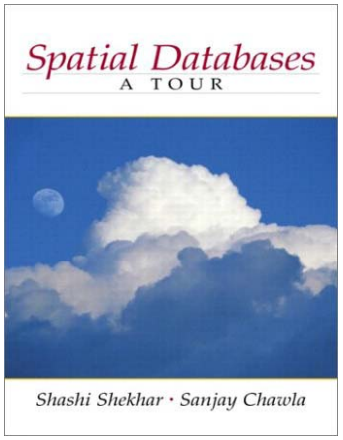
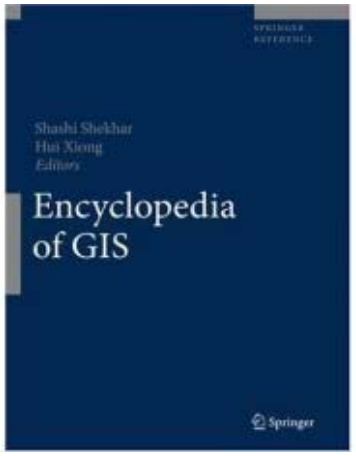
Shashi Shekhar

Affiliation: Faculty of Computer Sc. and Eng., Univ. of Minnesota
www.cs.umn.edu/~shekhar

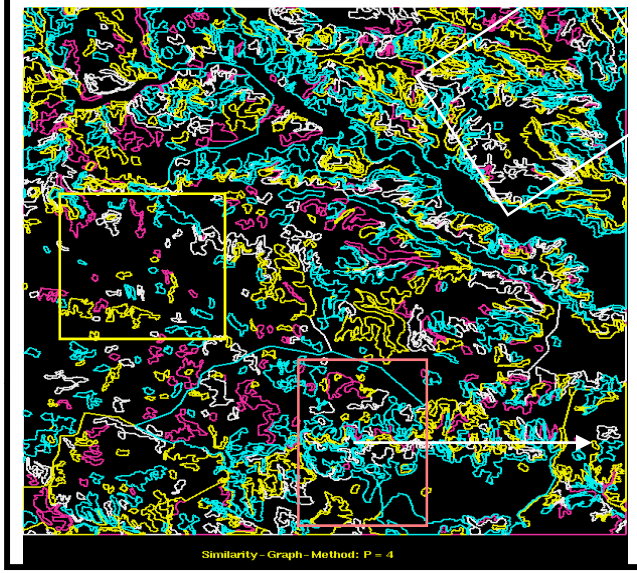
Collaboration Profile:

- Role : Build novel data analysis **tools**
- To facilitate science, engineering or medicine
- Near future opportunities
  - NSF Cyber-driven Discovery and Innovation, NSF/CISE/IIS, …
- Research Focus: Spatial Database, Spatial Data Mining



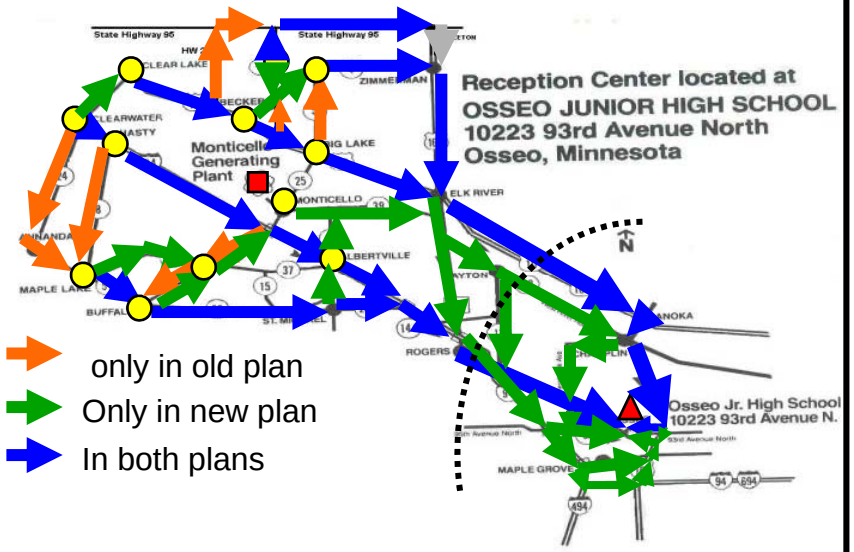


Spatial Databases: Example Projects

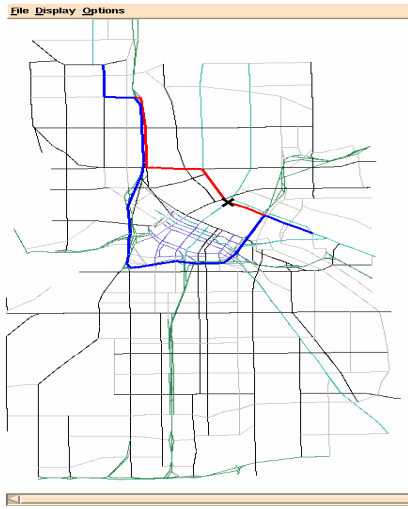


Parallelize Range Queries

Evacuation Route Planning



Shortest Paths

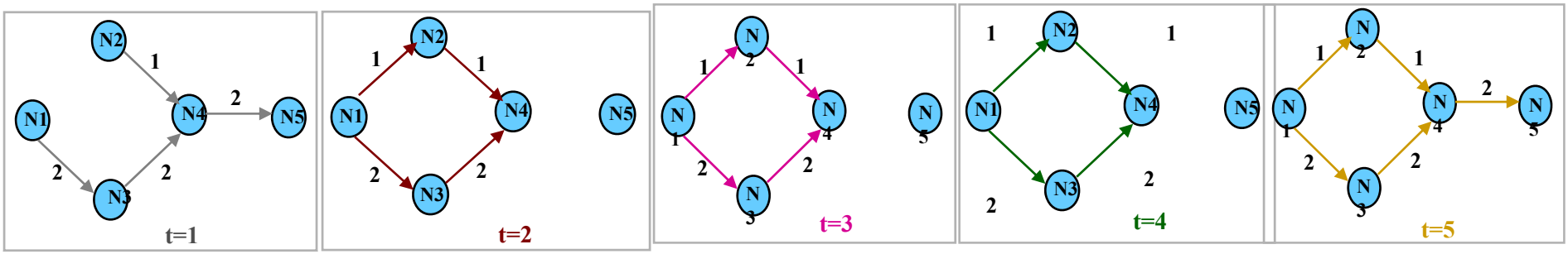


Storing graphs in disk blocks

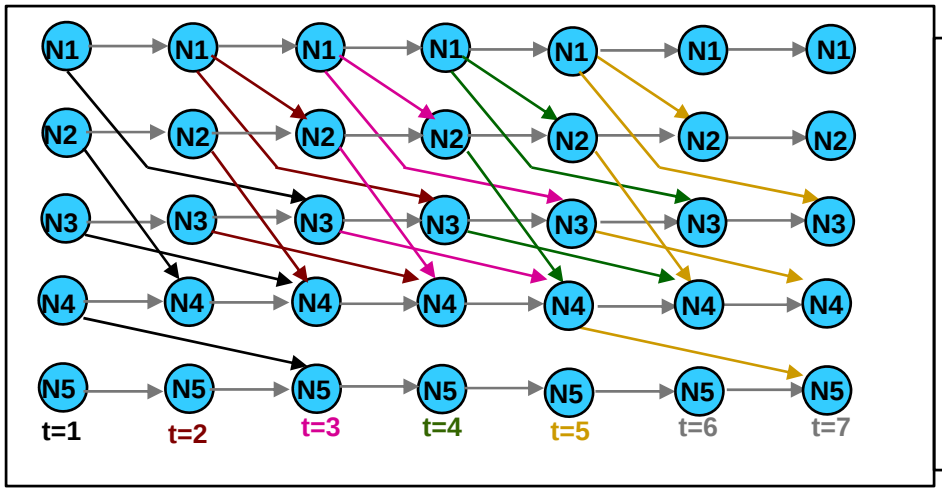
Secret Sauce: Representation of (Spatio-)temporal Networks

(1) Snapshot Model [Guting 04]

Node: N.. Edge: Travel time →



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

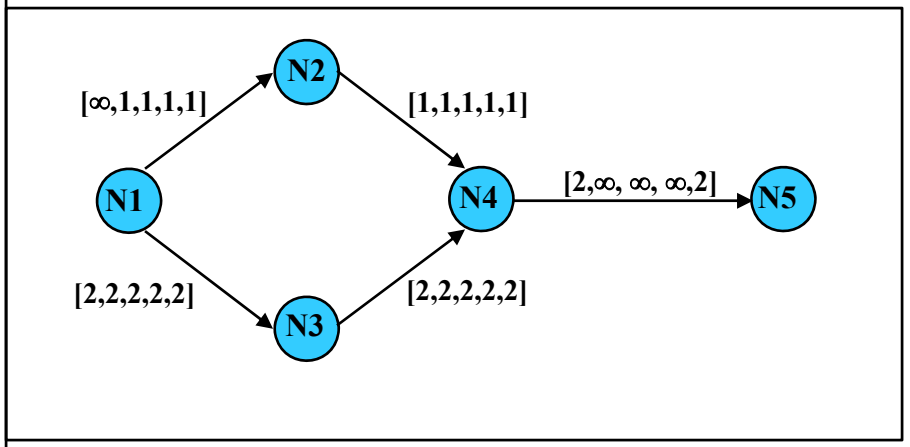


Holdover Edge

Transfer Edges

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

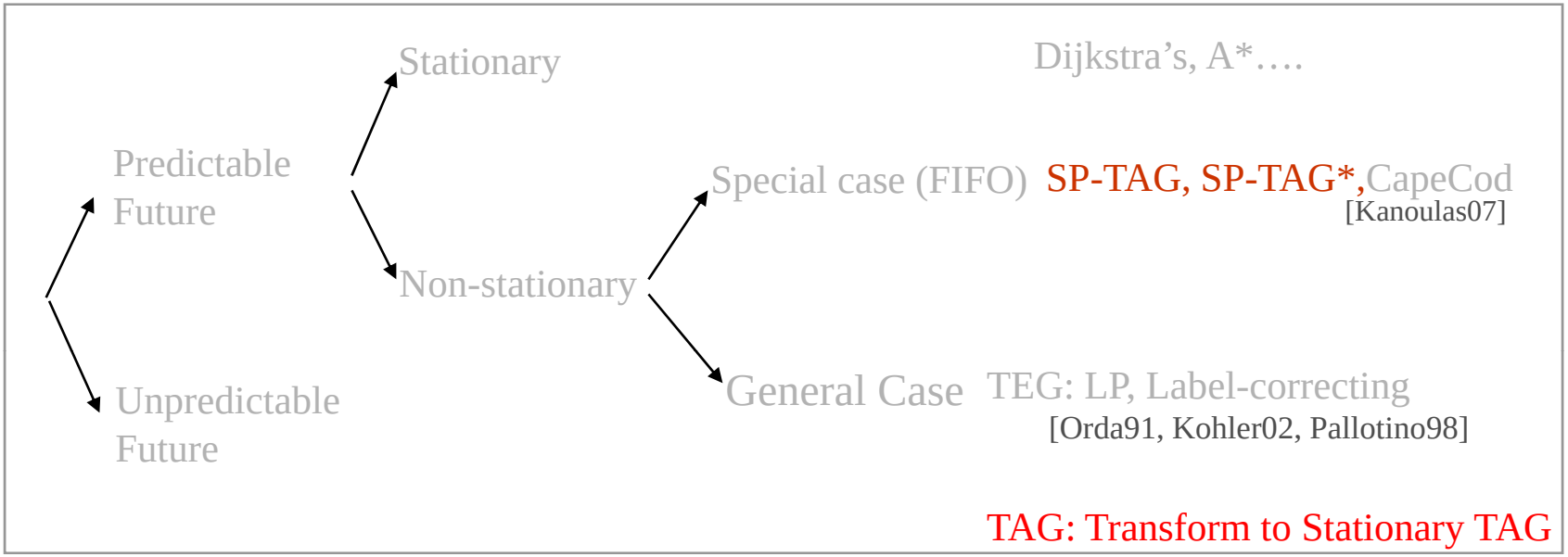
- Attributes aggregated over edges and nodes.



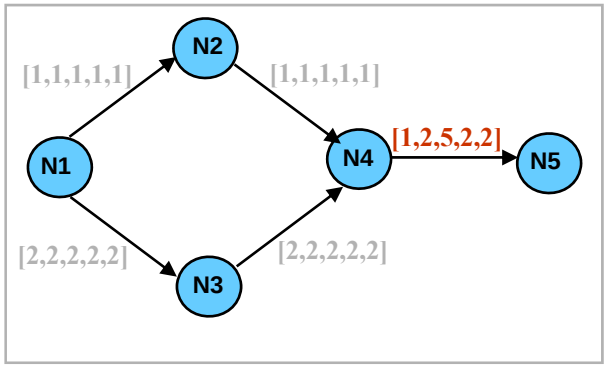
Edge → $[m_1,\ldots,(m_T]$

m_i- travel time at t=i

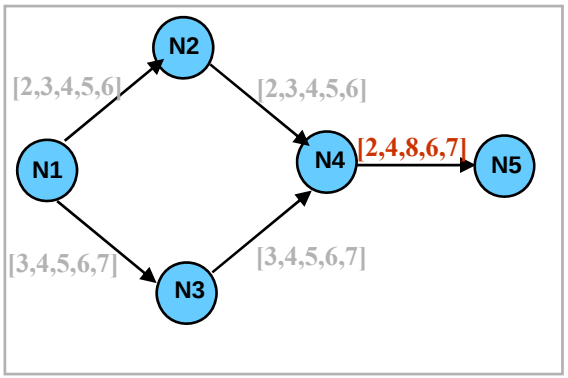
Power of Representation: Ex. Routing Algorithms



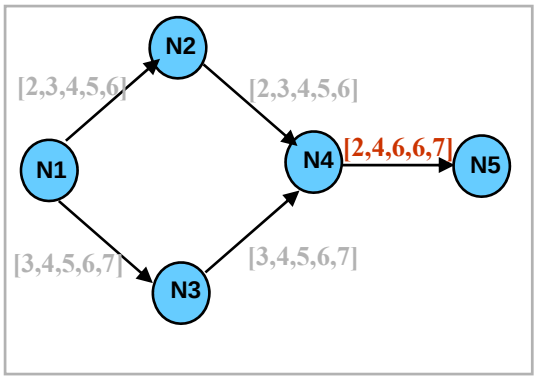
travel times → arrival times at end node → Min. arrival time series



Non-stationary TAG

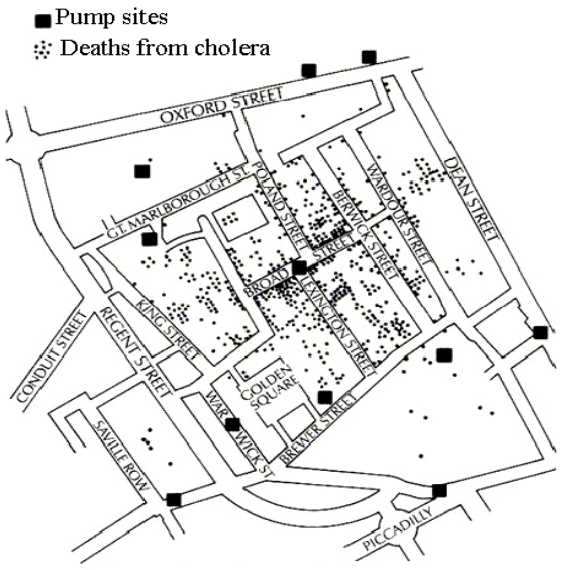


→



Stationary TAG

Spatial and Spatio-temporal Data Mining



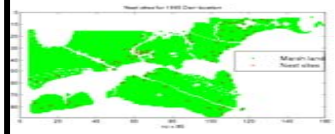
- What is it?
  - Identifying interesting, useful, non-trivial patterns
    - Hot-spots, discontinuities, co-locations, trends, …
  - in large spatial or spatio-temporal datasets
    - Satellite imagery, geo-referenced data, e.g. census
    - gps-tracks, geo-sensor network, …

- Why is it important ?
  - Potential of discoveries and insights to improve human lives
    - Environment: How is Earth system changing? Consequences for humans?
    - Public safety: Where are hotspots of crime? Why?
    - Public health: Where are cancer clusters? Environmental reasons?
    - Transportation, National Security, …
  - However, (d/dt) (Spatial Data Volume) >> (d/dt) (Number of Human Analysts)
    - Need automated methods to mine patterns from spatial data
    - Need tools to amplify human capabilities to analyze spatial data

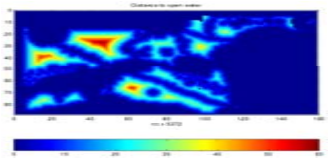
Spatial Data Mining: Example Projects

Location prediction: nesting sites

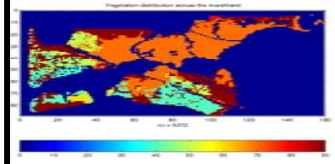
Nest locations



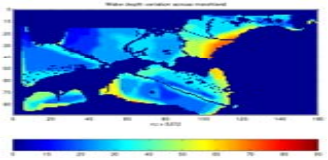
Distance to open water



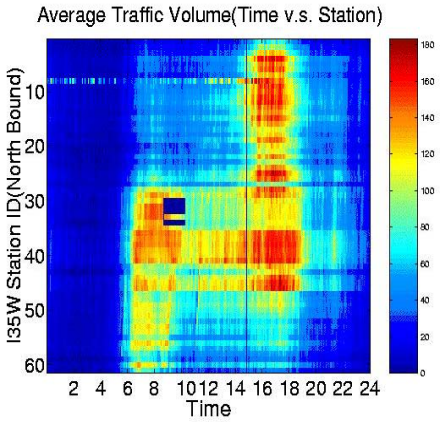
Vegetation durability



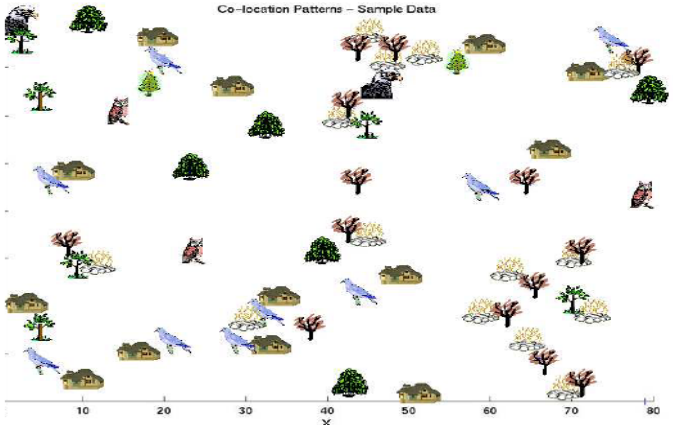
Water depth



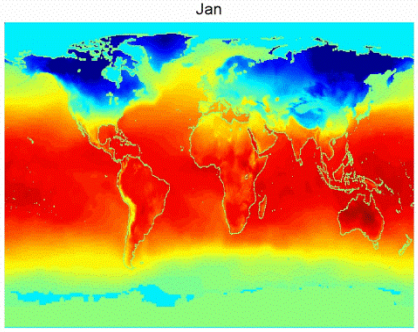
Spatial outliers: sensor (#9) on I-35

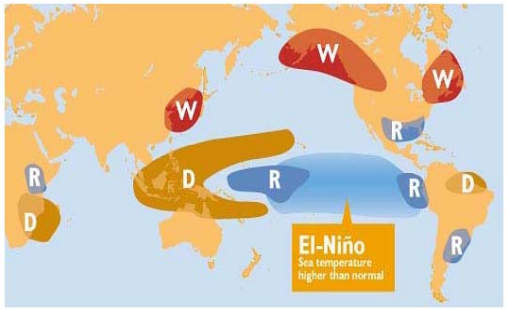


Co-location Patterns



Tele connections

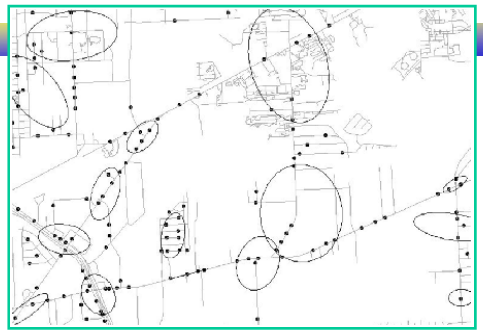


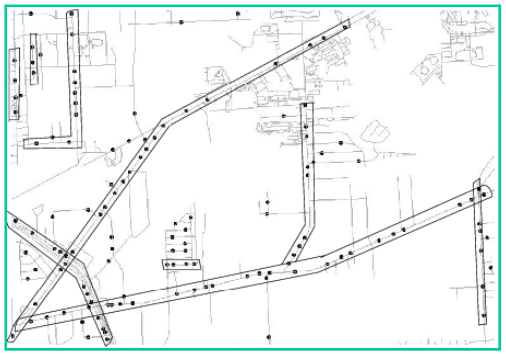


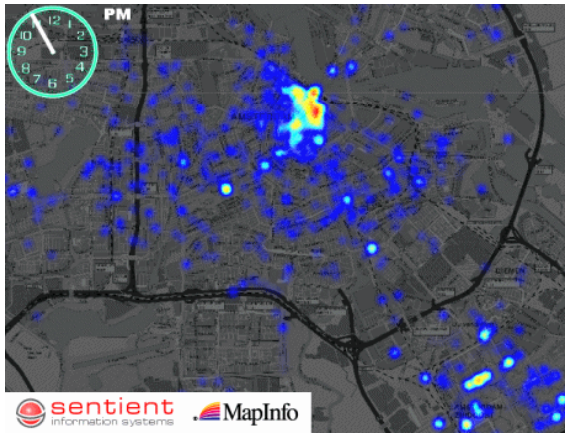
(Ack: In collaboration w/V. Kumar, M. Steinbach, P. Zhang)

HotSpots

- What is it?
  - Unusally high spatial concentration of a phenomena
    - Cancer clusters, crime hotspots
- Traditional Approach:
  - Spatial statistics based ellipsoids
- Our Recent Focus:
  - Computational Structure
    - Spatial Join-index reduces computational costs
  - Transportation network based hotspots
- Next: Spatio-temporal
  - Ex. Emerging hot-spots

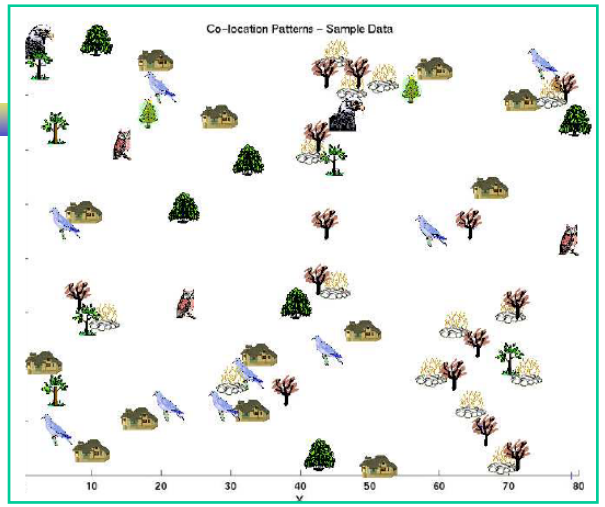


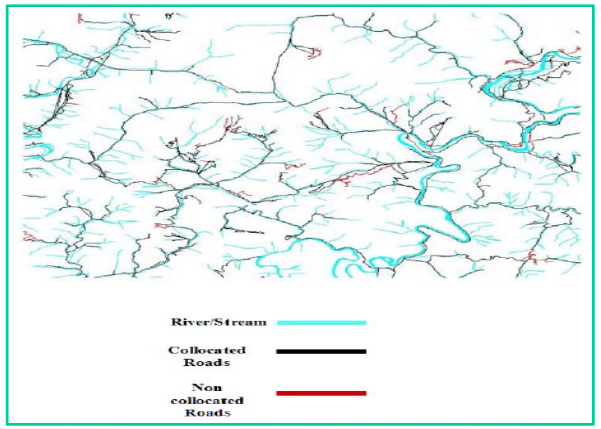


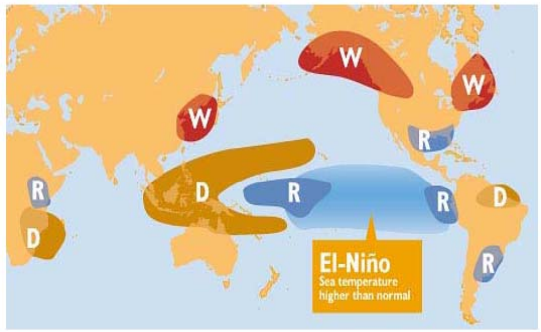


Colocation, Co-occurrence, Interaction

- What is it?
  - Subset of event types, whose instances occur together
  - Ex. Symbiosis, (bar, misdemeanors), …
- Traditional Approach:
  - Neighbor-unaware Transaction based approaches
- Our Approach:
  - Aggregate Functions on Neighbor relationships
  - Balance statistical rigor and computational cost
- Next: Spatio-temporal interactions
  - Item-types that sell well before or after a hurricane
  - Object-types that move together
  - Tele-connections

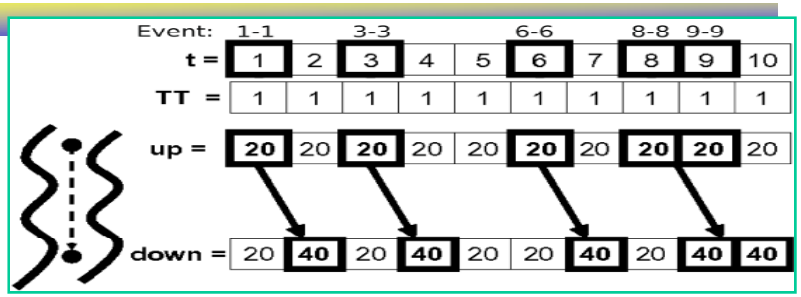


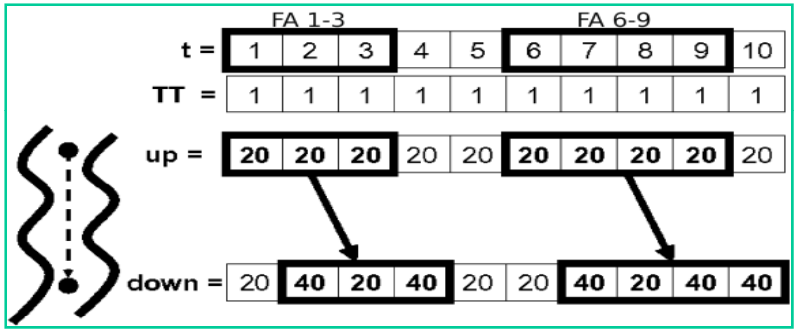


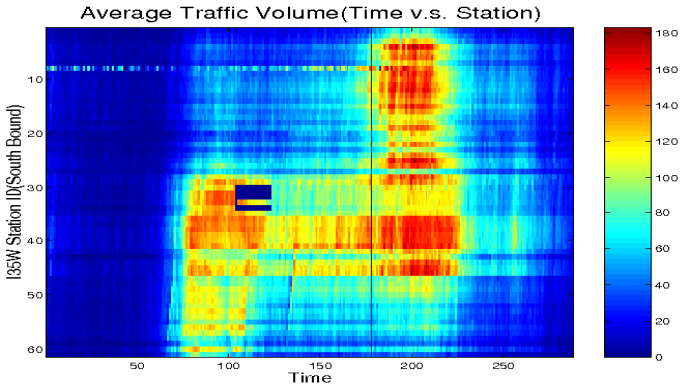


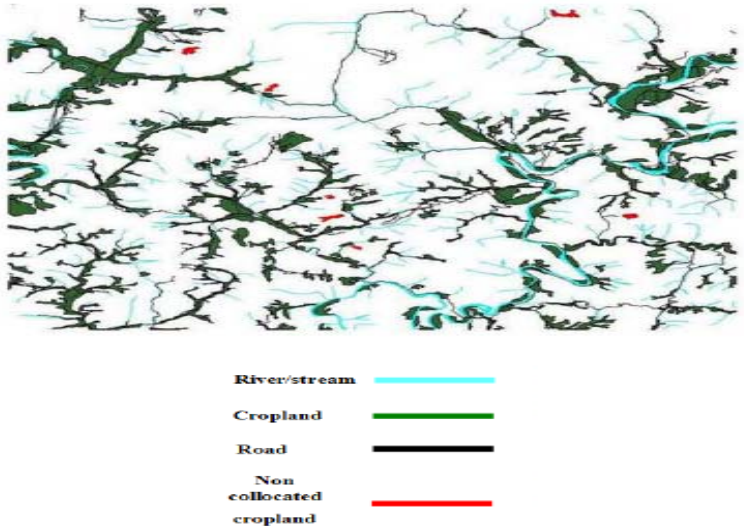
Spatial/Spatio-temporal Outliers, Anamolies

- What is it?
  - Location different from their neighbors
    - Discontinuities, flow anomalies
- Related Work
  - Transient spatial outliers
  - Anomalous trajectories
  - Computational Structure: Spatial Join
    - Very scalable using spatial DBMS
- Next
  - (Dominant) Persistent anomalies
  - Multiple object types, Scale

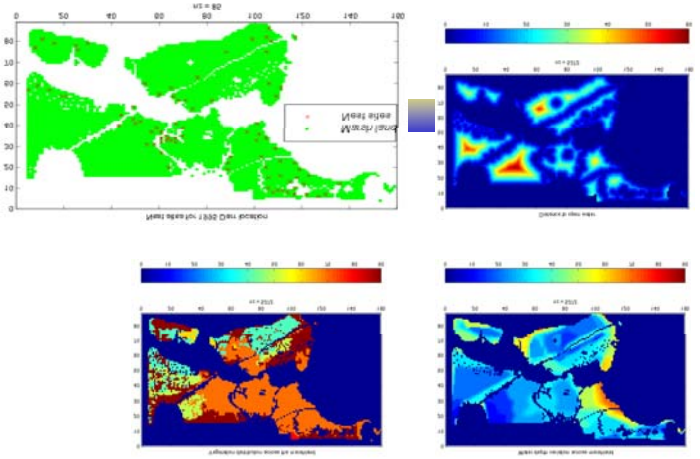








Space/Time Prediction



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

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

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