Gap Analysis for Spatial and Spatio-temporal Data Mining

Shashi Shekhar

McKnight Distinguished University Professor

Faculty of Computer Sc. and Eng., University of Minnesota

www.cs.umn.edu/~shekhar

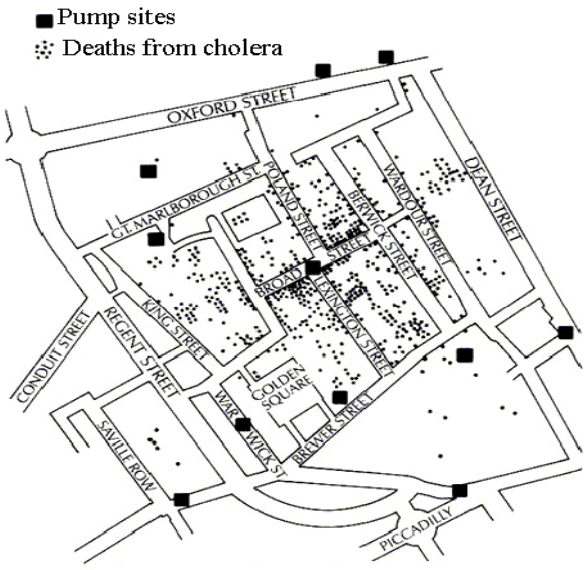
Presented at

NSF Workshop on GeoSpatial and GeoTemporal Informatics

January 8th-9th, 2009

Washington D.C.

Spatial and Spatio-temporal Data Mining



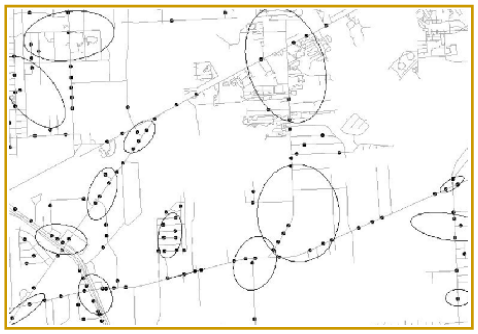
- What is it?
  - Identifying interesting, useful, non-trivial pattern
    - Hot-spots, discontinuities, co-locations, trends, …
  - in large spatial or spatio-temporal datasets
    - Satellite imagery, geo-referenced data, e.g. census
    - gps-tracks, geo-sensor network, …

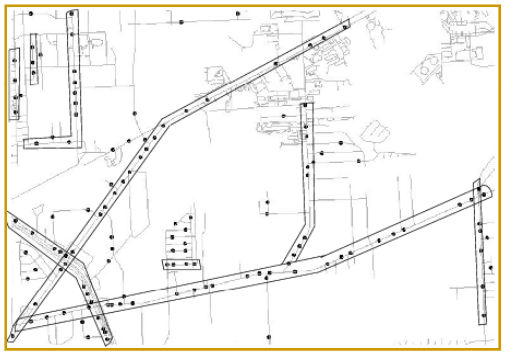
- Why is it important ?
  - Potential of discoveries and insights to improve human lives
    - Environment: How is Earth system changing? Consequences for humans?
    - Public safety: Where are hotspots of crime? Why?
    - Public health: Where are cancer clusters? Environmental reasons?
    - Transportation, 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

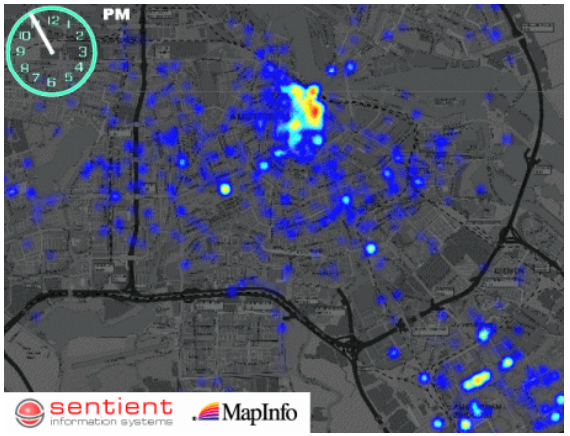
SPATIAL DATABASE RESEARCH GROUP

HotSpots

- What is it?
  - Unusally high spatial concentration of a phenomena
    - Cancer clusters, crime hotspots
- Solved
  - Spatial statistics based ellipsoids
- Almost solved
  - Transportation network based hotspots
- Failed
  - Classical clustering methods, e.g. K-means
- Missing
  - Spatio-temporal
- Next
  - Emerging hot-spots

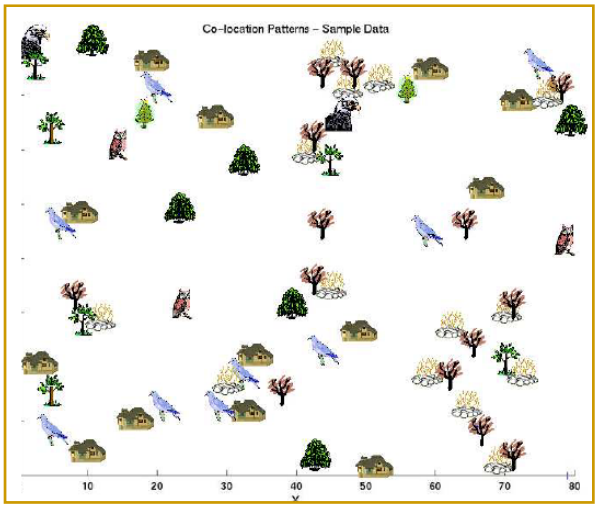


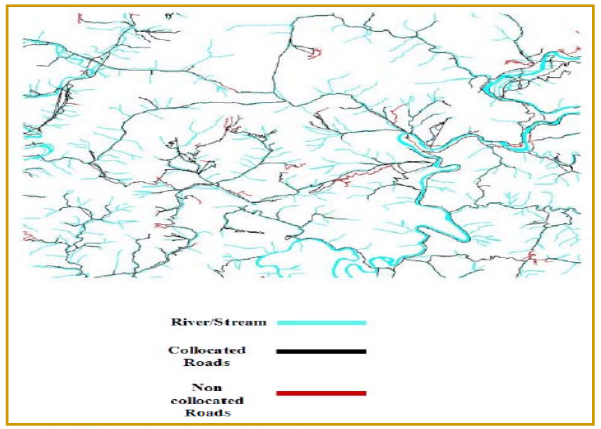


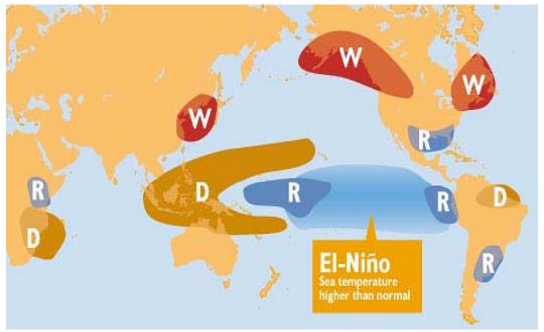


Colocation, Co-occurrence, Association

- What is it?
  - Subset of event types, whose instances occur together
  - Ex. Predator-prey species, (bar, misdemeanors)
- Solved
  - Colocation of point event-types
- Almost solved
  - Co-location of extended (e.g.linear) objects
- Failed
  - Neighbor-unaware Transaction based approaches
- Missing
  - Consideration of flow, motion, evolution
- Next
  - Tele-connection
  - Object-types that move / evolve together



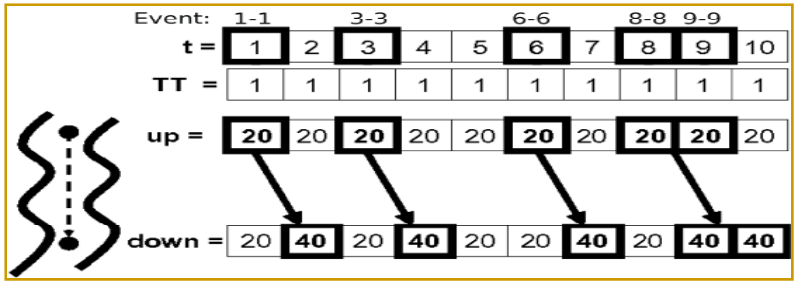


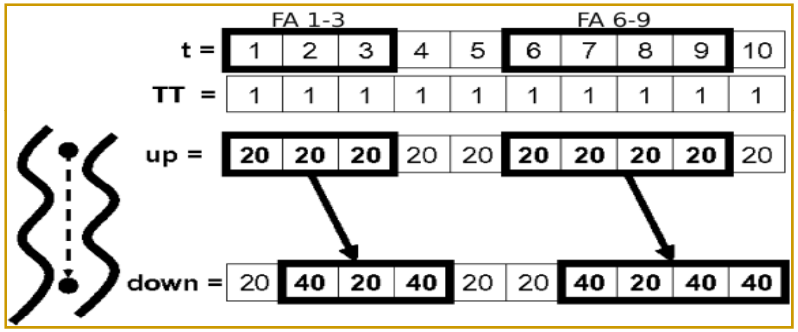


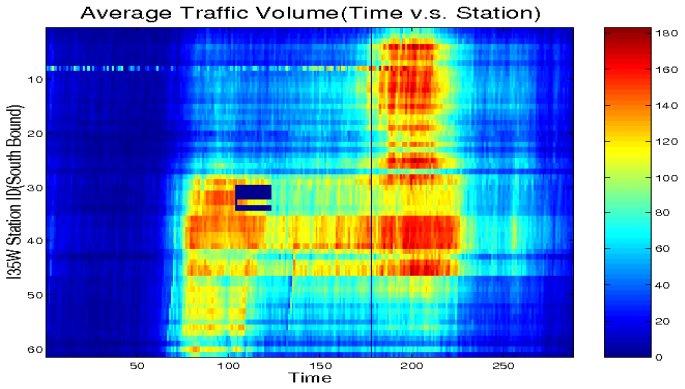
SPATIAL DATABASE RESEARCH GROUP

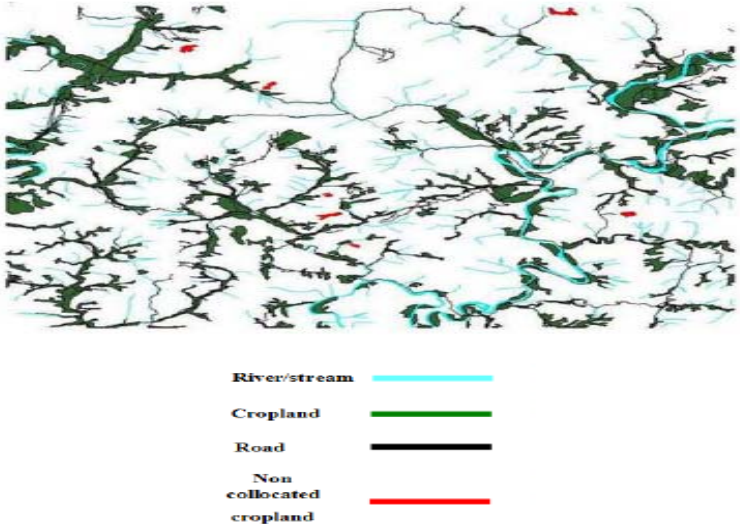
Spatial/Spatio-temporal Outliers, Anamolies

- What is it?
  - Location different from their neighbors
    - Discontinuities, flow anomalies
- Solved
  - Transient spatial outliers
- Almost solved
  - Anomalous trajectories
- Failed
- Missing
  - Persistent anomalies
  - Multiple object types, Scale
- Next
  - Dominant Persistent Anomalies

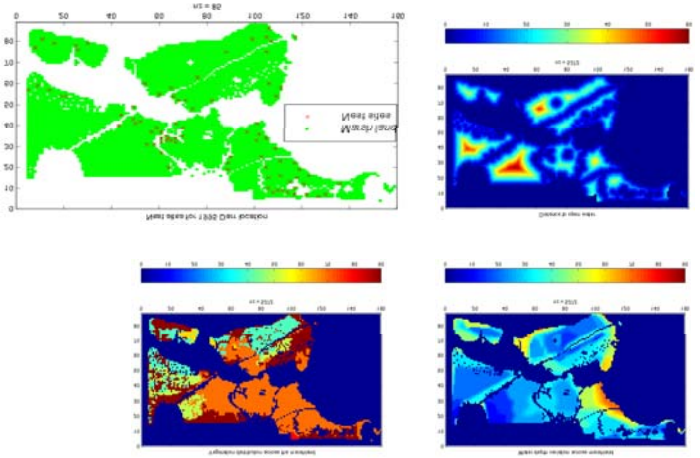








Space/Time Prediction



- 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

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

- Failed
  - Independence assumption
    - Models, e.g. Decision trees, linear regression, …
    - Measures, e.g. total square error, precision, recall

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

- Missing
  - Spatio-temporal vector fields (e.g. flows, motion), physics
- Next
  - Distance based errors, Scalable algorithms for parameter estimation