CONTRIBUTIONS OF SOCIAL NETWORK ANALYSIS TO UNDERSTANDING LEPROSY TRANSMISSION

Social Networks

• A **social network** is a social structure made of nodes (which are generally individuals or organizations) that are tied by one or more specific types of relations or “edges”

• Social network Analysis:
  – Relationship of nodes
  – Relations can be real or pseudo (generated post hoc)
Introduction

• Real world networks that are not regular or random are called complex networks
• Study of complex networks took off in the late 1990’s
• Networks can be analyzed as one-mode or two-mode.
• People in a shared space are two-mode:
Social network 2-mode to 1-mode

Places can be neighborhood or minimum distance

Hidden challenges
Klovdahl et al. Networks and tuberculosis: an undetected community outbreak involving public places (2001)
Our study: social networks defined by geographic distance

• A node or relationship means that two individuals occupy a space within a predefined geographic distance at the same time (1 Year)
Steps in our Project

– Collection of geolocated data about residence and other activities over 10 years in cases and controls
– Generation of networks of shared places
– Characterization and comparative analysis of these networks
Methods

- We developed a new computational system (GRAPHTUBE) to generate activity/geographic networks of 397 cases and 211 controls.
- Using filters the software builds person-to-person (PP) two-mode networks combining activity (housing, school, workplaces and leisure activities), geographic location, time and sociodemographic characteristics.
Methods

Network properties analyzed included degree, density, average topological distance between nodes, and clustering coefficient.

The results took into account:

a) differing network parameters for cases and controls
b) the effect of geographical distance in locales frequented by cases and controls using both individual geolocation and neighborhood
Methods

• c) analysis of the proximity matrices of cases and controls
• d) comparison of the topological distances in the networks of cases and controls
• e) comparison of critical distances in the networks of cases and controls for the establishment of dynamic states
Methods

• Network properties analyzed
  – Degree
  – Density
  – Average topological distance
  – Clustering coefficient
Methods

The results took into account:
- different network parameters for cases and controls
- geographic effects using both individual geolocation and neighborhood
- comparison of the topological distances in the networks of cases and controls
- comparison of critical distances in the networks of cases and controls
Results

The results demonstrate differences in the topology of case and control networks, especially in school and workplace networks, demonstrating the potential for systematic differences in contact
Social Networks generated by geographic distance

- Which cutoff better represents the interaction between infected individuals?
Which cutoff better represents the interaction between infected individuals?

• Using the concept of the average minimum path ($x_{ave}$) among the nodes of the network

• The minimum path between two nodes is the minimum number of steps needed to go from one node to another

• $x_{ave}$ is the mean among all the pairs
Social Networks generated by geographic distance

• If the cutoff is high $\rightarrow$ the network has too many connections (everybody is connected to everybody else)

• If the cutoff is low $\rightarrow$ the network has too few connections and the network is divided into small sub-networks with bottlenecks
Social Networks generated by geographic distance

• The distance that maximizes $x_{\text{ave}}$ generates a network with the minimum number of connections to connect all nodes

• This distance is called the critical distance
Evaluation of critical distance

Residence → 600m

Maximum distance between nodes
Another consideration

• These values are affected by network size
• We have fewer controls than cases
• Standardization through random generation of control networks of equivalent size
## Evaluation of critical distance

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Critical distance</th>
<th>N</th>
<th>Critical distance (randomly)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Residence</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cases</td>
<td>401</td>
<td>200</td>
<td>226*</td>
<td>400</td>
</tr>
<tr>
<td>Controls</td>
<td>226</td>
<td>600</td>
<td>226</td>
<td>600</td>
</tr>
<tr>
<td><strong>Workplace</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cases</td>
<td>309</td>
<td>300</td>
<td>133*</td>
<td>500</td>
</tr>
<tr>
<td>Controls</td>
<td>233</td>
<td>500</td>
<td>133</td>
<td>500</td>
</tr>
<tr>
<td><strong>School</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cases</td>
<td>251</td>
<td>200</td>
<td>177*</td>
<td>500</td>
</tr>
<tr>
<td>Controls</td>
<td>177</td>
<td>600</td>
<td>177</td>
<td>600</td>
</tr>
</tbody>
</table>

When network size is standardized, controls \(\approx\) cases \(\rightarrow\) can compare the networks.
Comparison of activity networks by social network measures

In general, there are differences in the structure of the residence network for cases and controls.
What is the probability (P) that a control shares a site with a case?

Where P is the number of edges that connect cases to controls divided by the total number of edges in the network.
Randomized test

Since the number of cases and controls are not the same, the networks were randomized 10,000 times to calculate the probability $P$ that the probability of a case shares a place with a control at the $i^{\text{th}}$ randomization.

The result is a p-value that demonstrates the difference between the observed probability and the null model in the random network of the same size.

![Histogram showing the distribution of $P$ values with observed and randomized data.](image)
### Matrix of significance/distance

#### Residence: Clusters of cases closer than controls

<table>
<thead>
<tr>
<th></th>
<th>Case</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case</td>
<td>0**</td>
<td>1</td>
</tr>
<tr>
<td>Control</td>
<td>0.957</td>
<td>0.096</td>
</tr>
</tbody>
</table>

#### Workplace: Cases surrounded by controls

<table>
<thead>
<tr>
<th></th>
<th>Case</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case</td>
<td>0.016*</td>
<td>0.995</td>
</tr>
<tr>
<td>Control</td>
<td>0.551</td>
<td>0.702</td>
</tr>
</tbody>
</table>

#### School: Clusters of controls closer than cases

<table>
<thead>
<tr>
<th></th>
<th>Case</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case</td>
<td>0.65</td>
<td>0.475</td>
</tr>
<tr>
<td>Control</td>
<td>0.999</td>
<td>0.009**</td>
</tr>
</tbody>
</table>

### Matrix of distance (m)

<table>
<thead>
<tr>
<th></th>
<th>Case</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case</td>
<td>6479</td>
<td>7040</td>
</tr>
<tr>
<td>Control</td>
<td>7429</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Case</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case</td>
<td>5018</td>
<td>5494</td>
</tr>
<tr>
<td>Control</td>
<td>5818</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Case</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case</td>
<td>6876</td>
<td>6862</td>
</tr>
<tr>
<td>Control</td>
<td>6645</td>
<td></td>
</tr>
</tbody>
</table>

* Hipótese nula é a de que as probabilidades observadas sejam menores ou iguais a probabilidades do modelo nulo

Hidden challenges

** p<0.01  * p<0.05
Relationship among cases and controls for residence

Cases meet cases

Controls meet controls
Relationship among cases and controls at workplace and school

At work: “cases are surrounded by controls”

At school: “Controls are surrounded by “cases”
Conclusions

• The residence network of cases and controls is the most different on all the parameters

• The school and workplace network do not behave like a random network
  – The same is not observed for residence network
Conclusions

• Taking into account the critical distance networks of cases and controls we concluded:
  – For residence: clusters of cases are closer
  – For school: cases are surrounded by controls
  – For workplaces: clusters of controls are closer
Conclusions

• Preventive measures should be expanded not only to the neighborhood of cases but also to workplaces of cases