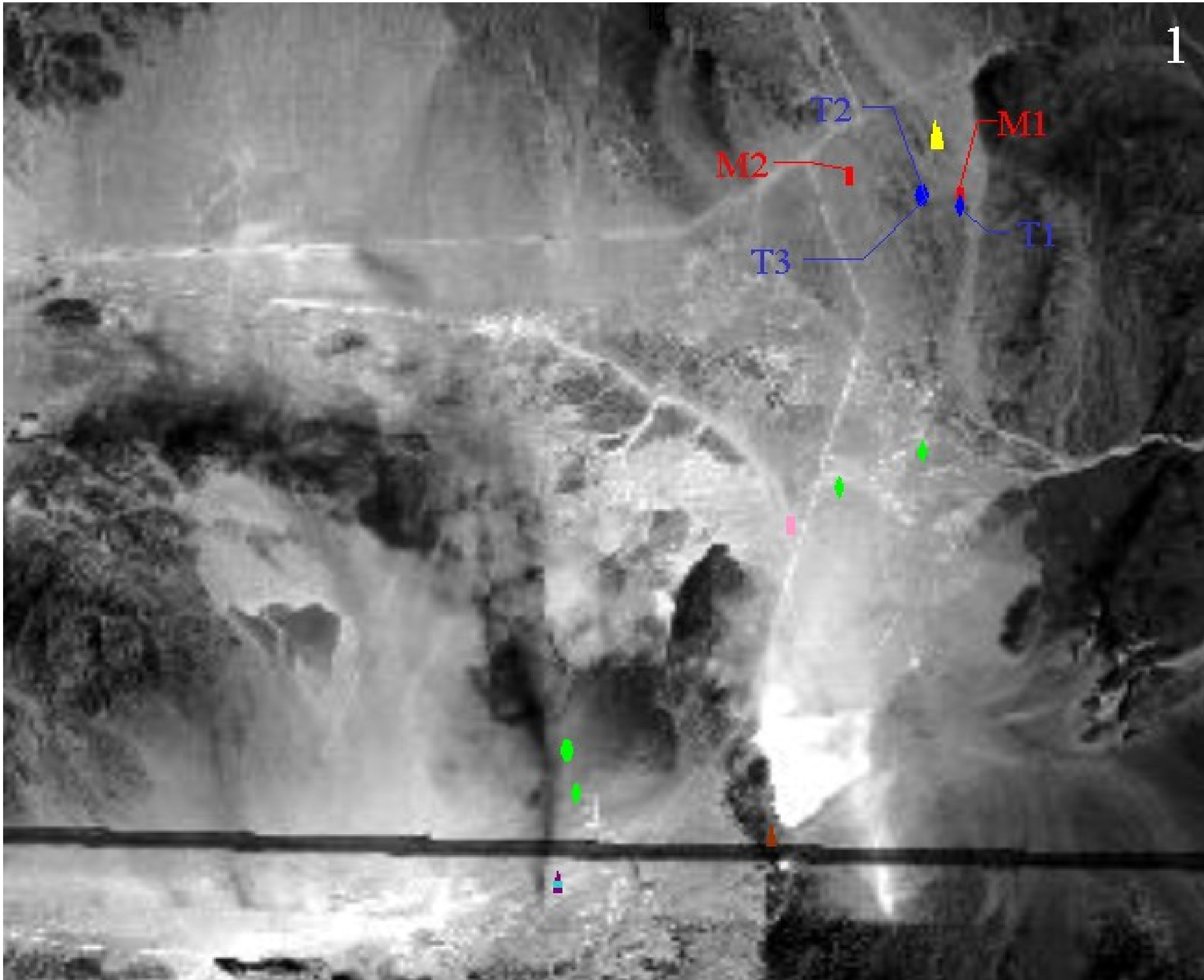
Field_Marker
(6 Objects)

T80_tank
(2 Objects)

BRDM_AT5
(enemy) (1 Object)

BMP1
(1 Object)

Co-occurring moving object-types



Manpack stinger
(2 Objects)

M1A1_tank
(3 Objects)

M2_IFV
(3 Objects)

Field_Marker
(6 Objects)

T80_tank
(2 Objects)

BRDM_AT5
(enemy) (1 Object)

BMP1
(1 Object)

Co-location: A Neighborhood based Approach

	Association rules	Colocation rules
underlying space	discrete sets	continuous space
item-types	item-types	events /Boolean spatial features
collections	Transactions	neighborhoods
prevalence measure	support	participation index
conditional probability measure	Pr.[A in T | B in T]	Pr.[A in N(L) | B at L]

Challenges:

1. Computational Scalability

   Needs a large number of spatial join, 1 per candidate colocation

2. Spatio-temporal Semantics

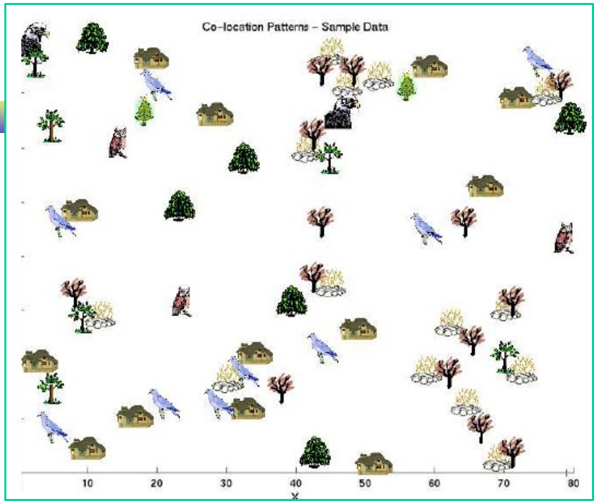
   Spatio-tempotal co-occurrences

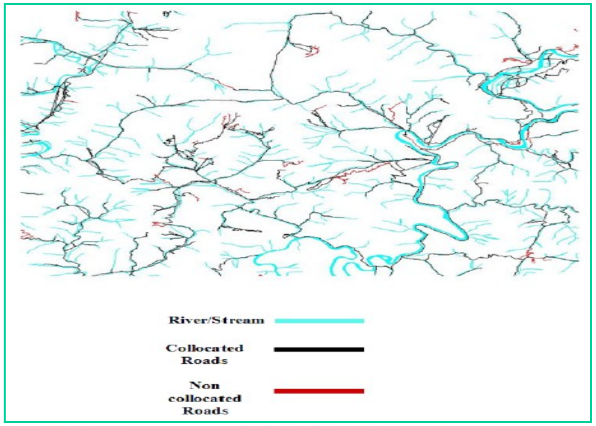
   Emerging colocations

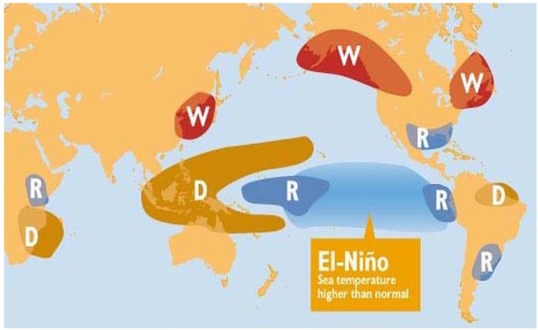
   …

Colocation, Co-occurrence, Interaction

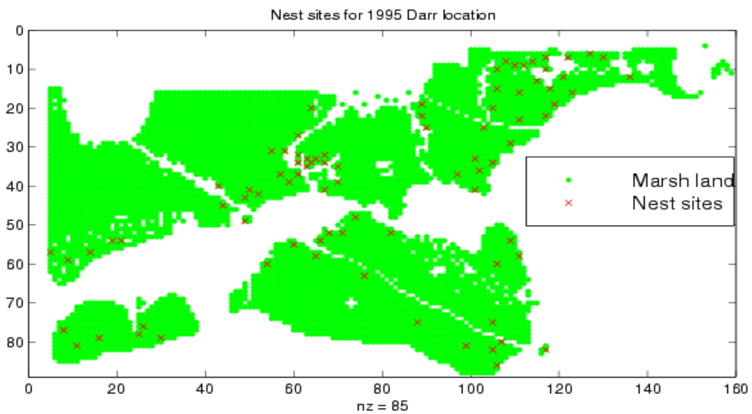
- What is it?
  - Subset of event types, whose instances occur together
  - Ex. Symbiosis, (bar, misdemeanors), …
- Solved
  - Colocation of point event-types
- Almost solved
  - Co-location of extended (e.g.linear) objects
  - Object-types that move together
- Failed
  - Neighbor-unaware Transaction based approaches
- Missing
  - Consideration of flow, richer interactions
- Next
  - Spatio-temporal interactions, e.g. item-types that sell well before or after a hurricane
  - Tele-connections



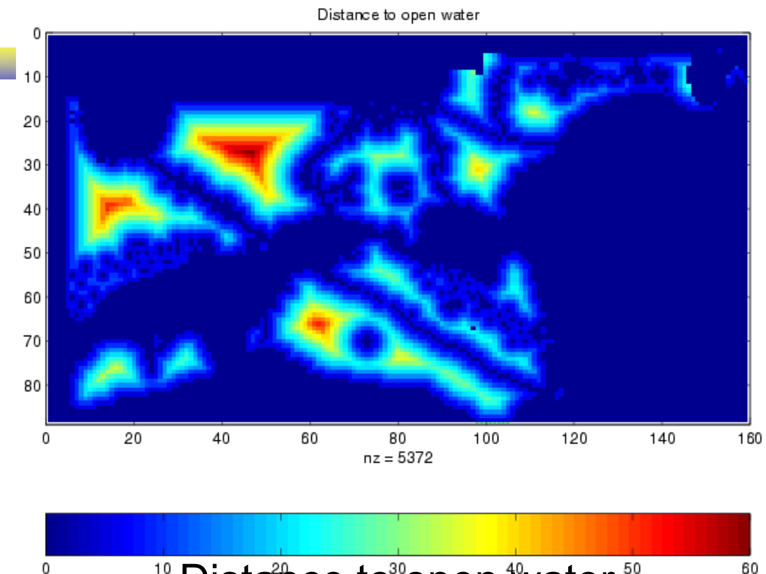




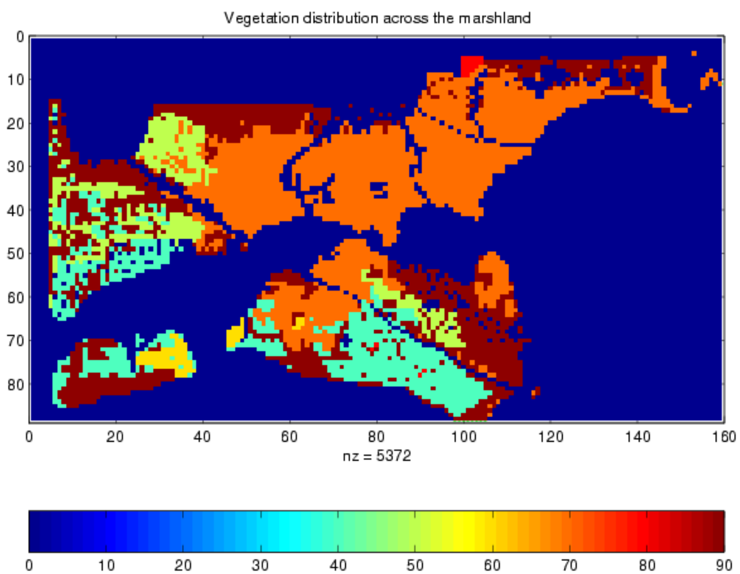
Spatial Prediction



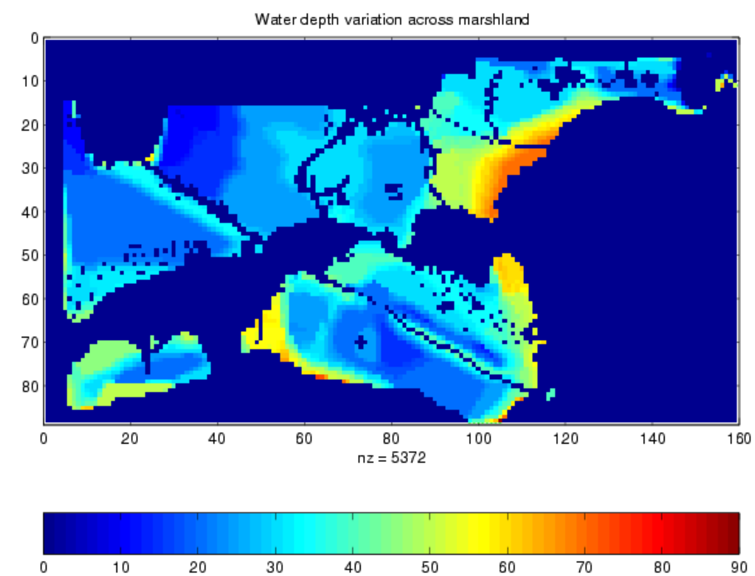
Nest locations



Distance to open water



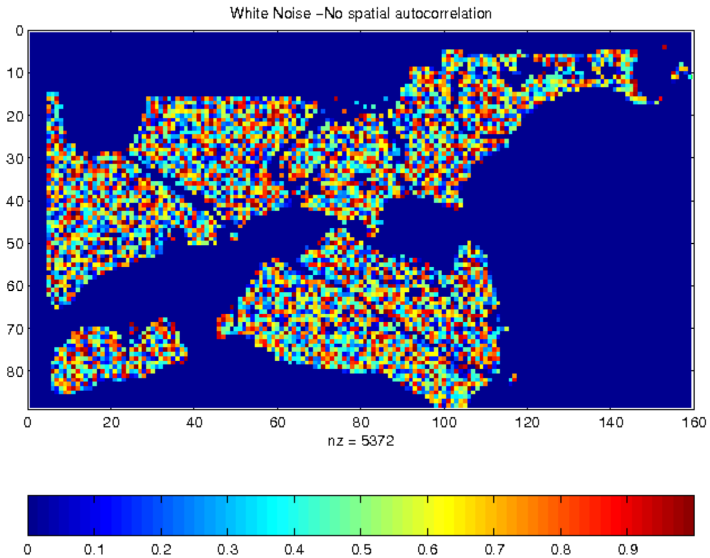
Vegetation durability



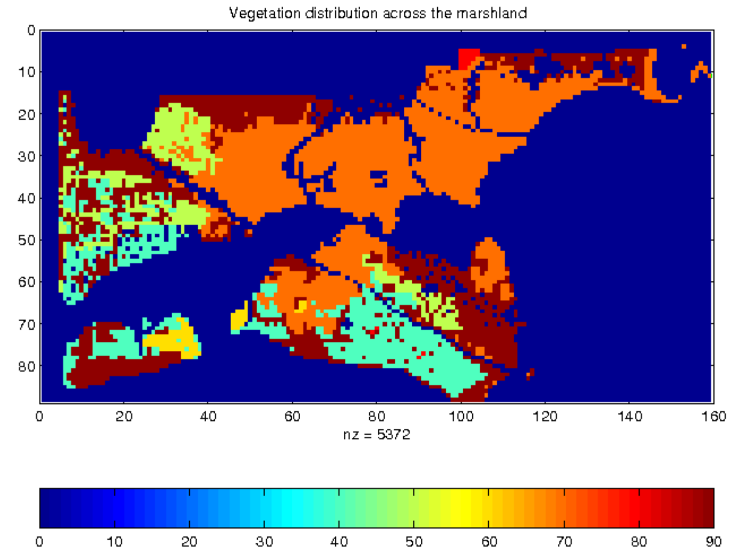
Water depth

Autocorrelation

- First Law of Geography
  - "All things are related, but nearby things are more related than distant things. [Tobler, 1970]"



Pixel property with independent identical distribution



Vegetation Durability with SA

- Autocorrelation
  - Traditional i.i.d. assumption is not valid
  - Measures: K-function, Moran's I, Variogram, …

Implication of Auto-correlation

Name	*Model*	*Classification Accuracy*
Classical Linear Regression	$\mathbf{y} = \mathbf{x}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$	Low
Spatial Auto-Regression	$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{x}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$	High

ρ : the spatial auto-regression (auto-correlation) parameter

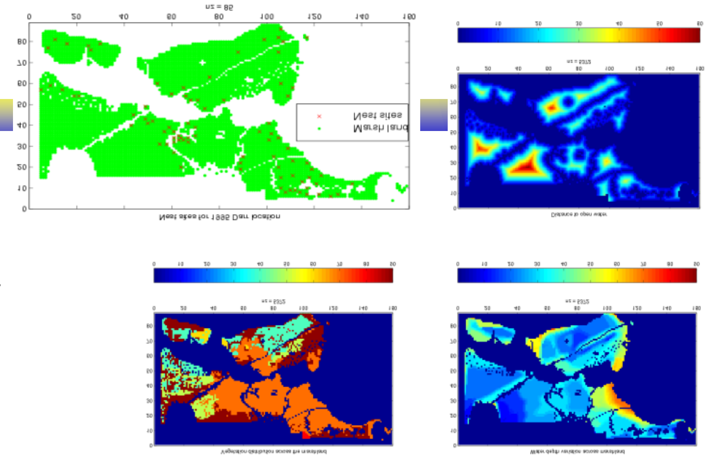
$\mathbf{W}$: n-by-n neighborhood matrix over spatial framework

Computational Challenge:

Computing determinant of a very large matrix in the Maximum Likelihood Function:

$$\ln(L) = \ln\left|\mathbf{I} - \rho\mathbf{W}\right| - \frac{n\ln(2\pi)}{2} - \frac{n\ln(\sigma^2)}{2} - SSE$$

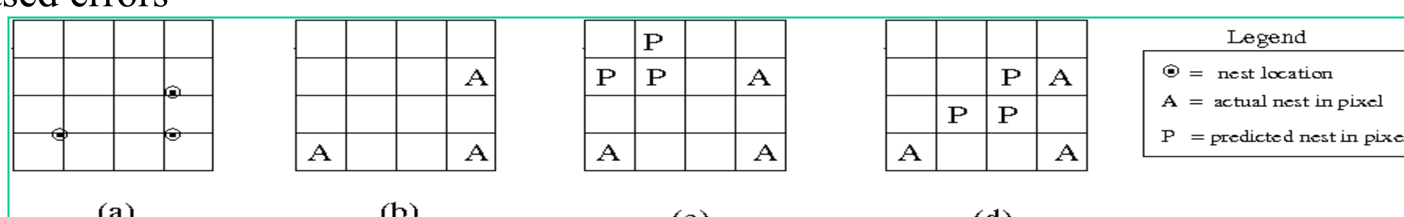
Space/Time Prediction



- What is it?
  - Models to predict location, time, path, …
    - Nest sites, minerals, earthquakes, tornadoes, …
- Solved
  - Interpolation, e.g. Krigging
  - Heterogeneity, e.g. geo. weighted regression
- Almost solved
  - Auto-correlation, e.g. spatial auto-regression
- Failed: Independence assumption
  - Models, e.g. Decision trees, linear regression, …
  - Measures, e.g. total square error, precision, recall
- Missing
  - Spatio-temporal vector fields (e.g. flows, motion), physics
- Next
  - Scalable algorithms for parameter estimation
  - Distance based errors

$$\mathbf{y} = \rho \mathbf{W} \mathbf{y} + \mathbf{x}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\ln(L) = \ln\left|\mathbf{I} - \rho \mathbf{W}\right| - \frac{n \ln(2\pi)}{2} - \frac{n \ln(\sigma^2)}{2} - SSE$$



Summary

- Spatial Computing is critical to many societal grand challenges
  - Sustainable development , Environment, Energy, Water, Public Safety …

- Time is ripe for broader participation from CSE!

  ACM Special Interest Group : SIG Spatial

- Challenges many CSE assumptions
  - Linearity assumption in relational DBMS
    - B+ tree, Sort-merge equi-join, …
  - Stationary assumption behind Dynamic Programming
    - Shortest Path problem
    - DBMS query optimization (Selinger style)
  - Independence assumption in Statistics, Machine Learning, …
    - Decision trees, Linear Regression, …

- Many disciplines are addressing spatial challenges
  - Spatial Statistics, Spatial Economics, Environmental Epidemiology
  - Is it time for greater CSE participation?

Spatial Thinking Across Disciplines!