Mapping collective human activity in an urban environment based on mobile phone data

Department of Geography and Earth Sciences
University of North Carolina, Charlotte, U.S.A.
Motivation

1. Proliferation of a broad array of geographically referenced data derived from GPS receivers or georeferenced user-generated data.
   - Enables new opportunities in the analysis of human spatial behavior
2. Digital traces include mobile phone data and data and information from social media such as Twitter, Instagram, or Facebook, which has provided a plethora of human generated data.
3. User-generated mobile network traffic data may serve as a proxy to characterize the behavior of the society.
Motivation

1. Proliferation of a broad array of geographically referenced data derived from GPS receivers or georeferenced user-generated data.
   - Enables new opportunities in the analysis of human spatial behavior

2. Digital traces include mobile phone data and information from social media such as Twitter, Instagram, or Facebook, which has provided a plethora of human generated data.

3. User-generated mobile network traffic data may serve as a proxy to characterize the behavior of the society
Motivation

1. Proliferation of a broad array of geographically referenced data derived from GPS receivers or georeferenced user-generated data.
   - Enables new opportunities in the analysis of human spatial behavior

2. Digital traces include mobile phone data and data and information from social media such as Twitter, Instagram, or Facebook, which has provided a plethora of human generated data.

3. User-generated mobile network traffic data may serve as a proxy to characterize the behavior of the society
Motivation

1. Proliferation of a broad array of geographically referenced data derived from GPS receivers or georeferenced user-generated data.
   - Enables new opportunities in the analysis of human spatial behavior

2. Digital traces include mobile phone data and information from social media such as Twitter, Instagram, or Facebook, which has provided a plethora of human generated data.

3. User-generated mobile network traffic data may serve as a proxy to characterize the behavior of the society
Illustration: twitter activity

- Worldwide twitter activity during a world-cup soccer match between NED and BRA (July 12 2014)

https://vimeo.com/113202585

Preprocessing was done in Python, Plotting with R's ggplot2 and ggmaps libraries.

https://vimeo.com/113202585
Motivation continued.

In contrast to VGI such as Twitter, user-generated traffic in mobile phone networks is - from the perspective of the user - typically provided involuntary and lacking in content. For instance:

- The number of text messages sent/received is known but not the text itself.
- The number and duration of calls is known but not the topic of the talk itself.
In contrast to VGI such as Twitter, user-generated traffic in mobile phone networks is - from the perspective of the user - typically provided involuntary and lacking in content. For instance:

- The number of text messages sent/received is known but not the text itself.
- The number and duration of calls is known but not the topic of the talk itself.
Urban environments are of particular interest due to the typically higher degree and complexity of social dynamics.

- Research on such dynamics on the basis of mobile phone data has been conducted by the MIT SENSEable City Lab.

Such data-driven studies explore diverse characteristics of urban environments - from general urban activity patterns to individual mobility preferences or characteristics of group behavior.
Why urban environments?

1. Urban environments are of particular interest due to the typically higher degree and complexity of social dynamics.
   - Research on such dynamics on the basis of mobile phone data has been conducted by the MIT SENSEable City Lab.

2. Such data driven studies explore diverse characteristics of urban environments - from general urban activity patterns to individual mobility preferences or characteristics of group behavior.
Why urban environments?

1. Urban environments are of particular interest due to the typically higher degree and complexity of social dynamics.
   - Research on such dynamics on the basis of mobile phone data has been conducted by the MIT SENSEable City Lab.

2. Such data driven studies explore diverse characteristics of urban environments - from general urban activity patterns to individual mobility preferences or characteristics of group behavior.
Digital traces in urban environments

- Tourist activity in Acre, Israel.
- Tourists equipped with GPS devices.

Digital traces in urban environments

- Boston 311 incident reports

http://senseable.mit.edu/bos311/
Digital traces in urban environments

- Taxi drop-offs and pickups in New York City

http://hubcab.org/
Digital traces in urban environments

- Tourist activity around Milan, Italy
- Uses a combination of telecom, social media data and official statistics
Reflecting on the previous application, we can answer questions such as:

- What are the preferred itineraries of tourists?

- How much do tourists spend and where?

- Where did tourists come from and where do they go next?

- How do these patterns change with time, seasonally and in the long term?
Reflecting on the previous application, we can answer questions such as:

- What are the preferred itineraries of tourists?
- How much do tourists spend and where?
- Where did tourists come from and where do they go next?
- How do these patterns change with time, seasonally and in the long term?
Digital traces in urban environments

Reflecting on the previous application, we can answer questions such as:

- What are the preferred itineraries of tourists?
- How much do tourists spend and where?
- Where did tourists come from and where do they go next?
- How do these patterns change with time, seasonally and in the long term?
Reflecting on the previous application, we can answer questions such as:

- What are the preferred itineraries of tourists?
- How much do tourists spend and where?
- Where did tourists come from and where do they go next?
- How do these patterns change with time, seasonally and in the long term?
Explore spatial and temporal variations cell phone activity at different times of the day and days of the week in an urban setting.

Will enable an enhanced understanding of human behavior in the context of the spatial configuration of the city.

Underlying hypothesis: cell phone data can explain variations in intensity and similarity of collective human activity.
Problem statement

Research Objective & Hypothesis

1. Explore spatial and temporal variations in cell phone activity at different times of the day and days of the week in an urban setting.

2. Will enable an enhanced understanding of human behavior in the context of the spatial configuration of the city.

3. Underlying hypothesis: cell phone data can explain variations in intensity and similarity of collective human activity.
1. Explore spatial and temporal variations in cell phone activity at different times of the day and days of the week in an urban setting.

2. Will enable an enhanced understanding of human behavior in the context of the spatial configuration of the city.

3. **Underlying hypothesis**: cell phone data can explain variations in intensity and similarity of collective human activity.
Relevance & Contribution

1 **Relevance:** Enhanced understanding of collective human activity is useful to urban planners in facilitating sustainable decision making
   - Housing, zoning
   - Public services (e.g. transit, emergency medical services)

2 **Contribution:** Advances the investigating about the rhythms of social urban systems
   - Elaborate on spatiotemporal variations within collective human activity patterns based on mobile phone data
Relevance & Contribution

1. **Relevance:** Enhanced understanding of collective human activity is useful to urban planners in facilitating sustainable decision making
   - Housing, zoning
   - Public services (e.g. transit, emergency medical services)

2. **Contribution:** Advances the investigating about the rhythms of social urban systems
   - Elaborate on spatiotemporal variations within collective human activity patterns based on mobile phone data
Case study using user-generated mobile phone data for the city of Udine, a medium-size city located in Northern Italy.

- Mobile penetration rate is 155% (155 subscribers per 100 people).
- The operator that provided the data has a market share of 34.2%.
Data

1. The user-generated mobile phone dataset fully covers the study area;
2. The time period of interest is from July 20\textsuperscript{th} to September 30\textsuperscript{th} 2009;
3. Each of the attributes is spatially and temporally aggregated (250m x 250m grid cell and a 15-minutes time interval). 440 neighborhoods.

<table>
<thead>
<tr>
<th></th>
<th>Incoming</th>
<th>Outgoing</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Voice Calls</strong></td>
<td>number of calls terminated in a given grid cell but originating from another, adjacent cell</td>
<td>number of calls originating from a given grid cell but terminated in another, adjacent cell</td>
</tr>
<tr>
<td><strong>Text Messages SMS</strong></td>
<td>number of text messages terminated in a given grid cell but originating from another, adjacent cell</td>
<td>number of calls originating from a given grid cell but terminated in another, adjacent cell</td>
</tr>
<tr>
<td><strong>Total Traffic</strong></td>
<td>The total network traffic is measured in Erlang\textsuperscript{2} and implicitly contains certain types of VGI and is therefore used as a proxy for the overall collective mobile communication activity; the total traffic refers to a given single cell</td>
<td></td>
</tr>
</tbody>
</table>
Data

1. The user-generated mobile phone dataset fully covers the study area;
2. The time period of interest is from July 20th to September 30th 2009;
3. Each of the attributes is spatially and temporally aggregated (250m x 250m grid cell and a 15-minutes time interval). 440 neighborhoods.

<table>
<thead>
<tr>
<th></th>
<th>Incoming</th>
<th>Outgoing</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Voice Calls</strong></td>
<td>number of calls terminated in a given grid cell but originating from another, adjacent cell</td>
<td>number of calls originating from a given grid cell but terminated in another, adjacent cell</td>
</tr>
<tr>
<td><strong>Text Messages SMS</strong></td>
<td>number of text messages terminated in a given grid cell but originating from another, adjacent cell</td>
<td>number of calls originating from a given grid cell but terminated in another, adjacent cell</td>
</tr>
<tr>
<td><strong>Total Traffic</strong></td>
<td>The total network traffic is measured in Erlang and implicitly contains certain types of VGI and is therefore used as a proxy for the overall collective mobile communication activity; the total traffic refers to a given single cell</td>
<td></td>
</tr>
</tbody>
</table>
### Data

1. The user-generated mobile phone dataset fully covers the study area;
2. The time period of interest is from July 20th to September 30th 2009;
3. Each of the attributes is spatially and temporally aggregated (250m x 250m grid cell and a 15-minutes time interval). 440 neighborhoods.

<table>
<thead>
<tr>
<th></th>
<th>Incoming</th>
<th>Outgoing</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Voice Calls</strong></td>
<td>number of calls terminated in a given grid cell but originating from another, adjacent cell</td>
<td>number of calls originating from a given grid cell but terminated in another, adjacent cell</td>
</tr>
<tr>
<td><strong>Text Messages SMS</strong></td>
<td>number of text messages terminated in a given grid cell but originating from another, adjacent cell</td>
<td>number of calls originating from a given grid cell but terminated in another, adjacent cell</td>
</tr>
<tr>
<td><strong>Total Traffic</strong></td>
<td>The total network traffic is measured in Erlang(^2) and implicitly contains certain types of VGI and is therefore used as a proxy for the overall collective mobile communication activity; the total traffic refers to a given single cell</td>
<td></td>
</tr>
</tbody>
</table>
Data: gridded city
3D visualization

Space-time variation in the five variables of mobile phone usage - i.e., SMS in and out, voice calls in and out, and total internet traffic.
Filtering data

1. Data is stratified along weekdays (Monday-Friday) and weekends (Saturday and Sunday)
   - Take into account the two most evident differences in weekly mobile communication pattern.

2. To strike a balance between temporal granularity and temporal representativeness, we aggregated the 15 minutes intervals to 1 hour intervals.
Filtering data

1. Data is stratified along weekdays (Monday-Friday) and weekends (Saturday and Sunday)
   - Take into account the two most evident differences in weekly mobile communication pattern.

2. To strike a balance between temporal granularity and temporal representativeness, we aggregated the 15 minutes intervals to 1 hour intervals.
Methods: Geovisual Analytics

- Support explorative spatial data analysis and allow to investigate user-generated mobile phone data in space and time.

<table>
<thead>
<tr>
<th>Raw mobile phone data</th>
<th>Computing mobile communication profiles</th>
<th>Generating temporal trajectories</th>
<th>Assessing geospatial dimensions of temporal trajectories</th>
<th>Spatio-temporal information</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mapping collective human activity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOM</td>
<td>LISA</td>
<td>Geovisual Analytics</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Data clustering and projection procedure

1. Takes input data of multiple dimensions and arranges it on an output space of a lower dimension (most often 2)
   - Most similar observations are arranged in close proximity to one another on the output space
   - Inherently visual output of this procedure is an ideal starting point for exploring temporal variations across multiple attribute dimensions
   - The number of nodes is user defined

   - Small number of nodes: algorithm or clustering procedure
   - Large number of nodes: Emergent structures in the data can be visualized
Methods: Self-Organizing Maps

1. Data clustering and projection procedure

2. Takes input data of multiple dimensions and arranges it on an output space of a lower dimension (most often 2)
   - Most similar observations are arranged in close proximity to one another on the output space
   - Inherently visual output of this procedure is an ideal starting point for exploring temporal variations across multiple attribute dimensions
   - The number of nodes is user defined
     - Small number of nodes: algorithm ~ clustering procedure
     - Large number of nodes: Emergent structures in the data can be visualized
Methods: Self-Organizing Maps

1. Data clustering and projection procedure

2. Takes input data of multiple dimensions and arranges it on an output space of a lower dimension (most often 2)
   - Most similar observations are arranged in close proximity to one another on the output space
   - Inherently visual output of this procedure is an ideal starting point for exploring temporal variations across multiple attribute dimensions
   - The number of nodes is user defined
     - Small number of nodes: algorithm ~ clustering procedure
     - Large number of nodes: Emergent structures in the data can be visualized
Methods: Self-Organizing Maps

1. Data clustering and projection procedure

2. Takes input data of multiple dimensions and arranges it on an output space of a lower dimension (most often 2)
   - Most similar observations are arranged in close proximity to one another on the output space
   - Inherently visual output of this procedure is an ideal starting point for exploring temporal variations across multiple attribute dimensions
   - The number of nodes is user defined
     - Small number of nodes: algorithm simulates clustering procedure
     - Large number of nodes: Emergent structures in the data can be visualized
Methods: Self-Organizing Maps

1. Data clustering and projection procedure

2. Takes input data of multiple dimensions and arranges it on an output space of a lower dimension (most often 2)
   - Most similar observations are arranged in close proximity to one another on the output space
   - Inherently visual output of this procedure is an ideal starting point for exploring temporal variations across multiple attribute dimensions
   - The number of nodes is user defined
     - Small number of nodes: algorithm $\sim$ clustering procedure
     - Large number of nodes: Emergent structures in the data can be visualized
Methods: Self-Organizing Maps
The SOM training procedure is an iterative process where at each step, a random input vector $x$ is presented to the output space where nodes compete for $x$ based on Euclidean distance similarity between $x$ and all other vectors.

When the best matching node is identified, the values of its attributes are updated to reflect the placement of $x$ on the grid.

The end result of this process is an ordering of the output space so that neighboring nodes have similar values across the $n$ dimensions.
Methods: Self-Organizing Maps

1. The SOM training procedure is an iterative process where at each step, a random input vector $x$ is presented to the output space where nodes compete for $x$ based on Euclidean distance similarity between $x$ and all other vectors.

2. When the best matching node is identified, the values of its attributes are updated to reflect the placement of $x$ on the grid.

3. The end result of this process is an ordering of the output space so that neighboring nodes have similar values across the $n$ dimensions.
Methods: Self-Organizing Maps

1. The SOM training procedure is an iterative process where at each step, a random input vector $x$ is presented to the output space where nodes compete for $x$ based on **Euclidean distance similarity** between $x$ and all other vectors.

2. When the best matching node is identified, the values of its attributes are updated to reflect the placement of $x$ on the grid.

3. The end result of this process is an ordering of the output space so that neighboring nodes have similar values across the $n$ dimensions.
Methods: Self-Organizing Maps

- A 2,992 node rectangular grid (input: 105,600 records)
- Increasing the number of nodes produces identical trajectories

The movement of each trajectory therefore represents changes in the attribute space for a single geographic cell across the 24-hour time span; a greater distance between vertices implies a larger change in attribute values.
Temporal trajectory describes the typical daily spatiotemporal movement of a given geographic grid cell within the SOM output space (not the geography space!)

- **Length**: indicates the variation of the changes within and among the five input attributes.
  - Longer trajectories denote higher variability in the underlying communication profile.
  - Short trajectories: dormant neighborhoods.

- **Location and extent**: Where the trajectory is located within the SOM output and how big its bounding box is indicates communication preference.

- Shape and geometry.
Temporal trajectory describes the typical daily spatiotemporal movement of a given geographic grid cell within the SOM output space (not the geography space!)

- **Length**: indicates the variation of the changes within and among the five input attributes.
  - Longer trajectories denote higher variability in the underlying communication profile.
  - Short trajectories: dormant neighborhoods.

- **Location and extent**: Where the trajectory is located within the SOM output and how big its bounding box is indicates communication preference.

- **Shape and geometry**.
Temporal trajectory describes the typical daily spatiotemporal movement of a given geographic grid cell within the SOM output space (not the geography space!)

- **Length**: indicates the variation of the changes within and among the five input attributes.
  - Longer trajectories denote higher variability in the underlying communication profile.
  - Short trajectories: dormant neighborhoods.

- **Location and extent**: Where the trajectory is located within the SOM output and how big its bounding box is indicates communication preference.
  - Shape and geometry.
Temporal trajectory describes the typical daily spatiotemporal movement of a given geographic grid cell within the SOM output space (not the geography space!)

- **Length**: indicates the variation of the changes within and among the five input attributes.
  - Longer trajectories denote higher variability in the underlying communication profile.
  - Short trajectories: dormant neighborhoods.
- **Location and extent**: Where the trajectory is located within the SOM output and how big its bounding box is indicates communication preference.
- Shape and geometry.
Self-Organizing Maps: trajectories

Temporal trajectories on the SOM output space. a), b) and c) have the same lengths but different location and extent and shape.
Self-Organizing Maps: trajectories

440 trajectories (one per geographic cell). Longer trajectories: higher variations of collective human activity.
Results

Self-Organizing Maps: trajectories in geographic space

- Trajectories length linked back to the geographic space
- Local Indicators of Spatial Association (LISA, Anselin).

![Image of geographic map showing trajectory lengths and LISA results for different days of the week.]
Self-Organizing Maps: trajectories in geographic space

Comparing trajectory lengths for the weekend and weekday.
Virtual ground truthing
What have we done?

1. Illustrated an approach for exploring variations in intensity and similarity of collective human activity.

2. Analyzed vast volumes of user-generated mobile network traffic data with Geovisual Analytics, Self-Organizing Maps and Local Indicators of Spatial Association.

3. We distinguished voice from text and incoming from outgoing communication and took the overall network traffic, which includes data transfer, video communications, social media interaction etc., as a base reference.
What have we done?

1. Illustrated an approach for exploring variations in intensity and similarity of collective human activity.

2. Analyzed vast volumes of user-generated mobile network traffic data with Geovisual Analytics, Self-Organizing Maps and Local Indicators of Spatial Association.

3. We distinguished voice from text and incoming from outgoing communication and took the overall network traffic, which includes data transfer, video communications, social media interaction etc., as a base reference.
Conclusions

What have we done?

1. Illustrated an approach for exploring variations in intensity and similarity of collective human activity

2. Analyzed vast volumes of user-generated mobile network traffic data with Geovisual Analytics, Self-Organizing Maps and Local Indicators of Spatial Association.

3. We distinguished voice from text and incoming from outgoing communication and took the overall network traffic, which includes data transfer, video communications, social media interaction etc., as a base reference.
What have we done?

1. We estimated an average 24 hour mobile communication profile for every single 250m x 250m grid cell (the spatial unit to which the data were spatially aggregated and provided by the network operator) for weekdays (Monday to Friday) and weekends (Saturday and Sunday).

2. Input for the analysis based on SOM. The resulting temporal trajectories were characterized within the SOM output space and linked back to geographic space.
What have we done?

1. We estimated an average 24 hour mobile communication profile for every single 250m x 250m grid cell (the spatial unit to which the data were spatially aggregated and provided by the network operator) for weekdays (Monday to Friday) and weekends (Saturday and Sunday).

2. Input for the analysis based on SOM. The resulting temporal trajectories were characterized within the SOM output space and linked back to geographic space.
What are the findings?

1. **Underlying hypothesis**: variations in intensity and similarity of collective human activity can be revealed from mobile phone data.

2. Virtual ground truthing shows that for instance parking lots as a dedicated functional infrastructure tend to imply different levels of variation within cell phone activity (and hence human activity) on weekdays as compared to weekends.

3. Higher variations of collective human activity (i.e., longer trajectories) are located closer to the city center.
   - Could suggest a mono-centric city structure in terms of social urban dynamics (Batty 2008)
What are the findings?

1 **Underlying hypothesis**: variations in intensity and similarity of collective human activity can be revealed from mobile phone data.

2 Virtual ground truthing shows that for instance parking lots as a dedicated functional infrastructure tend to imply different levels of variation within cell phone activity (and hence human activity) on weekdays as compared to weekends.

3 Higher variations of collective human activity (i.e., longer trajectories) are located closer to the city center.
   - Could suggest a mono-centric city structure in terms of social urban dynamics (Batty 2008)
What are the findings?

1. **Underlying hypothesis**: variations in intensity and similarity of collective human activity can be revealed from mobile phone data.

2. Virtual ground truthing shows that for instance parking lots as a dedicated functional infrastructure tend to imply different levels of variation within cell phone activity (and hence human activity) on weekdays as compared to weekends.

3. Higher variations of collective human activity (i.e., longer trajectories) are located closer to the city center.
   - Could suggest a mono-centric city structure in terms of social urban dynamics (Batty 2008)
Conclusions

Take-home message

1. User-generated mobile phone data can serve as the basis for investigating variations of collective human activity.

2. Geovisual Analytics as an inductive reasoning framework can successfully be applied to explore large volumes of multivariate data and therefore supports all phases of the analysis.

3. Variations of collective human activity in diverse neighborhoods at different times correlate with the functional configuration of the city.

4. This research thus contributes to a better understanding of the dynamic social component in complex urban systems.
Conclusions

Take-home message

1. User-generated mobile phone data can serve as the basis for investigating variations of collective human activity.

2. Geovisual Analytics as an inductive reasoning framework can successfully be applied to explore large volumes of multivariate data and therefore supports all phases of the analysis.

3. Variations of collective human activity in diverse neighborhoods at different times correlate with the functional configuration of the city.

4. This research thus contributes to a better understanding of the dynamic social component in complex urban systems.
Take-home message

1. User-generated mobile phone data can serve as the basis for investigating variations of collective human activity.

2. Geovisual Analytics as an inductive reasoning framework can successfully be applied to explore large volumes of multivariate data and therefore supports all phases of the analysis.

3. Variations of collective human activity in diverse neighborhoods at different times correlate with the functional configuration of the city.

4. This research thus contributes to a better understanding of the dynamic social component in complex urban systems.
Take-home message

1. User-generated mobile phone data can serve as the basis for investigating variations of collective human activity.
2. Geovisual Analytics as an inductive reasoning framework can successfully be applied to explore large volumes of multivariate data and therefore supports all phases of the analysis.
3. Variations of collective human activity in diverse neighborhoods at different times correlate with the functional configuration of the city.
4. This research thus contributes to a better understanding of the dynamic social component in complex urban systems.
Thank you

Eric Delmelle, Ph.D.
Department of Geography and Earth Sciences
Center for Applied Geographic Information Science (CAGIS)
University of North Carolina at Charlotte
Eric.Delmelle@uncc.edu