Social Data Moves Towards Spatio-Temporal Future

Shashi Shekhar
McKnight Distinguished University Professor
Department of Computer Science and Eng., University of Minnesota
www.cs.umn.edu/~shekhar

Human-Centric/Social Data Session
ARO Workshop on Human-Centric Computation & Collective Intelligence
Arizona State University, Phoenix, AZ.
October 30th-31th, 2012
Outline

A. Social Data

B. Time

C. Space

D. Large-Scale Collaboration
Social Data

• History of Social Data
  – Pre-1900 – Census, meetings (e.g., public squares, conferences)
  – 1900s – Phone calls
  – 1970s – Internet, ftp, email, ...
  – 1980s – Usenet, gopher, cell-phone, email, ...
  – 1990s – www, On-line community, chat room, virtual world & games, ...
  – 2000s – Social networking, collaboration, participatory planning,
  – Workflow systems, e.g. Blackboard, Moodle, ...

• What is new in last few years?
  – Critical mass, e.g., audience size > 1 Billion
  – Social Influence, e.g. Tunisia, Egypt, India, Libya, ...
  – Real-time updates (e.g., Twitter) for Emergency Situation Assessment
  – Geo-location, e.g., HTML5, smart-phone, E911
  – Large Scale Collaboration: wikiipedia, crowd-sourcing, ...
Use Case 1: Sociology

1. Centrality:
2. (Information) Diffusion, Network Contagion
3. Community Detection, Network Structure
4. Bursts and Event Detection

Tracking Flu Transmission
Researchers following the spread of swine flu at an elementary school in 2009 found that students were more likely to catch the flu from playmates of their own sex than from the students seated near them. Below, a chart shows how the flu spread through students and their households.

Source: PNAS

The New York Times
Use Case 2: Social Sciences – Big Questions

1. How can we induce people to look after their health?
2. How do societies create effective and resilient institutions, such as governments?
3.
4.
5. How can we aggregate information possessed by individuals to make the best decisions?
6. How can we understand the human capacity to create and articulate knowledge?
7.
8. How and why does the ‘social’ become ‘biological’?
9.
10. Why do social processes, e.g. civil violence, either persist over time or suddenly change?

Sources: Harvard Meeting on Social Science (2010-2011), NSF SBE
Use Cases Beyond Sociology: Temporal Questions

Social Data

Social Trends, Dynamics, …

Time
Temporal Use Cases of Social Data

- Temporal Question for Society
  - Flu trend
  - Trend in crime, economy, …

- Temporal Questions in Sociology
  - How is trust or leadership changing over time?
  - Who are the emerging leaders in a group?
  - How long is the tenure of a leader in a group?
  - How long does it take to elevate the level of trust, e.g. visitor to friend or loyal customer?
Simpson’s Paradox in Time-varying Centrality

- How is trust or leadership changing over time?
  - Who is the most central node using between centrality?
  - Traditional Answer: D
- How may we compute it efficiently by avoiding unnecessary computations?
  - Challenge: violates dynamic programming principle
  - Critical time-point approach
Computational Modeling of Geo-social networks

That's the topic of discussion in a new paper by Shashi Shekhar and research assistant Dev Oliver, spatial data scientists at the University of Minnesota, titled Computational Modeling of Spatio-temporal Social Networks: A Time-Aggregated Graph Approach (PDF). The paper was highlighted on the blog GIS and Science today. We've excerpted and put in context key points below.
Compact Representation: Time-Aggregated Graph (TAG)

Snapshot Representation of a Trust Network at time = 1-10

As Yet Unanswered Time-based Questions of Value for Any Community

- How is trust or leadership changing over time?
- Who are the emerging leaders in a group?
- What are the recurring changes in a group?
- How long is the tenure of a leader in a group?
- How long does it take to elevate the level of trust such as a relationship changing from visitor to friend?

TAG Representation at time=1-10

[m_1, ..., m_T] - trust level at time = i
Use Cases Beyond Sociology: *Geo-Social*

Social Data

Geo-Social data & Spatial variation

Space
Geo-Social Data & Army

- Army Geo-social Questions
  - Where are friends? (Avoid friendly fire)
  - Where are enemies? (Geo-Targetting)

- Operational Security Advice by Army: Avoid Geo-tags!
  - Q. Why?

Geo-tags can show enemies your location

*ArmyTimes* Monday Dec 20, 2010

The Army is warning troops to be careful when using Facebook and other popular social networking sites because their geo-tagging features may show where U.S. forces are located in war zones.
Was Bin Laden Really 'Off the Grid'? [FOX NEWS] May 03, 2011

U.S. officials say Al Qaeda chieftain Usama bin Laden would have remained "off the grid" -- electronically and digitally invisible, without phone or Internet connection -- and free to plot more terror attacks if not for a Kuwaiti courier's cell-phone signal.

- What is adversary advising its people if we avoid Geo-tags!
- What can tradition Social Data Analytics do in this situation?
- Examine geo-social data from smart phone, cameras, videos, etc.
- However, U.S. officials were stunned to realize that whenever Kuwaiti or others left the compound to make a call, they drove some 90 minutes away before even placing a battery in a cellphone.
Denial & Deception Challenge: Black-holes

- **Grand Challenge:** Infer enemy-location despite denial and deception
  - Black hole patterns
  - Mine lack of data, *i.e.*, patterns of evasion, *e.g.*, donut hole, black-hole-
  - => Meld Socio-Physical Theories with Data Mining

(a) Blackhole detection using expected vs observed
(b) Occluded route prediction to identify potential key locations.
Geo-Social Data

• Explicit Geo
  – Geo-Socialize, Geo-social shopping & bargaining, ...
  – Crowd sourcing, VGI, citizen science, ...
  – Smart mobs, smart mobile workforce, ...
  – Supply chain, Sourcing: Local food & mood, ...

• Implicit Geo
  – Communicate: calls, messages, blogs, marketing, ...
  – Socialize: friends, family, co-workers, ...
  – Professional networks, free-lance, corporate SNs, ...

• Future
  – Find patterns despite denial and deception
  – Influence behavior via GSM
  – For challenges facing humanity
Are (smart-phone) GPS tracks social data?

- Ex. Weekday GPS track for 3 months
  - May provide Patterns of social life
  - Usual places & visits, routine activity theory
  - Rare places, Rare visits

<table>
<thead>
<tr>
<th></th>
<th>Morning 7am – 12am</th>
<th>Afternoon 12noon – 5pm</th>
<th>Evening 5pm – 12pm</th>
<th>Midnight 12midnight – 7pm</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Home</td>
<td>10</td>
<td>2</td>
<td>15</td>
<td>29</td>
<td>54</td>
</tr>
<tr>
<td>Work</td>
<td>19</td>
<td>20</td>
<td>10</td>
<td>1</td>
<td>50</td>
</tr>
<tr>
<td>Club</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td></td>
<td>15</td>
</tr>
<tr>
<td>Farm</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>120</td>
</tr>
</tbody>
</table>
Geo-social Data Has Tremendous Value!

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.

New Ways to Exploit Raw Data May Bring Surge of Innovation, a Study Says
Use Cases Beyond Sociology: Spatio-temporal

Social Data

Social Dynamics

Geo-social

Persistent Surveillance

Time

Space
Persistent Surveillance at American Red Cross

• Even before cable news outlets began reporting the tornadoes that ripped through Texas on Tuesday, a map of the state began blinking red on a screen in the Red Cross' new social media monitoring center, alerting weather watchers that something was happening in the hard-hit area. (AP, April 16th, 2012)
Spatio-temporal Social Data

• Social data sub-genre
  – Geo-tagged tweet, Ushahidi report, checkin, ...
  – Often more recent and richer than news

• Use cases: Persistent Surveillance
  – Outbreaks of disease, Disaster, Unrest, Crime, ...
  – Hot-spots, emerging hot-spots
  – Spatial Correlations: co-location, teleconnection
Spatio-temporal Social Questions

General Questions:
• Predict spatio-temporal spread of insurgency.
• How is trust varying over space and time?
• Where are emerging leaders coming from?
• Geo-dynamics of crowd behavior, unrest, …
• How is global footprint of international crime changing?
• Where are adversaries migrating from?
• …

Theoretical Questions:
Which patterns are accurately be inferred with statistical confidence despite duplicates and missing data?
• outbreaks
• peaks
• associations
Use Cases Beyond Sociology: Data Collection

- Social Data
- Large-Scale Collaboration
- Space
- Social map
Volunteered Geographic Information (VGI)

The harnessing of tools to create, assemble, and disseminate geographic data provided voluntarily by individuals (Goodchild, 2007). e.g., Wikimapia, Flickr, OpenStreetMap, Cyclopath

OpenStreetMap map of Haiti

Cyclopath: A volunteered bike route in Minneapolis, MN.
What motivates VGI volunteers?

- OSM:Haiti Earthquake :: Facebook:Egypt

**Method**
- Email survey (Spring 2010)
- 100+ OSM volunteers over the world

**Volunteer Motivation**
- Personal Satisfaction, Altruism,
- Learn and use Geo-knowledge
- Professional networking, socialize

**Volunteer Background**
- Non-academic non-GIS-professionals!
- Gender Skew
- Middle-age

**Incentive Preference**
- None
- Technical training
- Recognition

**Age Distribution**

- Professionals, but not Geo-professionals

- Unstated
VGI Questions

• How can soldiers help assess battlefield situation without revealing their locations to adversaries?

• Data Quality: How accurate and complete is crowd-sourced data?
  – How may one improve VGI data quality?
  – How do we detect multiple perspectives and adversarial views?
  – What can be learned from untrusted data?

• How does one recruit volunteers with local and social knowledge?
  – E.g., Road-name (official names vs. popular names)
  – How can we increase participation in crowd-sourcing?

• What are the limits of crowd-sourcing?
  – Wisdom of crowd vs. madness of crowd (e.g. witch-hunting, mis-perceptions)

• Time-critical VGI: How quickly can one map a disaster area to support relief?
  – What are time-critical bottleneck steps? How do we speed those up?
Summary

• Value of Social Data
  – High-frequency samples of large geo-social behavior
  – Analogies: Hubble telescope and sensor networks
  – Potential: persuade adversaries and friends, transform social science
  – Limitations: Denial and Deception, e.g., Bin Laden off-grid

• Way Forward: Spatio-temporal (ST) adds tremendous value
  – Many Army questions are spatio-temporal
    • understand enemy movements, protests, crowds, smart-mobs, …
  – ST-ignorance is an Operational Security hazard, e.g., geo-tagged social data!

• Grand challenges
  – How can soldiers help assess battlefield situation despite denial and deception
    • without revealing their locations to adversaries?
  – Computational Challenges:
    • Melding Data Mining with Social-Physical Theories
    • Violation of dynamic programming, Non-stationarity, Auto-correlation, …