A Geospatial Perspective

Michael F. Goodchild
University of California
Santa Barbara
Embedded social networks

• Embedded in geographic space (and time)
• What can we learn from the embedding?
• What constraints does the embedding impose?
• What do we know about embedded systems that can inform research?
Geospatial technologies

• GPS
  – measurement of position is now trivial

• Remote sensing
  – massive resources of imagery
  – ubiquitous, fine-resolution base maps
  – near real-time
    • days

• Geographic information systems
  – formal methods of representation
  – analysis and modeling
Interoperability

• Easy exchange of data
  – primarily syntactic

• Formal and informal location referencing
  – 120.12456 W, 34.89176 N
  – 909 West Campus Lane, Goleta, CA 93117, USA
  – 5789654N, 314654E, Zone 11, Northern Hemisphere
  – NE 1/4, Section 12, Township 23 Range 5 of the Second Principal Meridian
  – National Grid reference 11SKU36151156

• Mike Goodchild's house
Weaknesses

• Time
  – legacy of static map-based information

• 3D
  – recorded elevation (2.5D)
  – lack of support for full 3D structures

• Binary knowledge
Spatial knowledge

- Knowledge about properties $z$ present at locations $x$ in space-time (unary knowledge)
  - expressed as maps
    - when that knowledge is relatively static in time
  - increasingly dynamic
- Knowledge about the properties $z$ of pairs of places in space-time $x_1, x_2$ (binary knowledge)
  - e.g. distance, social affinity and interaction, travel time, flow, proximity
  - not ideally suited to mapping
Generic Flow Model

Glennon, TGIS 2010
Links

• Real or implied
• Attributed
• Directed
• Planar or non-planar
Geospatial data modeling

• Point, line, area classes
  – attributes and methods

• Association classes
  – attributes of pairs of objects
Learning from embedding

• Inferences from spatial and spatiotemporal form
  – footprints of process

• Context
  – vertical
    • what else is known about this location?
  – horizontal
    • what is known about nearby locations?
    • TFL
Laws of geography

• Nearby things are more similar than distant things
  – spatial dependence
  – distance decay

• Spatial heterogeneity
  – statistical non-stationarity
  – uncontrolled variance
  – spatial sampling designs
1843 map of London from David Rumsey collection

Pump and death locations from Snow
Swing rebellion of 1832
Daily patterns of georeferenced tweets, Los Angeles, August 2010
Distance decay

• A general pattern observed in processes embedded in geographic space

\[ I_{ij} = O_i D_j e^{-bd_{ij}} \]

• Wilson: the most likely distribution of interaction with distance if the total or mean distance is known

• Darren Hardy’s work on Wikipedia authorship
Articles with geotags

# of articles per unit area (log scale, 0.1° resolution)

988,522 articles
103,291 distinct locations
Wikipedia authorship

- **Registered** authors
  - Only username required
  - Name, email, etc. optional
  - IP address kept hidden

- **Anonymous** authors
  - IP address made public
  - But nothing else

### Contributions to “Copenhagen Opera House”

<table>
<thead>
<tr>
<th># of Contributions</th>
<th>Username or IP</th>
<th>Most Recent</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>Dybdahl</td>
<td>18-Sep-2005</td>
</tr>
<tr>
<td>6</td>
<td>85.233.237.71 (anon)</td>
<td>12-Jan-2008</td>
</tr>
<tr>
<td>3</td>
<td>Viva-Verdi</td>
<td>8-Sep-2006</td>
</tr>
<tr>
<td>1</td>
<td>Hemmingsen</td>
<td>3-Jan-2007</td>
</tr>
<tr>
<td>4</td>
<td>81.62.92.47 (anon)</td>
<td>15-Apr-2006</td>
</tr>
<tr>
<td>1</td>
<td>Thue</td>
<td>28-Feb-2006</td>
</tr>
<tr>
<td>2</td>
<td>Ghent</td>
<td>30-Apr-2006</td>
</tr>
<tr>
<td>3</td>
<td>Valentinian</td>
<td>7-Jan-2007</td>
</tr>
<tr>
<td>3</td>
<td>83.77.92.205 (anon)</td>
<td>10-Apr-2006</td>
</tr>
<tr>
<td>3</td>
<td>130.226.234.229 (anon)</td>
<td>29-Sep-2007</td>
</tr>
<tr>
<td>2</td>
<td>86.149.109.196 (anon)</td>
<td>15-Oct-2007</td>
</tr>
<tr>
<td>2</td>
<td>Uppland</td>
<td>24-Dec-2005</td>
</tr>
<tr>
<td>2</td>
<td>87.48.100.222 (anon)</td>
<td>12-Jan-2006</td>
</tr>
</tbody>
</table>
University of California, Santa Barbara
135 anonymous authors with 719 revisions; signature distance = 533 km
64% of articles at 2,000 km or less
Distribution of signature distances over 2,000 km
(n = 159,016 articles; 36%)

Distance between Europe and North America
A mixture?

- **Negative exponential distance decay**
  - for some entries
  - driven by familiarity, proximity-based interest
  - some fraction of contributors $\alpha$
- **Flat**
  - $b$ goes to 0
  - the death of distance
  - some fraction of contributors $1-\alpha$
The embedding space

• Invert to infer distance

\[ d_{ij} = -\frac{1}{b} \log \left[ \frac{I_{ij}}{O_i D_j} \right] \]

• Scale to obtain a space