Social and Technological Network Analysis: Spatial Properties

Salvatore Scellato
In this lecture we introduce social and technological networks embedded in metric space.

We discuss their properties and the metrics to study them.

We introduce recent research results on location-based social networks and user mobility.
Spatial Networks: examples and properties
Spatial networks

- Complex systems are very often organized under the form of networks where nodes and edges are embedded in metric space, with important consequences on their topological properties and on the processes which take place on them.

- transportation and mobility networks
- Internet router physical connections
- mobile phone networks
- urban road networks
- electric power grids
- social and contact networks
Spatial networks

- In all these examples space is relevant and network topology alone does not contain all the information.

- **Metric distance** directly influences the network structure by imposing higher costs on the connections between distant nodes.

- Longer links must be compensated by some other advantage.
Network planarity

- For most practical applications, the space is the 2-dimensional space and the metric is the usual Euclidean distance.

- A **planar graph** is a graph that can be embedded in the plane in such a way that its edges intersect only at their endpoints. Not all spatial networks are planar: we will mainly focus on **non-planar graphs**.

<table>
<thead>
<tr>
<th>Planar</th>
<th>Non-planar</th>
</tr>
</thead>
<tbody>
<tr>
<td>highway networks</td>
<td>mobile phone networks</td>
</tr>
<tr>
<td>railroads networks</td>
<td>flights and cargo ships networks</td>
</tr>
<tr>
<td>urban road networks</td>
<td>(online) social networks</td>
</tr>
</tbody>
</table>
Basic properties

• Given the metric distance associated to a spatial network, each couple of nodes has a spatial distance, whether they are connected or not. Thus, each link has an associated length.

• Noteworthy properties of a spatial network:

  • probability distribution of node degrees
  • probability distribution of spatial distances
  • probability distribution of link lengths
  • probability of connection as a function of distance
Basic properties

Distribution of distances (in km) between airports linked by a direct connection for the North American network. The straight line indicates an exponential decay with scale of order of 1,000 km.
Spatial properties of online social networks
More and more people want to share their geographic position with their friends.
Social ties and geographic distance

- A popular assumption is that most individuals try to **minimize the efforts** to maintain a friendship by interacting more with their spatial neighbors.

- The **connection costs** imposed by distance in spatial networks are not as important in social networks.

- Online tools and long-distance travel might result in the “**Death of Distance**”.

---

**UNIVERSITY OF CAMBRIDGE**

---

**THE DEATH OF DISTANCE**

*How the Communications Revolution Will Change Our Lives*

**FRANCES CAIRNCROSS**
Recent works on spatial social networks


- L. Backstrom, E. Sun, C. Marlow. Find me if you can: improving geographical prediction with social and spatial proximity. WWW 2010

- D. J. Crandall, L. Backstrom, D. Cosley, S. Suri, D. Huttenlocher, and J. M. Kleinberg. Inferring social ties from geographic coincidences. PNAS 2010


Effect of distance on social connections

- One fundamental spatial property of social networks is that the probability of friendship between two individuals decays as an inverse power of their geographic distance.


\[ P(d) \propto d^{-1} + \epsilon \]  \[ P(d) \propto d^{-2} \]  \[ P(d) \propto d^{-1} \]
Interesting questions...

- How is **geographic distance** affecting social ties in online location-based networks?

- Do users exhibit **homogeneous** or **heterogeneous** socio-spatial properties?

- What are the spatial properties of **social triads**?

- How are **spatial and social factors** simultaneously shaping how individuals create their connections?
Approach

• We have acquired data about the socio-spatial network of 3 real-world location-based services

• We design two randomized models of a socio-spatial network to better understand which factors shape the real networks.

• We study how individual users create their social links and their social triangles over space.
## Datasets

<table>
<thead>
<tr>
<th>Service</th>
<th>brightkite</th>
<th>foursquare</th>
<th>Gowalla</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nodes</strong></td>
<td>54,190</td>
<td>258,706</td>
<td>122,414</td>
</tr>
<tr>
<td><strong>Social links</strong></td>
<td>213,668</td>
<td>2,854,957</td>
<td>580,446</td>
</tr>
<tr>
<td><strong>Average degree</strong></td>
<td>7.88</td>
<td>22.07</td>
<td>9.48</td>
</tr>
<tr>
<td><strong>Average clustering coefficient</strong></td>
<td>0.181</td>
<td>0.191</td>
<td>0.254</td>
</tr>
<tr>
<td><strong>Average distance between friends [km]</strong></td>
<td>2,041</td>
<td>1,442</td>
<td>1,792</td>
</tr>
<tr>
<td><strong>Average distance between users [km]</strong></td>
<td>5,651</td>
<td>8,494</td>
<td>5,663</td>
</tr>
</tbody>
</table>
Distance between users and between friends

- **Friends tend to be much closer than random users**: about 50% of social links span less than 100 km, while about 50% of users are more than 4,000 km apart.
The decay is less sharp than in other systems: **location-based services appear affected by distance in a weaker way.**
Network randomization

Two randomized models, which capture either the geographic or the social properties of the original social networks and randomize everything else.

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
<th>Social properties</th>
<th>Spatial properties</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Original data</strong></td>
<td>No modification.</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td><strong>Geo model</strong></td>
<td>Fix node locations and reassign all links according to probability P(d).</td>
<td>✘</td>
<td>✔</td>
</tr>
<tr>
<td><strong>Social model</strong></td>
<td>Fix links and shuffle all node locations.</td>
<td>✔</td>
<td>✘</td>
</tr>
</tbody>
</table>
Average friend distance

\[ w_i = \frac{1}{k_i} \sum_{j \in \Gamma_i} l_{ij} \]

- Link length
- Node degree
- Node neighborhood

Graphs showing the CDF of link length for different datasets:
- Foursquare
- Brightkite
- Gowalla
Distance strength and correlation with degree

\[ s_i = \sum_{j \in \Gamma_i} l_{ij} = k_i w_i \]

\[ s(k) \propto k^\beta \quad \beta \in [1.10, 1.18] \]
A link is equally likely to belong to a social triangle regardless of its geographic length.
Users exhibit **heterogeneity**, as there are users with smaller triangles and users with wider ones.
Correlation with degree

- Users with many friends belong to triangles with longer links.
A gravity model for spatial social networks?

Gravity model

\[ P(i, j) \propto \frac{N_i N_j}{f(d_{ij})} \]

- Links connecting popular users tend to be much longer, while a user might connect to an unimportant one only when they are close to each other.
Gravitational attachment: temporal evolution of online social networks
Temporal evolution of a geospatial social network

- **Temporal daily snapshots** of Gowalla data between May and August 2010, with information about all user accounts, their profiles, their friends and their check-ins.

- We study three main elements of temporal network growth:
  - how new social edges are created
  - how social triangles are created
  - how fast social edges are added by users

- Methodology based on **Maximum Likelihood Estimation** (MLE).

### Properties at the end of measurement period

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Nodes</td>
<td>122,414</td>
</tr>
<tr>
<td>Social links</td>
<td>580,446</td>
</tr>
<tr>
<td>Average degree</td>
<td>9.48</td>
</tr>
<tr>
<td>Average clustering coefficient</td>
<td>0.254</td>
</tr>
<tr>
<td>Average distance between friends [km] A</td>
<td>1,792</td>
</tr>
<tr>
<td>Average distance between users [km]</td>
<td>5,663</td>
</tr>
</tbody>
</table>
Edge attachment models

- $D$: proportional to a power $\alpha$ of its degree
- $A$: proportional to a power $\alpha$ of its age
- $G$: inversely proportional to a power $\alpha$ of its geographic distance
- $DG$: proportional to its degree and inversely proportional to a power $\alpha$ of its geographic distance

The main factors driving in social edge attachment are node degree and geographic distance and that a gravity model which combines them is the most suitable option.
Predominance of triangle-closing links

- Triadic closure is the predominant factor shaping network growth over time: **new edges are exponentially more likely to connect people sharing at least one friend**, closing social triangles.

- Social connections at shorter social distance tend to have higher geographic distances, while links spanning more hops have lower spatial distance: **geographic proximity is complementary to social closeness.**
Triadic closure is mainly driven by social processes, while geographic distance is not an important factor: having many connections in common results in higher probability of connection.
Temporal evolution: inter-edge time gap.

- There is a wide range of variability in how quickly nodes start adding new edges after they join the network, captured by the truncated power-law of inter-edge time gaps.

- Nodes with higher degree add links at a higher pace: given a fixed temporal period, higher degree nodes add more links than lower degree ones.
Gravitational attachment model

1. A new node joins the network according to a certain arrival discipline and positions itself over the space;

2. The new node samples its lifetime from an exponential distribution;

3. The new node adds its first edge according to a gravity model;

4. A node with a given degree k samples a time gap from the degree-dependant distribution and then goes to sleep for that time gap;

5. When a node wakes up, if its lifetime has not expired yet it creates a two-hop new edge using the random–random triangle-closing model and repeats step 4.
Model evaluation: preferential attachment vs. gravitational model

- Preferential attachment mechanisms need to be turned into gravity-based mechanisms, which are able to correctly balance the effect of node attractiveness and the connection costs imposed by spatial distance.
User mobility: universal laws governing movements between places
Check-in concept

• The **check-in dynamics** involves different factors:

• **social**: broadcast your location to your friends

• **competition**: achieve goals and beat your friends

• **economic**: unlock benefits or deals with local businesses

• Unprecedented chances to understand how users actively engage with individuals **places**.
Check-in activity - weekdays
Check-in activity - weekends
Check-in map

New York - morning
New York - night
Data

- Data collected from public *checkins* generated in

- 925,030 users around the globe over a period of 6 months in 2010.

- 34 Cities that span 4 continents and 11 countries.

- For the first time analysis of human mobility among 5 million discrete spatial entities (places).

- All spatial points feature GPS-accuracy down to 10 meters.
Global vs urban mobility

2 problems:

i. Power-laws not sufficient to explain urban mobility. How do we model those movements?

ii. While the distributions from city to city feature similar shapes, scales may vary significantly. Why?
Distance is not enough

Stouffer's law of intervening opportunities states, "The number of persons moving over a given distance is directly proportional to the number of opportunities at that distance and inversely proportional to the number of intervening opportunities."

Empirically proven using data for migrating families in the city of Cleveland. We investigate the plausibility of this theory for urban movements in Foursquare.

S. Stouffer (1940) Intervening opportunities: A theory relating mobility and distance, American Sociological Review 5, 845-867
The importance of density

- Stouffer’s Theory of Intervening Opportunities motivated us to inspect the impact of places (=opportunities) in human mobility.

- Place density seems by far more important than the city area size with respect to mean length of human movements
- The rank for each transition between two places \( u \) and \( v \) is the number of places \( w \) that are closer in terms of distance to \( u \) than \( v \) is. Formally: \( \text{rank}(u,v) = |\{w : d(u, w) < d(u, v)\}| \).

- The rank essentially accounts for the relative density between two places \( u \) and \( v \).

- We have measured a power-law exponent \( \alpha = 0.84 \pm 0.07 \) for the rank distributions of all cities.
A model to capture human urban mobility

Transition probability between places according to rank

\[ Pr[u \rightarrow v] \propto \frac{1}{\text{rank}_u(v)^a} \]

Real spatial distribution of places in each city
Simulation

- For all cities we have used the average value of the $\alpha$ exponent in the rank relationship ($\alpha = 0.84$).

- We have considered all places in the city as potential starting points for our agents and have averaged the output.

- For each city we have employed the empirically recorded places within its territory.

- Hence, while the rank element of the model is universal, the set of places used in the simulations varies from city to city.
Capturing urban movements in 34 cities

Any heterogeneity observed in human mobility across cities appears due to **geographic variations**. The rank-based model can cope with the **complex spatial variations** in densities observed in urban environments.
Summary

• We have introduced spatial networks and described how to study them.

• We have analyzed research results on the spatial properties of social networks

• We have used geo-social analysis techniques to understand how spatial and social factors can be combined to describe and model social networks.

• We have presented the analysis of mobility trajectory study over geographical data in urban environment
References

• **Socio-spatial Properties of Online Location-based Social Networks**


• **A tale of many cities: universal patterns in human urban mobility**
  Anastasios Noulas, Salvatore Scellato, Renaud Lambiotte, Massimiliano Pontil, Cecilia Mascolo. arXiv:1108.5355