



Trails and Networks: From Trails to Networks and Higher-order Networks

Mihovil Bartulovic
mbartulovic@cmu.edu

Dr. Kathleen M. Carley
kathleen.carley@cs.cmu.edu



Carnegie Mellon

Center for Computational Analysis of
Social and Organizational Systems
<http://www.casos.cs.cmu.edu/>



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.



June 2020



Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

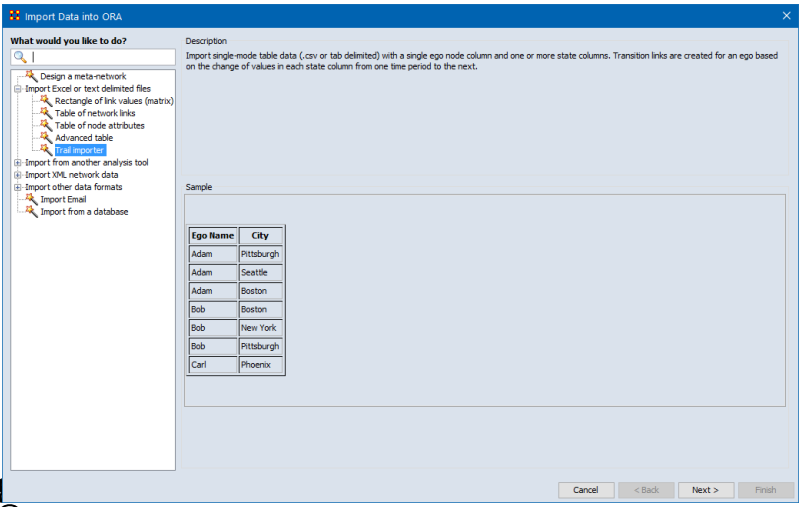
What are trails? (2)

- Question 1: How can networks be generated from trail data?
- Question 2: Can we always use classic network metrics on networks created from trails?

CASOS
June 2020

Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

Importing Trail Data (1)



What would you like to do?

- Design a meta-network
- Import Excel or text delimited files
 - Rectangle of link values (matrix)
 - Table of network links
 - Table of node attributes
 - Advanced table
 - Trail importer**
- Import from another analysis tool
- Import XML network data
- Import other data formats
- Import Email
- Import from a database

Description
Import single-mode table data (.csv or tab delimited) with a single ego node column and one or more state columns. Transition links are created for an ego based on the change of values in each state column from one time period to the next.

Sample

Ego Name	City
Adam	Pittsburgh
Adam	Seattle
Adam	Boston
Bob	Boston
Bob	New York
Bob	Pittsburgh
Carl	Phoenix

Cancel < Back Next > Finish

CASOS
June 2020

Carnegie Mellon
IST Institute for Software Research

Importing Trail Data (2)

Meta-Network Manager

ORA 3.0.0.9.87

File Edit Preferences Data Management Generate Networks Analysis Simulations Visualizations System Help

Import Data into ORA

Select a format and then one or more files to import:

Trail data /Users/Mihovil/Dropbox/Data for Trail Training & Case Studies/jihadist_groups data.csv Browse

Use the Ego and timestamp tab to define how to extract ego nodes and how to timestamp the trail. Choose one or more State Info tabs to choose columns with states through which the ego moves.

Ego and timestamp State Info Add State

Specify the Ego nodeset:

Nodeset Class: Agent

Nodeset ID: Jihadist Group

Specify how to determine the Ego node for each row:

Column: gname

Fixed name:

Specify how to define the Timestamp for each row:

Column: Date Date pattern: yyyy-MM-dd

File is sorted increasing by time

Create new Ego nodes during import

Clear Dataset Import Files

Cancel < Back Next > Finish

Find: Larger sizes:

CASO June 2020

Carnegie Mellon
IST Institute for Software Research

Importing Trail Data (3)

Meta-Network Manager

ORA 3.0.0.9.87

File Edit Preferences Data Management Generate Networks Analysis Simulations Visualizations System Help

Import Data into ORA

Select a format and then one or more files to import:

Trail data /Users/Mihovil/Dropbox/Data for Trail Training & Case Studies/jihadist_groups data.csv Browse

Use the Ego and timestamp tab to define how to extract ego nodes and how to timestamp the trail. Choose one or more State Info tabs to choose columns with states through which the ego moves.

Ego and timestamp State Info Add State

Specify the Ego nodeset:

Nodeset Class: Agent

Nodeset ID: Jihadist Group

Specify how to determine the Ego node for each row:

Column: gname

Fixed name:

Specify how to define the Timestamp for each row:

Column: Date Date pattern: yyyy-MM-dd

File is sorted increasing by time

Create new Ego nodes during import

Clear Dataset Import Files

Cancel < Back Next > Finish

Find: Larger sizes:

CASO June 2020



Carnegie Mellon
IST Institute for Software Research

Importing Trail Data (5)

Meta-Network Manager - No Meta-Network Loaded

Import Data into ORA

Select one or more Ego nodes to analyze by selecting the trails to consider and the states to pass through.

Select Ego Nodes | Select Trails | Select State Nodes | Select Options

Node ID

- Al-Qaida
- Taliban
- Other
- Al-Shabaab
- Boko Haram
- Islamic State ...

Select/Clear All | Select/Clear Visible

6 / 6 Selected, 6 / 6 Visible

Cancel < Back Next > Finish

Find: _____

CASO June 2020

Carnegie Mellon
IST Institute for Software Research

Importing Trail Data (6)

Meta-Network Manager - No Meta-Network Loaded

Import Data into ORA

Select one or more Ego nodes to analyze by selecting the trails to consider and the states to pass through.

Select Ego Nodes | Select Trails | Select State Nodes | Select Options

Include?	Trail ID	Target Node/et Class	Target Node/et ID	Target Node Count
<input checked="" type="checkbox"/>	targtype_1_bst	Unknown	Targets	25

Cancel < Back Next > Finish

Find: _____

CASO June 2020



Carnegie Mellon
IST Institute for Software Research

Importing Trail Data (7)

The screenshot shows the 'Import Data into ORA' dialog box in the 'Select State Nodes' step. The dialog has tabs for 'Select Ego Nodes', 'Select Trails', 'Select State Nodes', and 'Select Options'. The 'Select State Nodes' tab is active, displaying a list of node categories with checkboxes. The categories listed are: Business, Police, Airports & Airports, Government, Private Citizen, Refugee Camp, Government, Journalists & Journalists, Religious Figures, Military, Educational Institutions, Maritime, Telecommunications, Unknown, and Transportation. All 25 categories have their checkboxes checked. At the bottom of the list, it says '25 / 25 Selected, 25 / 25 Visible'. There are 'Select/Clear All' and 'Select/Clear Visible' buttons at the bottom left of the list area. The dialog also has 'Cancel', '< Back', 'Next >', and 'Finish' buttons at the bottom right.

Meta-Network Manager - No Meta-Network Loaded

Import Data into ORA

Select one or more Ego nodes to analyze by selecting the trails to consider and the states to pass through.

Select Ego Nodes Select Trails **Select State Nodes** Select Options

Targets: Targets

Node ID

- Business
- Police
- Airports & Airports
- Government
- Private Citizen
- Refugee Camp
- Government
- Journalists & Journalists
- Religious Figures
- Military
- Educational Institutions
- Maritime
- Telecommunications
- Unknown
- Transportation

25 / 25 Selected, 25 / 25 Visible

Select/Clear All Select/Clear Visible

Cancel < Back Next > Finish

Find: Larger sizes:

CASO

June 2020

Carnegie Mellon
IST Institute for Software Research

Importing Trail Data (8)

The screenshot shows the 'Import Data into ORA' dialog box in the 'Select Options' step. The dialog has tabs for 'Select Ego Nodes', 'Select Trails', 'Select State Nodes', and 'Select Options'. The 'Select Options' tab is active, displaying configuration options for dimensions and transitions. The options include: 'Dimensions to create' (set to 1), 'Transition duration time unit' (set to Days), 'State node separator', 'Remove gaps in the trail from filtered states', 'Create transitions meta-network' (checked, with 'Transitions-Days' selected), 'Create trails dynamic meta-network' (checked, with 'Trails-Days' selected), 'Start trails at same timestamp', and 'Restart trail interval' (set to 1). There are 'Cancel', '< Back', 'Next >', and 'Finish' buttons at the bottom right.

Meta-Network Manager - No Meta-Network Loaded

Import Data into ORA

Select one or more Ego nodes to analyze by selecting the trails to consider and the states to pass through.

Select Ego Nodes Select Trails Select State Nodes **Select Options**

Dimensions to create: 1 State node separator:

Transition duration time unit: Days

Remove gaps in the trail from filtered states

Create transitions meta-network: Transitions-Days

Create trails dynamic meta-network: Trails-Days

Start trails at same timestamp

Restart trail interval: 1

Cancel < Back Next > Finish

Find: Larger sizes:

CASO

June 2020



Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

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

CASOS
June 2020

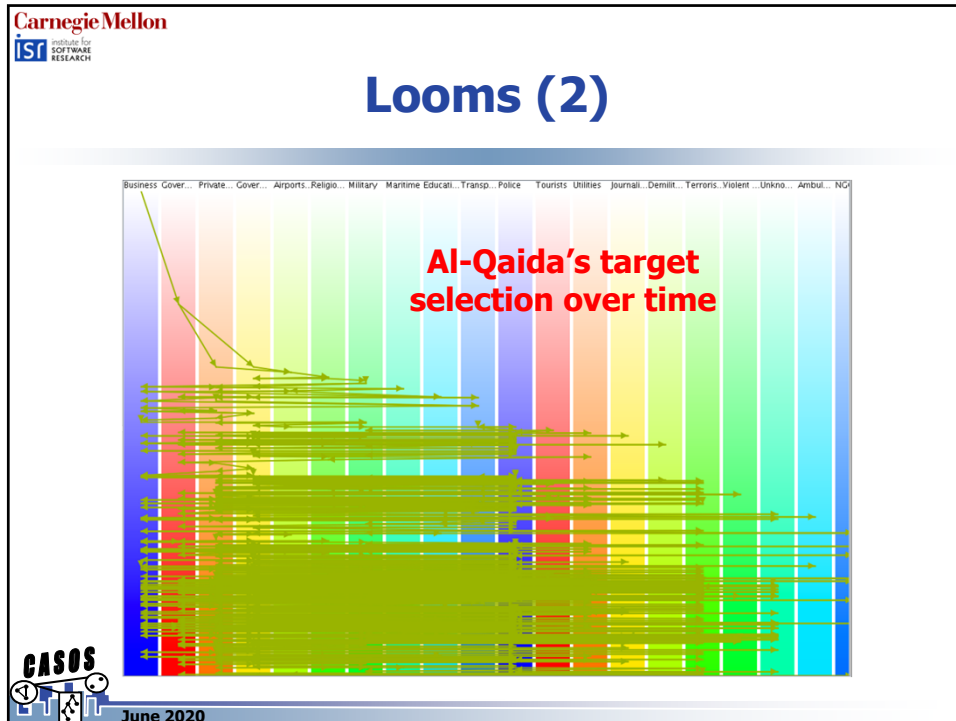
Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

Looms (1)

- Visualization depends on what we wish to observe
- Good indicator of timeline
- Sometimes cluttered

CASOS
June 2020





Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

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

CASOS
June 2020

Carnegie Mellon
IST Institute for Software Research

Networks From Trails (2)

Time	4 pm@Apr. 1	3 pm@Apr. 2	9 am@Apr. 3	1 pm@Apr. 3	2 pm@Apr. 4	4 pm@Apr. 5
Trail 1	F1	F2	F3	F2	F1	F2
Trail 2	F2	F3	F4	F2	F1	F1
Trail 3	F2	F3	F1	F1	F2	F3

	F1	F2	F3	F4
F1	2	3	0	0
F2	2	0	4	0
F3	1	1	0	1
F4	0	1	0	0

Traffic flow network

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

	F1	F2	F3	F4
F1	0.4	0.6	0	0
F2	0.33	0	0.67	0
F3	0.33	0.33	0	0.33
F4	0	1	0	0

Markov transition network

CASOS June 2020 15

Carnegie Mellon
IST Institute for Software Research

From Trails to Transitional Networks

- Observe ego's transitions from one state to another
- Aggregate the observed transitions
- Create probabilities from the aggregated values


CASOS June 2020



Carnegie Mellon
 ISI Institute for SOFTWARE RESEARCH

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.

CASOS
 June 2020

Carnegie Mellon
 ISI Institute for SOFTWARE RESEARCH

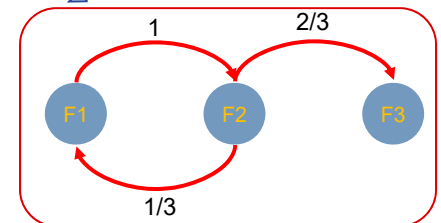
Example 1 – Function Calling

Function Caller	Function Called
F2	F3
F2	F1
F2	F3
F1	F2
F1	F2


Time ↓

Function Caller	Function Called
F1	F2
F2	F1
F1	F2
F2	F3
F2	F3

Time ↓



We lost the temporal component!

CASOS
 June 2020



Carnegie Mellon
IST Institute for Software Research

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)

CASOS
June 2020

Carnegie Mellon
IST Institute for Software Research

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

CASOS
June 2020



Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

Example 2 - Jihadist Groups (1)

Group Name	Target
ISIL	Business
Al-Qaida	Police
ISIL	Military
Al-Qaida	Military
Al-Qaida	Government (General)
ISIL	NGO
...	...

Time

CASOS June 2020

Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

Example 2 - Jihadist Groups (2)

Group Name	Target
ISIL	Business
Al-Qaida	Police
ISIL	Military
Al-Qaida	Military
Al-Qaida	Government
ISIL	NGO
...	...

Time

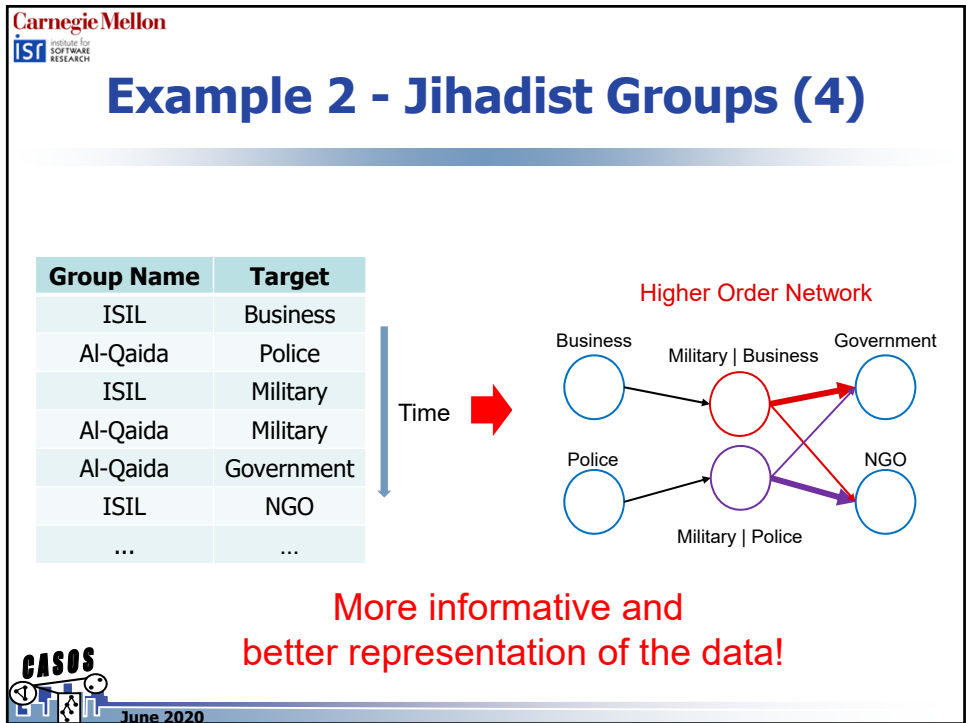
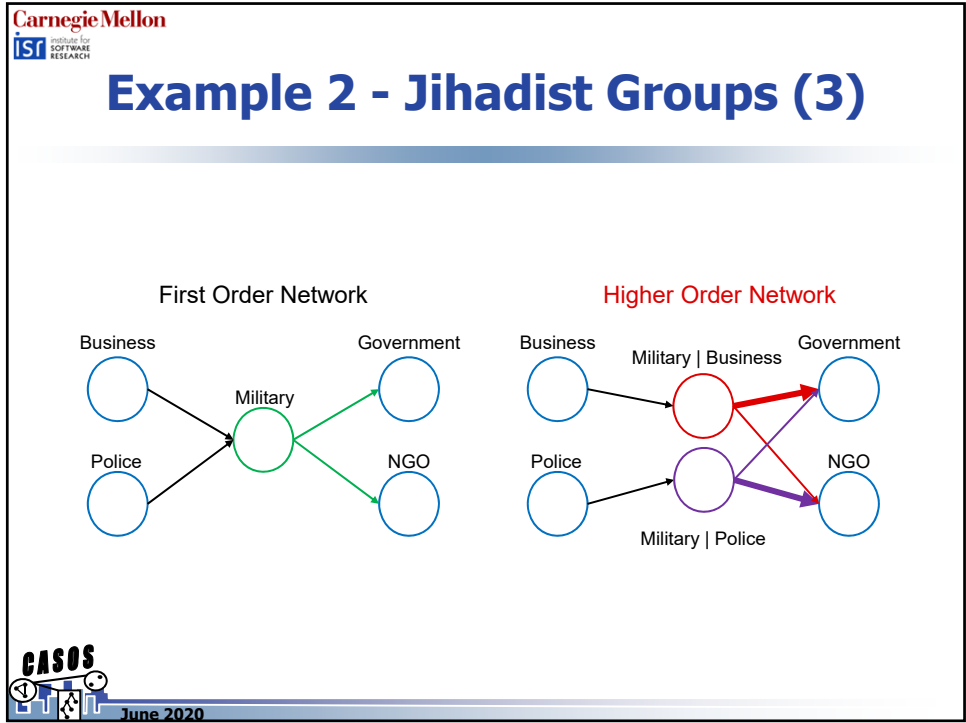
First Order Network

```

    graph LR
      Business((Business)) -- black --> Military((Military))
      Police((Police)) -- black --> Military
      Military -- green --> Government((Government))
      Military -- green --> NGO((NGO))
  
```

CASOS June 2020





Carnegie Mellon
IST Institute for Software Research

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.

CASOS
June 2020

Carnegie Mellon
IST Institute for Software Research

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

CASOS
June 2020



Carnegie Mellon
IST Institute for Software Research

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, ...

CASOS
June 2020

Carnegie Mellon
IST Institute for Software Research

Importing High-Dimensional Trails

ORA 3.0.0.0.07
File Edit Preferences Data Management Generate Networks Analysis Simulations Visualizations System Help

Meta-Network Manager No Meta-Network Loaded

Import Data into ORA

Select one or more Ego nodes to analyze by selecting the trails to consider and the states to pass through.

Select Ego Nodes Select Trails Select State Nodes Select Options

Dimensions to create: 1,2,3,4 State node separator: *

Transition duration time unit: Sequential

Remove gaps in the trail from filtered states

Create transitions meta-network: Transitions-Sequential

Create trails dynamic meta-network: Trails-Sequential

Start trails at same timestamp

Restart trail interval: 1

Cancel < Back Next > Finish

CASOS
June 2020

Carnegie Mellon
ISI Institute for SOFTWARE RESEARCH

Trail Report

ORA 3.0.9.9.87

File Edit Preferences Data Management Generate Networks Analysis Simulations Visualizations System Help

Generate Reports - Trails

Reports: select a report to run from the list or by category.

Trails

Description: Input Requirements | Output Formats

Load...

Analyzes data imported with the data import wizard's Trail Importer option.

Meta-Networks: select one or more to analyze in the report.

Select

- Transitions-Sequential - dimension 1
- Trails-Sequential - dimension 1
- Trails-Days - dimension 1
- Trails-Days - dimension 1
- Trails-Dataset

< Back Next > Cancel

Meta-Network Manager

- Transitions-Sequential - dimension 1
- Trails-Sequential - dimension 1
- Transitions-Days - dimension 1
- Trails-Days - dimension 1
- TrailsDataset

Find:

CASOS

June 2020

29

