A Study of Brightkite
connecting people through their visited places

Guanling Chen
University of Massachusetts Lowell

glchen@cs.uml.edu

6/14/2010, WPI
Workshop on Opportunistic RF Localization for Next Generation Wireless Devices

Emerging LSN

Smartphone  OSN  Location Services

Location-based OSN
**LSN Features**

- Users sharing their location
- Discover users and content based on location
- Check-in models
  - Foursquare, Gowalla, Brightkite
- Tracking models
  - Loopt, Latitude
Data Collection from Brightkite

- About two years updates 4,460,161
  - Checkin update 42.3%
  - Majority Scope is address level 56.2%
- Daily friend lists 168 days
- User profiles 70,337
  - Male 57.7%
  - Female 15.2%
User Classification

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Prop %</th>
<th>Total Updates</th>
<th>Active Days</th>
<th>Loyalty</th>
<th>Unique Places</th>
<th>Friends Number</th>
<th>PoPU</th>
</tr>
</thead>
<tbody>
<tr>
<td>average</td>
<td>N/A</td>
<td>95.6</td>
<td>27.0</td>
<td>0.25</td>
<td>16.3</td>
<td>7.35</td>
<td>0.09</td>
</tr>
<tr>
<td>trial</td>
<td>41</td>
<td>2.8</td>
<td>1.8</td>
<td>0.00</td>
<td>1.7</td>
<td>2.11</td>
<td>0</td>
</tr>
<tr>
<td>temporary</td>
<td>24</td>
<td>15.9</td>
<td>8.3</td>
<td>0.42</td>
<td>6.0</td>
<td>2.50</td>
<td>0.09</td>
</tr>
<tr>
<td>normal</td>
<td>23</td>
<td>72.0</td>
<td>32.7</td>
<td>0.35</td>
<td>20.9</td>
<td>9.58</td>
<td>0.12</td>
</tr>
<tr>
<td>active</td>
<td>11</td>
<td>658.3</td>
<td>153.3</td>
<td>0.70</td>
<td>85.6</td>
<td>32.90</td>
<td>0.15</td>
</tr>
<tr>
<td>privacy</td>
<td>2</td>
<td>35.3</td>
<td>3.1</td>
<td>0.01</td>
<td>2.5</td>
<td>3.01</td>
<td>0.45</td>
</tr>
</tbody>
</table>

Friendship Modeling

- Three-layer
  - Tag layer -- interests
  - Social graph layer
  - Location layer
**Tag Distance**

- Top 1000 tags

\[ m_t(i,j) = \sum_{n=1}^{N} \sum_{m=1}^{M} w_{im} \cdot t_{jn} \]

**Social Distance**

\[ D_{S:i,j} = \text{geodesic distance}(v_i, v_j) \text{ in } G'_s(V_s, E_s - e_{i,j}) \]
Location Distance

- Grid-based Location Metric
  \[ m_t(i,j) = \sum_n sng(c_{in}) \times sng(c_{jn}) \times (c_{in} + c_{jn}) \]

- 36,269 cities
- Threshold 50,000

Evaluation Results

<table>
<thead>
<tr>
<th>Top (^a)</th>
<th>Friend Pairs Number and Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Multi-layer</td>
</tr>
<tr>
<td>10,000</td>
<td>8,672</td>
</tr>
<tr>
<td>5,000</td>
<td>4,804</td>
</tr>
<tr>
<td>1,000</td>
<td>982</td>
</tr>
<tr>
<td>500</td>
<td>487</td>
</tr>
<tr>
<td>100</td>
<td>98</td>
</tr>
<tr>
<td>50</td>
<td>49</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Social Dist.</th>
<th>Loc. Metric</th>
<th>Tag Metric</th>
<th>Gender Diff</th>
<th>Age Diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Info. Gain</td>
<td>0.549</td>
<td>0.405</td>
<td>0.034</td>
<td>0.020</td>
<td>0.011</td>
</tr>
</tbody>
</table>
Privacy Protection

Privacy Metrics

- Proportion of protected updates (PoPU)

\[ PoPU = \frac{N_p}{N_t} \]

- Scopes
  - address, street, zip+4, zip, city, state, country
Gender and Age

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
<th>Unspecified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion (%)</td>
<td>57.7</td>
<td>15.2</td>
<td>27.1</td>
</tr>
<tr>
<td>Average PoPU</td>
<td>0.1226</td>
<td>0.1502</td>
<td>0.1464</td>
</tr>
</tbody>
</table>

Mobility

Average Visited Locations = 15.6
Friends

Sources
Regions

Community Analysis
More Information

- glchen@cs.uml.edu
- PS. Nan Li is looking for jobs.