Shashi Shekhar

McKnight Distinguished University Professor
Faculty of Computer Sc. and Eng., Univ. of Minnesota
www.cs.umn.edu/~shekhar

Outline:

• Spatial/Spatio-temporal Database
• Spatial/Spatio-temporal Data Mining
Spatial / Spatio-temporal Databases: Example Projects

- **Evacuation Route Planning**

  - **Parallelize Range Queries**
  - Storing graphs in disk blocks

  - **Shortest Paths**
  - **Evacuation Route Planning**
    - Reception Center located at OSSEO JUNIOR HIGH SCHOOL 10223 93rd Avenue North Osseo, Minnesota
    - Orange: only in old plan
    - Green: Only in new plan
    - Blue: In both plans
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 )
Evacuation Route Planning: A Scenario

Nuclear Power Plants in Minnesota

[Map of Minnesota with nuclear power plants marked: Monticello, Twin Cities, Prairie Island]
Monticello Emergency Planning Zone

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

<table>
<thead>
<tr>
<th>Subarea</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>4,675</td>
</tr>
<tr>
<td>5N</td>
<td>3,994</td>
</tr>
<tr>
<td>5E</td>
<td>9,645</td>
</tr>
<tr>
<td>5S</td>
<td>6,749</td>
</tr>
<tr>
<td>5W</td>
<td>2,236</td>
</tr>
<tr>
<td>10N</td>
<td>391</td>
</tr>
<tr>
<td>10E</td>
<td>1,785</td>
</tr>
<tr>
<td>10SE</td>
<td>1,390</td>
</tr>
<tr>
<td>10S</td>
<td>4,616</td>
</tr>
<tr>
<td>10SW</td>
<td>3,408</td>
</tr>
<tr>
<td>10W</td>
<td>2,354</td>
</tr>
<tr>
<td>10NW</td>
<td>707</td>
</tr>
<tr>
<td>Total</td>
<td>41,950</td>
</tr>
</tbody>
</table>

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.
Computer Sc. Challenge: Time-varying Networks

"U.P.S. Embraces High-Tech Delivery Methods - (by Claudia Deutsch)
The research at U.P.S. is paying off. Last year, it cut 28 million miles from truck routes — saving roughly three million gallons of fuel — in good part by mapping routes that minimize left turns”
- New York Times (July 12, 2007)

Congestion Levels in road networks change with time resulting in travel time changes.

Increasing availability of traffic data through sensors on road networks.
Example queries on a time-varying network

1) Transportation network Routing
   - Varying Congestion Levels and turn restrictions ⇒ travel time changes.

<table>
<thead>
<tr>
<th>Static</th>
<th>Time-Variant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which is the shortest travel time path from downtown Minneapolis to airport?</td>
<td>Which is the shortest travel time path from downtown Minneapolis to airport at different times of a work day?</td>
</tr>
<tr>
<td>What is the capacity of Twin-Cities freeway network to evacuate downtown Minneapolis?</td>
<td>What is the capacity of Twin-Cities freeway network to evacuate downtown Minneapolis at different times in a work day?</td>
</tr>
</tbody>
</table>

2) Crime Analysis
   - Identification of frequent routes (i.e.) Journey to Crime

3) Knowledge discovery from Sensor data.
   - Spreading Hotspots
Secret Sauce: **Representation** of (Spatio-)temporal Networks

(1) **Snapshot Model** [Guting 04]

Node:  
Edge: Travel time

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

Node:  
Edge: [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

Predictable Future

Dijkstra’s, A*….

Non-stationary

SP-TAG, SP-TAG*, CapeCod

[Kanoulas07]

Unpredictable Future

Special case (FIFO)

General Case TEG: LP, Label-correcting

[Orda91, Kohler02, Pallotino98]

TAG: Transform to Stationary TAG

travel times $\rightarrow$ arrival times at end node $\rightarrow$ Min. arrival time series

Non-stationary TAG $\rightarrow$ Stationary TAG
Spatial / Spatio-temporal 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**
- Average Traffic Volume (Time vs. Station)

**Co-location Patterns**

**Tele connections**

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

■ What is it?
  ■ Unusually 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, Anomalies

- **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
Location Prediction – An Example

Nest locations

Vegetation distribution across the marshland

Distance to open water

Water depth variation across marshland

Vegetation durability

Water depth
Implication of Auto-correlation

<table>
<thead>
<tr>
<th>Name</th>
<th>Model</th>
<th>Classification Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classical Linear Regression</td>
<td>$y = x\beta + \varepsilon$</td>
<td>Low</td>
</tr>
<tr>
<td>Spatial Auto-Regression</td>
<td>$y = \rho W y + x\beta + \varepsilon$</td>
<td>High</td>
</tr>
</tbody>
</table>

$\rho$: the spatial auto-regression (auto-correlation) parameter

$W$: $n$-by-$n$ neighborhood matrix over spatial framework

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

$$
\ln(L) = \ln|I - \rho W| - \frac{n \ln(2\pi)}{2} - \frac{n \ln(\sigma^2)}{2} - SSE
$$
Space/Time Prediction

- What is it?
  - Models to predict location, time, path, ...
    - Nest sites, minerals, earthquakes, tornadoes, ...

- 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

\[
y = \rho W y + x \beta + \epsilon
\]

\[
\ln(L) = \ln|\mathbf{I} - \rho \mathbf{W}| - \frac{n \ln(2\pi)}{2} - \frac{n \ln(\sigma^2)}{2} - \text{SSE}
\]