What are trails? (1)

- Graph theory: A trail in a walk with no repeated edge. The length of a trail is constrained by the number of edges.

- Trail is a path of an ego through time and space
  - people, ideas, diseases etc.

- It is a time-ordered sequence, i.e., a sequence of observations taken at different times.
What are trails? (2)

- Question 1: How can networks be generated from trail data?

- Question 2: Can we always use classic network metrics on networks created from trails?

Importing Trail Data (1)
Importing Trail Data (2)

Select a format and then one or more files to import:
- **Trail data**: [Specify the EPGA nodes and their nodes](filename). Nodes nodes are defined as such: (node A, node B, node C, node D, node E).
- **Nodes ID**: [Specify the EPGA nodes](filename). Nodes nodes are defined as such: (node A, node B, node C, node D, node E).
- **Node ID**: [Specify the EPGA nodes](filename). Nodes nodes are defined as such: (node A, node B, node C, node D, node E).

Specify how to determine the EPGA nodes for each row:
- **Column**: [Specify the EPGA nodes](filename). Nodes nodes are defined as such: (node A, node B, node C, node D, node E).
- **Filenames**: [Specify the EPGA nodes](filename). Nodes nodes are defined as such: (node A, node B, node C, node D, node E).

Specify how to determine the Timestamp for each row:
- **File Name**: [Specify the EPGA nodes](filename). Nodes nodes are defined as such: (node A, node B, node C, node D, node E).

Create new EPGA nodes during import
- **Create New EPGA nodes**: [Specify the EPGA nodes](filename). Nodes nodes are defined as such: (node A, node B, node C, node D, node E).

---

Importing Trail Data (3)

Select a format and then one or more files to import:
- **Trail data**: [Specify the EPGA nodes](filename). Nodes nodes are defined as such: (node A, node B, node C, node D, node E).
- **Nodes ID**: [Specify the EPGA nodes](filename). Nodes nodes are defined as such: (node A, node B, node C, node D, node E).
- **Node ID**: [Specify the EPGA nodes](filename). Nodes nodes are defined as such: (node A, node B, node C, node D, node E).

Specify how to determine the EPGA nodes for each row:
- **Column**: [Specify the EPGA nodes](filename). Nodes nodes are defined as such: (node A, node B, node C, node D, node E).
- **Filenames**: [Specify the EPGA nodes](filename). Nodes nodes are defined as such: (node A, node B, node C, node D, node E).

Specify how to determine the Timestamp for each row:
- **File Name**: [Specify the EPGA nodes](filename). Nodes nodes are defined as such: (node A, node B, node C, node D, node E).

Create new EPGA nodes during import
- **Create New EPGA nodes**: [Specify the EPGA nodes](filename). Nodes nodes are defined as such: (node A, node B, node C, node D, node E).
Importing Trail Data (5)

Importing Trail Data (6)
Importing Trail Data (7)

Importing Trail Data (8)
Importing Trail Data (9)

- Data is imported both as a sequence of “per time slice” networks and aggregated transitional networks (number of transitions ego has between two nodes)
  - “Per time slice” networks → Looms
  - Aggregated transitional networks → Markov Chains

Looms (1)

- Visualization depends on what we wish to observe
- Good indicator of timeline
- Sometimes cluttered
Looms (2)

Al-Qaida’s target selection over time

Networks From Trails (1)

- Question 1: How can networks be generated from trail data?
  - Markov Chains - network of transitional probabilities (or cumulative weights) among nodes i.e. each node represents a location or an individual
Networks From Trails (2)

<table>
<thead>
<tr>
<th>Time</th>
<th>4 pm@Apr. 1</th>
<th>3 pm@Apr. 2</th>
<th>9 am@Apr. 3</th>
<th>1 pm@Apr. 3</th>
<th>2 pm@Apr. 4</th>
<th>4 pm@Apr. 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trail 1</td>
<td>F1</td>
<td>F2</td>
<td>F3</td>
<td>F2</td>
<td>F1</td>
<td>F2</td>
</tr>
<tr>
<td>Trail 2</td>
<td>F2</td>
<td>F3</td>
<td>F4</td>
<td>F2</td>
<td>F1</td>
<td>F1</td>
</tr>
<tr>
<td>Trail 3</td>
<td>F2</td>
<td>F3</td>
<td>F1</td>
<td>F1</td>
<td>F2</td>
<td>F3</td>
</tr>
</tbody>
</table>

\[
P(F_i \rightarrow F_j) = \frac{N(F_i \rightarrow F_j)}{\sum_j N(F_i \rightarrow F_j)}
\]

From Trails to Transitional Networks

- Observe ego’s transitions from one state to another
- Aggregate the observed transitions
- Create probabilities from the aggregated values
Why do we care about high dimensional networks?

- Both sequential and “memory” property of the data has to be accounted for
  - network-analytic methods make the fundamental assumption that paths are transitive, i.e. the existence of paths from a to b and from b to c implies a transitive path from a via b to c.

Example 1 – Function Calling

<table>
<thead>
<tr>
<th>Function Caller</th>
<th>Function Called</th>
</tr>
</thead>
<tbody>
<tr>
<td>F2</td>
<td>F3</td>
</tr>
<tr>
<td>F2</td>
<td>F1</td>
</tr>
<tr>
<td>F2</td>
<td>F3</td>
</tr>
<tr>
<td>F1</td>
<td>F2</td>
</tr>
<tr>
<td>F1</td>
<td>F2</td>
</tr>
</tbody>
</table>

Time

Function Caller: F1, F2, F3
Function Called: F2, F1, F3

Time

1/3 2/3

We lost the temporal component!
Why do we care about high dimensional networks?

- Agent’s paths and previous actions matter
  - First-order network is built by taking the number of transitions between pairs of nodes as edge weights (or scaled to transitional probabilities)

Why do we care about high dimensional trails?

- Agent’s paths and previous actions matter
  - First-order network is built by taking the number of trails between pairs of nodes as edge weights (or scaled to transitional probabilities) → PROBLEM!!
    - Same nodes could be used by different entities coming from different nodes following their own path
  - Solution → splitting the “crossroad” nodes
    - We care about where ego comes from
    - More accurate simulation of movement patterns observed in the original data
### Example 2 - Jihadist Groups (1)

<table>
<thead>
<tr>
<th>Group Name</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISIL</td>
<td>Business</td>
</tr>
<tr>
<td>Al-Qaida</td>
<td>Police</td>
</tr>
<tr>
<td>ISIL</td>
<td>Military</td>
</tr>
<tr>
<td>Al-Qaida</td>
<td>Military</td>
</tr>
<tr>
<td>Al-Qaida</td>
<td>Government (General)</td>
</tr>
<tr>
<td>ISIL</td>
<td>NGO</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

**Time**

### Example 2 - Jihadist Groups (2)

<table>
<thead>
<tr>
<th>Group Name</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISIL</td>
<td>Business</td>
</tr>
<tr>
<td>Al-Qaida</td>
<td>Police</td>
</tr>
<tr>
<td>ISIL</td>
<td>Military</td>
</tr>
<tr>
<td>Al-Qaida</td>
<td>Military</td>
</tr>
<tr>
<td>Al-Qaida</td>
<td>Government</td>
</tr>
<tr>
<td>ISIL</td>
<td>NGO</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

**First Order Network**

- Business
- Military
- Government
- Police
- NGO
- Time
Example 2 - Jihadist Groups (3)

First Order Network

Business   Military
Police     NGO

Higher Order Network

Business   Military | Business
Police     Military | Police
NGO

Example 2 - Jihadist Groups (4)

<table>
<thead>
<tr>
<th>Group Name</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISIL</td>
<td>Business</td>
</tr>
<tr>
<td>Al-Qaida</td>
<td>Police</td>
</tr>
<tr>
<td>ISIL</td>
<td>Military</td>
</tr>
<tr>
<td>Al-Qaida</td>
<td>Military</td>
</tr>
<tr>
<td>Al-Qaida</td>
<td>Government</td>
</tr>
<tr>
<td>ISIL</td>
<td>NGO</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Time

More informative and better representation of the data!
Higher Order Networks (1)

- Rethinking the building blocks of a network:

  - Instead of using a node to represent a single entity, we break down the node into different higher order nodes that carry different dependency relationships (each node can now represent a series of entities)

  - Military | Business and Military | Police → the edges can now involve multiple different targets as entities and carry different weights → second-order dependencies.

Higher Order Networks (2)

- Out-edges are in the form of $i|h \rightarrow k$ instead of $i \rightarrow k$, transitional probability from node $i|h$ to node $j$ is

  \[ P(X_{t+1} = j|X_t = (i|h)) = \frac{N(i|h \rightarrow j)}{\sum_k N(i|h \rightarrow k)} \]

- Movement depends on the current node and on one or more other entities in the new network representation
Higher Order Networks (3)

- This new representation is consistent with conventional networks and compatible with existing network analysis methods
  - We need to be careful when using the network metrics and have full graph of how network is created and what edges represent!

- PROBLEM – How to determine optimal order of the Higher Order Network?
  - Statistical analysis, Maximum likelihood, ...

Importing High-Dimensional Trails
Trail Report

[Image of a computer screen showing a Trail Report window]

CASOS