Geo-Social Research

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

Annual Retreat Panel - The Geo-Social Revolution: Hype or Real?
Integrated Media Systems Center, University of Southern California, Los Angeles.
February 17th, 2011.
Spatial / Spatio-temporal Databases: Example Projects

- Parallelize Range Queries
- Evacuation Route Planning
  - Only in old plan
  - Only in new plan
  - In both plans
  - Shortest Paths
  - Storing graphs in disk blocks
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**

**Co-location Patterns**

**Tele connections**

(Ack: In collaboration w/V. Kumar, M. Steinbach, P. Zhang)
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 professionals
  – But not GIS-professionals!
  – Middle-age

• Incentive Preference
  – None
  – Technical training
  – Recognition

• Acknowledgement: R. Tewari
VGI Questions?

- 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?
- 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? (Thanks Beth!)
- How does one recruit volunteers with local and social knowledge? (e.g. road-name)
  - Challenge: official names vs. popular names
- How to locate missing people?
  - Coordinate family members after disasters, emergencies, etc.
- How can we increase participation in crowd-sourcing?
- What are the killer applications for crowd-sourcing?
- What are the technical hurdles?
- What are the limits of crowd-sourcing?
  - Wisdom of crowd vs. madness of crowd (e.g. witch-hunting, common mis-perceptions)
Computational Modeling of GSM

• Perspective on Phenomena, Behavior, Event
  • Geo: Where is it?
  • Social: Who participates? Ring-leaders? Level of trust
• Geo-constraints
  • Geographic distance, schedule, diurnal cycles, …
• Geo-constraints on social tie formation
  • Opportunities to meet a person reduce with distance
  • Cost (Face to face interaction) goes up with distance
  • Travel and communication technology reduces cost
    • but does not eliminate it!
• Do we wish liberation from Geo-constraint?
  • Time-travel, light-speed travel, teleportation, …
• Temporal, Social
  • How is trust, reputation, leadership or friendship changing over time?
  • Why did Facebook took over mySpace?
  • 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 earn/lose trust, e.g. friendship, customer loyalty?
Snapshots vs. Time-Aggregated Graph (TAG) Model

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

TAG Representation at time=1-10

Node

Edge

$[m_1, \ldots, (m_T)]$

$m_i$: trust level at time = i
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.
Outline

• Value added by Geo-Social
  – What is novel? Is it old wine in a new bottle?
  – Commercial potential (market, business models)
  – Relative to geospatial and social networking tools
  – New capability for users

• Risks and Threats

• Research Challenges

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.

KEY
- Male student
- Female student
- Household member

Source: PNAS
Geo, Social and Geo-Social

- Different Perspectives on Phenomena, Behavior, Event
  - Social: Who participates? Ring-leaders?
  - Geo: Where is it? Which way is it migrating?
  - Geo-social:
    - Global vs. local leaders, Where are the ring-leaders?
    - Which is more effecting in getting message out?
      - Social: via critical people or a social network, e.g. Twitter
      - Geo: via critical places, e.g. bill-boards
Geo vs. Social vs. Geo-Social

- Which is more effective in Influencing behavior?
  - Social: Via critical people (e.g. stars) & social networks, e.g. Twitter, …
  - Geo: Via critical places (e.g. malls, high-traffic roads, gathering places, …) & events (e.g. football games, state-fair, …)

• Options
  • (Social) Media & Communication:
    - Cyber: Twitter, blogs, U-tube, Facebook, …
    - Printing Press (e.g. newspaper, books, flyers, …)
    - Large-scale media, e.g. TV-networks, Radio-networks, …
    - Word of mouth
  • Geo:
    - Location-based messages: cell-phones,
      - Bill-boards, notice-boards at high traffic places
    - Gather at key places, e.g. capitol, historic monuments, …
    - Protest march, conference, …
    - Local media, e.g. radio-stations, local TV-channels, local newspapers, …
  • Geo-social:
    - Use cyber-messages via cell-phone, Twitter, U-tube
    - Gather protestors at key locations
    - Smart mobs: observe opponents to find openings

WI State Capitol, Feb. 2011
<table>
<thead>
<tr>
<th>Consumers</th>
<th>Social</th>
<th>Geo</th>
<th>Geo-Social</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which friends are online? e.g. Social Networking</td>
<td>Where is nearest ATM? Location Based Services</td>
<td>Which friends are nearby? Location-based social networks, Check-in</td>
<td></td>
</tr>
<tr>
<td>Business</td>
<td>Social / viral Marketing, Key influencers, Customer sentiments, ... Crowd-sourcing, ...</td>
<td>Geo-Marketing, e.g. radio, TV, newspaper, bill-boards, Logistics: Routing, Facility location and allocation, ... Geo-Tracking, Mining, Transportation</td>
<td>Geo-social deal-making, e.g. Groupon Geo-crowd-sourcing, e.g. OpenStreetMap, VGI Participatory Planning</td>
</tr>
<tr>
<td>Government</td>
<td>US-HSS Public health: infectious disease management.</td>
<td>Dept. of Interior, USGS, BLM, Land-use Planning, ... USDOJ: Crime hotspots CDC: Cancer clusters Critical places: Suez Canal Geo-surveillance: cameras, remote sensing (air, satellite), gps-collars, ...</td>
<td>USDOS: migrations UCDOC: Census USDOT-BTS: OD-matrix, Travel survey, ... Policing smart mobs (e.g. Egypt) at G-20, political conventions, ... Geo-social surveillance of smart mobs</td>
</tr>
<tr>
<td>Emergency Management, Situation Assessment, ...</td>
<td>US-HSS Public health: infectious disease management. Critical People: community leaders, brokers, ... Social-surveillance: wire-taps, background checks, ...</td>
<td>Dept. of Interior, USGS, BLM, Land-use Planning, ... USDOJ: Crime hotspots CDC: Cancer clusters Critical places: Suez Canal Geo-surveillance: cameras, remote sensing (air, satellite), gps-collars, ...</td>
<td>USDOS: migrations UCDOC: Census USDOT-BTS: OD-matrix, Travel survey, ... Policing smart mobs (e.g. Egypt) at G-20, political conventions, ... Geo-social surveillance of smart mobs</td>
</tr>
</tbody>
</table>
Is it old wine in new bottle?

- **History of Geo-social**
  - 1800s – Census, public squares, conferences, ...
  - 1970s – Internet, ftp, email, ...
  - 1980s – Usenet, gopher, cell-phone, email, ...
  - 1990s – www, On-line community, chat room, virtual world & games, ...
  - 2000s – Geo-Social networking, crowd-sourcing, participatory planning, ...
  - Workflow systems, e.g. Blackboard, Moodle, ...

- **What is new in last few years?**
  - Critical mass for businesses and government
    - Audience size > 500 Milllion
  - Social Impact, e.g. Tunisia, Egypt, Emergency Management, US DOS, ...
  - Technology improvements
    - Almost real-time updates (Twitter, Facebook)
    - Cloud computing, uTube, Google Earth, OpenStreetMap, ...
    - Smart cell-phones, E911, Peer 2 peer, ...
Geo-Media Market

• Market size (approx.)
  – 4+ Billions - Cell-phones
  – 600+ Millions - Tencent QQ, Skype, Facebook
  – 100-200 Millions: Qzone, Twitter, Gmail, Orkut,
    • Windows Live Spaces, Bebo, Myspace
  – 50 – 100 Millions – LinkedIn, Groupon, ...
  – ...

• Frequent Interaction
  – Half of Facebook users login everyday
  – 500 Billion minutes spent per day

• May grow even more!

  From check-in to volunteered continuous tracing for E911, emergencies.
What is the potential?

• What behavior does GSM influence?
  • Purchase
  • Revolution
  • Social responsibility

• Challenges
  • Human behavior evolves over time
  • Need high-frequency large samples
  • Surveys are expensive

• GSM observation as an alternative!
  • Half the 500 Million users log in everyday!
Geo-Social-Media (GSM) Market Segments

- **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**
  - **Influence behavior** via GSM
  - For challenges facing humanity
Societal Potential of GSM

• Can GSM encourage socially responsible behavior?

• Ex. Reduce greenhouse gases
  – ride-share & public transport
  – Will Gen. Y prefer socialization to driving?
  – “Check-in” for vanpools, buses, commuter trains, ...
  – Award titles of captain, pilot, (or even pirates of Carribbean)

• Recall: Check-in model
  – Frequent flyer program
  – Perfected by airlines
  – For customer loyalty
  – Will it work for buses? Vans?

I sold my car. It was distracting me dangerously from my texting.
Outline

• Value added by Geo-Social

• Risks and Threats
  – What are the risks?
  – Probabilities, Exposure
  – How to address risks?

• Research Challenges

USA TODAY 03/02/2011

Sexual predators turn to social media, mobile gadgets
By Byron Acohido, USA TODAY

Sexual predators are capitalizing on the popularity of mobile devices and social media to victimize children like never before, say police...
Risks in Geo-Social Systems

- Risk identified by many
  - Privacy, Geo-slavery
  - Security, data theft
  - Incivility
  - Change in society
Big Risks for Geo-Social Media (GSM)

- Under-understood & Over-used!
  - understanding of GSM lags behind its adoption and evolution!
  - Over-driving our headlights in fog or darkness!
  - Recall Titanic!

- Lack of quality leadership

- Public, Businesses, policy makers lose trust
  - and turn against it!
How to address the risks?

- Encourage GSM research
  - Understand GSM
  - Benefits, Unintended consequences
- Educate public, policy makers
- Start a science based dialogue
  - across industry, government, academia
  - Promote quality leadership
Outline

• Value added by Geo-Social
• Risks and Threat
• Research Challenges in combining Geo and Social
  – New opportunities
  – Potential discoveries about human behavior
  – Barriers

Facebook killed Geography!
Internet killed Geography!
Cell-phones killed Geography!

Social Networks trump other influences
Geo-Social Media (GSM) Research

• Two meta-questions
  – Are we investigating the right questions?
  – Are we investigating the questions right?
Are we investigating the right questions?

- **Big Questions**
  - Hard
  - Societal impact
  - Long-standing open questions
  - Social norms

- **Examples**
  - GSM questions
  - Social Science Questions
  - Policy questions
  - Computational Questions
  - What is after Geo-social?
Understanding GSM

• How can GSM induce people to desired behavior, e.g. purchase, volunteer?
• How and why does the GSM become ‘biological’?
  – Health effects of friends, e.g. smoking, drinking, exercise, nutrition, optimism, …
• How do we weigh different evidences to infer social ties, e.g. friendship, trust, …?
  – Communicate (e.g. email, text), socialize (facebook, linked-in), geo (e.g. proximity, tracks)
• How and why does geo become social? Social become geo?
  – Is Geo is first law of Social? Do we friend ones we are with, if we can’t be with friends?
  – How and why does geo constrain social? Social constrain geo?
• Why do GSM processes (either persist or) abruptly change?
  – mySpace vs. Facebook, hotmail vs. Gmail, Yahoo vs. Google, …
• How do GSM create effective and resilient institutions, such as cell-phone?
  – How resilient is Facebook to change in public trust and sentiment?
  – How can avoid the mySpace fate?
• How can a person or humanity best use a collection of GSM?
  – Integrate data across many GSMS, e.g. moodle, linked-in, facebook, …
  – manage multiple GSMS via a uniform interface for a person or an organization?
• How do we reduce the GSM ‘comfort gap’ across elderly, middle-aged and young?
  – What are GSM social norms? E.g. welcome rule, decorum, …
• How can we aggregate information volunteered by individuals to make the best decisions?
  – e.g. crowd-sourcing, VGI, citizen science
• How can we understand and articulate the GSM capacity to create knowledge?
• How can we be robust against ‘black swans’ — rare events with extreme consequences?
  – fragmentation and vulcanization of GSM or society
Example Black Swan

- Will GSM fragment our society?
- What are implication for society, country? People?

Figure 1. Network snapshots for different values of $a$ (where $a = p/q$) when no state update occurs (i.e. $r = w = 0$). Different colours indicate different states. Three classes of stable system behaviour can be distinguished: (a) when the rate of random rewiring is high with respect to random rewiring (e.g. $a = 1$), network topology is random; (b) when the rate of random rewiring is low (e.g. $a = 0.01$), the network fractures into a set of disconnected, homogeneous components; (c) when homophilous and random rewiring are balanced (e.g. $a = 0.1$), densely connected homogeneous state groups are evident, but the network as a whole also remains connected.
Sample Computational Questions in GSM

• What are major GSM use-cases in business, government and consumer spaces?
• How do we monitor our GSM reputation? How does GSM reputation compare with other ones?
• How does one infer social (e.g. friendship) from geo (e.g. trajectories) ?
• How does one infer geo (e.g. location) from social (e.g. friend locations)?
• How do we measure GSM concepts, recognize them in observations? accuracy and quality?
• How does one balance privacy and societal benefits from GSM data mining etc.?
• How does one improve quality of crowd-sourced GSM data?
• What can be learned from distrusted or semi-trusted GSM data?
• How do we represent GSM concepts with incomplete/ uncertain information, with alternative data models, and possibly with multiple representations for the same data?
• How do we conceptualize GSM data and phenomena ?
• How do we think about GSM phenomena, and to seek explanations for GSM patterns?
• How do we explain GSM phenomena through the application of methods of science?
• How do we store, access, and transform GSM concepts, facilitating data sharing, data transfer, and data archiving, while ensuring minimum information loss?
• How do we visualize GSM concepts on different media, e.g. paper, phones, computers?
More Computational Questions in GSM

• Which geo-social properties are computable? Which ones are not? How do we design geo-social-computational systems to address undecidable problems?
  
  • Identify all tents after Haiti Earthquake using aerial geo-social-imagery
  
  • Identify all churches in Rome, Geo-code geo-social video shots, map human terrain

• Which geo-social problems are computationally challenging? Which geo-social problems can not be decomposed into geo and social sub-problems? Which ones exhibit uni-modularity to facilitate greedy algorithms? Which ones obey Dynamic Programming (DP) principle? What algorithm design paradigms are appropriate for problems violating DP principle?

• What are interesting, useful, but non-trivial geo-social pattern families? How do we quantify their interest measures? How do we design efficient algorithms to mine those from large datasets? Which geo-social datasets may contain these pattern families? Which societal problems may benefit?

• Which statistical assumptions (e.g. independence of data samples, stationarity of models over space and time, scale-free) are violated by geo-social data and phenomena? What are the consequences of violated assumptions? How can statistical models be generalized to address the violations?

• How do we computationally represent social relationships? Their evolution over time? Differences across cultures and geographies? What are limitations of traditional graph theory in this context? How can we design new representations to address those limitations?

• Is there a semantic gaps among geo, social, geo-social and computational concepts? How can we reduce the semantic gap?
Policy Questions on Geo-Social Media (GSM)

• Q? Is GSM a force for liberation or repression? (Source: Hon. Secretary of State H. Clinton)
• Q? Is GSM free-speech a fundamental human right?

• Q? Should anyone have kill GSM switch?
• Q? If so, who? President, CEO, Mayor, Principal, Parents, … ?

• Q? What were unintended consequences of kill-Internet in Egypt?
• Q? Will USDOS fight against Internet repression have intended effect?
  • Location authentication (e.g. IP2location), Electromagnetic jamming, license contracts
• Q? What may be unintended consequences?
Big Questions in Social Science

1. How can we induce people to look after their health?
2. How do societies create effective and resilient institutions, such as governments?
3. How can humanity increase its collective wisdom?
4. How do we reduce the ‘skill gap’ between black and white people in America?
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. Why do so many female workers still earn less than male workers?
8. How and why does the ‘social’ become ‘biological’?
9. How can we be robust against ‘black swans’ — rare events with extreme consequences?
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
Beyond Geo-Social

• Perspective on Phenomena, Behavior, Event
  • Geo: Where is it? Which way is it migrating?
  • Social: Who participates? Ring-leaders?
  • Geo-social: Global vs. local leaders

• Perspectives beyond Geo-social
  • Time
  • Economic
  • Psychology
  • Biology, Chemistry, Physics
Spatio-temporal future of GSM

• Perspective on Phenomena, Behavior, Event
  • Geo: Where is it?
  • Social: Who participates? Ring-leaders?
  • Geo-social: Global vs. local leaders

• Temporal, Social
  • 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, e.g. visitor to friend or loyal customer?

• Spatio-tempo-Social
  • How is global reaching of Facebook changing?
  • Where are customers migrating to/from? Why?
  • How is trust changing over space and time?
  • Where are emerging leaders?
  • Geo-dynamics of Travel behavior, unrest, ...
  • …
Geo-Social Media (GSM) Research

• Two meta-questions in Research
  – Are we investigating the right questions?
  – Are we investigating the questions right?

Peer-review Process
Are we investigating the questions right?

• Goal of Science
  – Reproducible
  – Reduce subjectivity

• Methods in Science
  – Theory, mathematics, deduction
  – Controlled Experiments, hypothesis testing
  – Computer Simulation
  – Data-intensive

• Other Methods
  – Observation
  – Expert Polling, e.g. Delphi method
  – Case studies
  – ...
GSM Potential for Research

• Past Hurdles in behavior research
  • Detailed datasets – hi-frequency large sample
  • Instruments for ongoing large controlled experiments

• Can GSM help behavioral science?
  – From infrequent surveys to high-frequency ones
    • Hand-grab samples to sensor networks
  – From coarser data (e.g. census) to finer data
    • Telescope to Hubble telescope
Summary

• Value added by Geo-Social
  – High-frequency samples of large geo-social behavior
  – Analogies: Hubble telescope and sensor networks
  – Potential: help society, facilitate social science

• Risks and Threats:
  – Under-understood & over-used!
  – May alienate government, individuals and businesses!

• Way Forward
  – Research and understand
  – Big questions: social science, computational, data, ...