



Trails and Networks: Loom; Network Representation of Trails

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/>



Overview

- What is a trail?
- How do we get trail data?
 - Characterize trail as network data
- Trails and Loom
 - Visualization
 - Networks from trails
 - Finding similar trails



June 2020

CASOS Summer Institute 2020

2



Carnegie Mellon
IST Institute for Software Research

What is a Trail?

- A **trail** is a **trace** of the **movement** of something **over time**
- For example, the movement of an attachment through a series of email communications creates a *trail*
- What are some other examples of trails?
 - People moving from place to place – geospatial trails
 - Twitter hashtags
 - ...

CASOS June 2020 CASOS Summer Institute 2020 3

Carnegie Mellon
IST Institute for Software Research

Geospatial Trails

- Usually geospatial trails represent agents travelling in continuous space and time.
- Network data: discrete node and discrete time.

Continuous space $\xrightarrow{\text{Aggregate}}$ Discrete location node

Continuous time $\xrightarrow{\text{Slice}}$ Discrete time

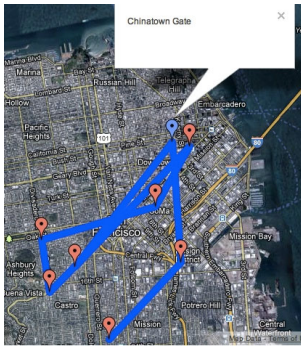
The diagram shows two vertical paths of red dots connected by arrows. The left path has three dots with arrows pointing down, up, and down. The right path has three dots with arrows pointing down, up, and down. The word 'vs' is placed between the two paths.

CASOS June 2020 CASOS Summer Institute 2020 4



Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

Geospatial Trails



Aggregate
Slice

Time	Location
2017, June 7, 9 am	Green St.
2017, June 7, 10 am	Design District
2017, June 7, 11 am	Chinatown Gate
2017, June 7, 12 am	16 th st.
.....

CASOS

June 2020 CASOS Summer Institute 2020 5

Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

Trails visualization

- ORA Over-time visualizer
 - Benefit: Can see changes in network structure over time
 - Drawback: For sparse trail data, not very effective
- ORA GIS Visualizer
 - Benefit: Can see the spatial distribution of trails
 - Drawback: Lose the temporal information
- Loom
 - Benefit: Can see the temporal distribution and the places travelled to
 - Drawback: Spatial distances, where they exist, are not preserved

CASOS


June 2020 CASOS Summer Institute 2020 6



Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

What we'll do


- Import a trail dataset with spatial information
- Visualization
 - Understand the benefits and drawbacks of different visualizations of trail data
 - ORA Over-time visualizer
 - ORA GIS visualizer
 - Loom
- Finding Similar trails
 - Use Loom to cluster trails
- Obtain networks from trails

 June 2020 CASOS Summer Institute 2020 7

Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

Import a dynamic meta-network

- Same as importing a regular meta-network
 - Drag-and-drop
 - File->Open Meta Network
- Import TrailsDataset.xml

 June 2020 CASOS Summer Institute 2020 8



Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

Importing

Trail Data

ORA 3.0.0.9.87

Meta-Network Manager

Dynamic-Meta-Network: TrailsDataset

Name: TrailsDataset

Filename: ... Dropbox/Data for Trail Training & Case Studies/Trail Data/TrailsDataset:

Generate Reports... Visualize Measure Charts... View Tra

Keyframes and deltas time stamped by: Date Period

Nodeset Agent nodes display by Node ID

Statistics:

Keyframe count: 109

Delta count: 0

Earliest date: 1603-01-01 00:00:00

Latest date: 2012-01-01 00:00:00

CASOS

June 2020 CASOS Summer Institute 2020 9

Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

The Data

- Our trail:
 - Locations are our nodes
 - Agents are what is moving between them
- Lets explore the data
 - In ORA
 - Networks over time visualizer
 - Geospatial visualizer

CASOS

June 2020 CASOS Summer Institute 2020 10



Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

ORA Main Window

The screenshot shows the ORA 3.0.0.9.8.7 interface. On the left is the Meta-Network Manager with a tree view of nodes from 1955 to 1989. On the right is the Network: Agent x Location window, which contains a table with columns for agents and locations.

	New York	Portland	Oakland	Montreal
The Commissioner	1	0	0	0
Mr. O'Brien	0	1	0	0
Vince Lombardi	0	0	1	0
Lord Stanley	0	0	0	1

Rows: 0 / 4 Selected, 4 / 4 Visible
Columns: 0 / 4 Selected, 4 / 4 Visible

CASOS
June 2020 CASOS Summer Institute 2020 11

Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

Networks Over Time Visualizer

The screenshot shows the ORA 3.0.0.9.8.7 interface with the Visualizations menu open. The 'View Networks Over Time...' option is circled in red. The background shows the Meta-Network Manager and a Dynamic-Meta-Network: Trails window with statistics.

Statistics:
Keyframe count: 109
Delta count: 0
Earliest date: 1903-01-01 00:00:00
Latest date: 2012-01-01 00:00:00

CASOS
June 2020 CASOS Summer Institute 2020 12



Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

Networks Over Time Visualizer

June 2020 CASOS Summer Institute 2020 13

Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

Geospatial Visualizer

June 2020 CASOS Summer Institute 2020 14



Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

Geospatial Visualizer

June 2020 CASOS Summer Institute 2020 15

Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

Geospatial Visualizer (older versions of ORA)

June 2020 CASOS Summer Institute 2020 16



Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

Geospatial Visualizer

June 2020 CASOS Summer Institute 2020 17

Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

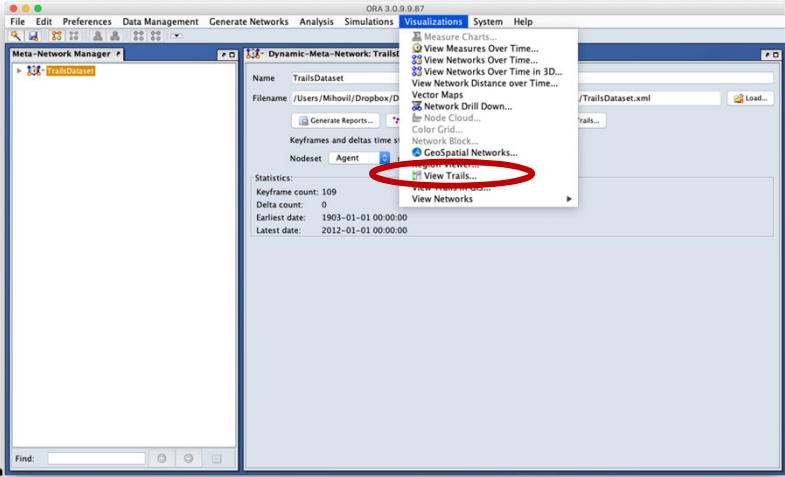
Geospatial Visualizer

June 2020 CASOS Summer Institute 2020 18



Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

Loom



The screenshot shows the Loom software interface. The 'Visualizations' menu is open, and the 'View Trails...' option is highlighted with a red circle. The interface includes a 'Meta-Network Manager' on the left, a central panel for 'Dynamic-Meta-Network: TrailsDataset', and a 'Statistics' section at the bottom. The statistics show a keyframe count of 109, a delta count of 0, and dates from 1903-01-01 to 2012-01-01.

ORA 3.0.9.9.87
File Edit Preferences Data Management Generate Networks Analysis Simulations Visualizations System Help

Meta-Network Manager
TrailsDataset

Dynamic-Meta-Network: TrailsDataset
Name TrailsDataset
Filename /Users/Mihovil/Dropbox/D...
Generate Reports...
Keyframes and deltas time stamped by: Date Period
Nodeset Agent
nodes display by Node ID

Statistics:
Keyframe count: 109
Delta count: 0
Earliest date: 1903-01-01 00:00:00
Latest date: 2012-01-01 00:00:00

Visualizations menu items:
Measure Charts
View Measures Over Time...
View Networks Over Time...
View Network Distance over Time...
Vector Maps
Network Drill Down...
Node Cloud...
Color Grid...
Network Back...
GeoSpatial Networks...
View Trails...
View Networks

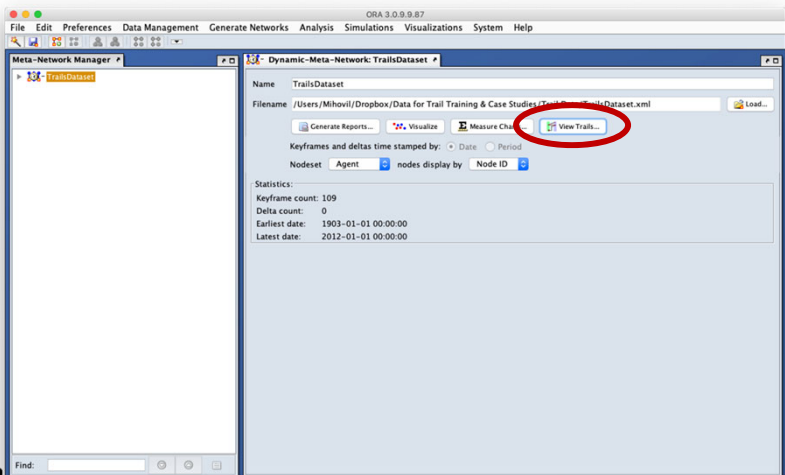
Find:

CASOS

June 2020 CASOS Summer Institute 2020 19

Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

Loom



The screenshot shows the Loom software interface. The 'View Trails...' button is highlighted with a red circle. The interface includes a 'Meta-Network Manager' on the left, a central panel for 'Dynamic-Meta-Network: TrailsDataset', and a 'Statistics' section at the bottom. The statistics show a keyframe count of 109, a delta count of 0, and dates from 1903-01-01 to 2012-01-01.

ORA 3.0.9.9.87
File Edit Preferences Data Management Generate Networks Analysis Simulations Visualizations System Help

Meta-Network Manager
TrailsDataset

Dynamic-Meta-Network: TrailsDataset
Name TrailsDataset
Filename /Users/Mihovil/Dropbox/Data for Trail Training & Case Studies/TrailsDataset.xml
Generate Reports... Visualize Measure Ch... View Trails...
Keyframes and deltas time stamped by: Date Period
Nodeset Agent nodes display by Node ID

Statistics:
Keyframe count: 109
Delta count: 0
Earliest date: 1903-01-01 00:00:00
Latest date: 2012-01-01 00:00:00

Find:

CASOS

June 2020 CASOS Summer Institute 2020 20



Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

Loom

The screenshot shows the Loom software interface. At the top, there are menu options: File, Options, Clustering, and Events. Below the menu is a list of locations: Ottawa, Quebec, Toronto, Vancouver, Montreal, Seattle, Victoria, New York, Boston, Chicago, Detroit, Philadelphia, Edmonton, Calgary, Pittsburgh, Newark, Denver, Dallas, and San Francisco. On the left, there is a 'Cluster' panel with a 'Find:' field and a list of agents: Lord Stanley, Mr. O'Brien, The Commissioner, and Vince Lombardi. Below that is a 'Location' panel with colored dots for each location. The main area is a grid of colored vertical bars representing time for each location. Green arrows represent trails between locations over time. Two red circles highlight specific clusters of trails. The bottom left corner has the CASOS logo and the date June 2020. The bottom right corner has the text CASOS Summer Institute 2020 and the page number 21.

Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

Trails and Loom

- **Visualization** over time is hard
 - State of the art revolves around animation
 - Loom allows us to visualize trails over time in a static, understandable environment
- Trails may have similar **patterns**, but these are difficult to observe
 - Loom allows us to cluster similar trails together
- We can get **networks from trails**, for example, who is connected by the given attachment?
 - Loom allows us to easily export such networks to ORA

CASOS
June 2020 CASOS Summer Institute 2020 22



Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

What we'll do

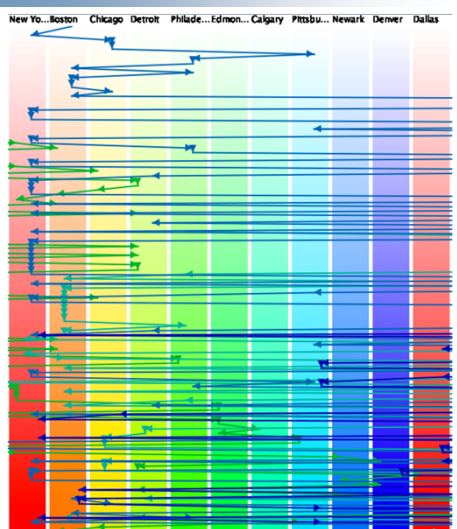
- ~~Import a "DynamicMetaNetwork" with spatial information~~
- ~~Visualization~~
 - ~~Understand the benefits and drawbacks of different visualizations of trail data~~
 - ~~ORA Over-time visualizer~~
 - ~~ORA GIS visualizer~~
 - ~~Loom~~
- **Finding Similar trails**
 - **Use Loom to cluster trails**
- Obtain networks from trails

CASOS
June 2020 CASOS Summer Institute 2020 23

Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

Why cluster?

- Why are we interested in trails and trail clustering?
 - Gain information by analyzing agents across space and time together.
 - Interested in grouping agents that display same behavior across time. E.g. visit the same locations across time.



CASOS
June 2020 CASOS Summer Institute 2020 24



Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

Feature vector representation using PFSA

βααβαββααββααββα.....

Depth = 1

Depth = 2

Depth = 3

$$\vec{p} = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{bmatrix} \quad \Pi = \begin{bmatrix} \pi_{11} & \pi_{12} & 0 & 0 \\ 0 & 0 & \pi_{23} & \pi_{24} \\ \pi_{31} & \pi_{32} & 0 & 0 \\ 0 & 0 & \pi_{43} & \pi_{44} \end{bmatrix}$$

State Probability Vector State Transition matrix

CASOS

June 2020 CASOS Summer Institute 2020 25

Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

Clustering of Trails using PFSA

- Each trail is now represented by a numerical feature vector, the state probability vector of the derived PFSA (the model of the generative process).
- To look at joint spatiotemporal behavior we now cluster the agent trails based on their feature vectors.
- This is done using a two step process.
 - A coarse clustering step : Trails are initially grouped coarsely according to the locations visited, irrespective of the frequency of the visits.
 - A cluster refining step : The coarse clusters are each then clustered using agglomerative clustering to derive groups of trails which visit "similar" locations with "similar" frequencies.

CASOS

June 2020 CASOS Summer Institute 2020 26



Carnegie Mellon
IST Institute for Software Research

Refining the Coarse Clustering

Depth = 1

Depth = 2

Depth = 3

Cluster

Trail pattern depth: 1

Cluster quality: 1.0

Clear Compute

CASOS

June 2020 CASOS Summer Institute 2020 27

Carnegie Mellon
IST Institute for Software Research

Viewing time sequences

- Each cluster contains trails with similar patterns in the sequences of locations visited
- Thus extract the longest common subsequence amongst all the trails belonging to a cluster.

BANANA
ATANA

ANA AANA

Longest common string Longest common subsequence

New York St. Louis San Franc...

CASOS

June 2020 CASOS Summer Institute 2020 28



Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

What we'll do

- ~~Import a "DynamicMetaNetwork" with spatial information~~
- ~~Understand the benefits and drawbacks of different visualizations of trail data~~
 - ~~ORA Over-time visualizer~~
 - ~~ORA GIS visualizer~~
 - ~~Loom~~
- ~~Use Loom to cluster similar trails~~
 - ~~The high-level concept~~
 - ~~The details~~
- **Obtain networks from trails**

CASOS
June 2020 CASOS Summer Institute 2020 29

Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

Generating Networks from Trails

- We can better understand how different cities relate via championships by getting networks out of them

What we'll do

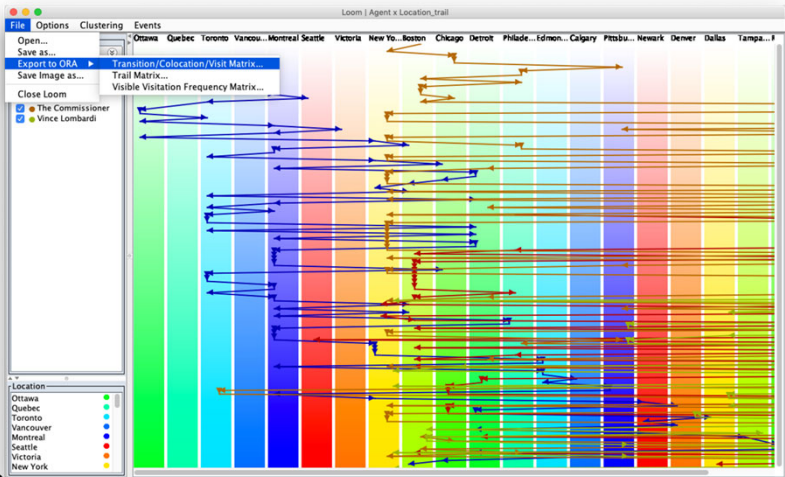
- Generate the networks
- View them in ORA
- Use ORA Network Visualizer

CASOS
June 2020 CASOS Summer Institute 2020 30



Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

Exporting the Matrices



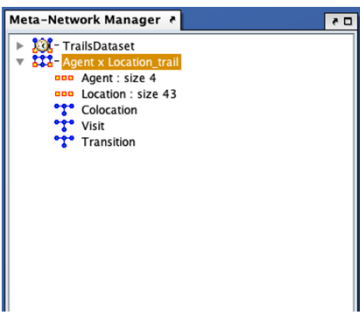
The screenshot shows the Loom software interface. The 'Export to ORA' menu is open, showing options: 'Transition/Colocation/Visit Matrix...', 'Trail Matrix...', and 'Visible Visitation Frequency Matrix...'. The main window displays a visualization of agent trails between cities, with arrows indicating movement between locations. The cities listed in the 'Location' panel are Ottawa, Quebec, Toronto, Vancouver, Montreal, Seattle, Victoria, and New York. The trails are color-coded by city.

CASOS

June 2020 CASOS Summer Institute 2020 31

Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

What we now have



The screenshot shows the Meta-Network Manager interface. The tree view shows the following structure:

- TrailsDataset
 - Agent x Location_trail
 - Agent : size 4
 - Location : size 43
 - Colocation
 - Visit
 - Transition

- ORA uses all of the trails and outputs a single meta-network
 - **Colocation** – An edge is created between the trophies if they ever existed at the same place at the same time
 - **Visit Matrix** – An edge is created between city and trophy if the city ever won that trophy
 - **Transition** – An edge is created between cities if a trophy ever traveled from one to the other in consecutive years



Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

Colocation

Agent x Location_trail

```
graph TD; Commissioner --- Lombardi; Commissioner --- OBrien; Commissioner --- Stanley; Lombardi --- OBrien; Lombardi --- Stanley; OBrien --- Stanley;
```

powered by ORA-NetScapes, CASOS Center @ CMU

CASOS
June 2020 CASOS Summer Institute 2020 33

Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

Transition

CASOS
June 2020 CASOS Summer Institute 2020 34





Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

Summary

- We discussed what a trail was – a trace of the movement of something through a network over time
- We used an example dataset and looked at trail data three different ways – in the Networks Over Time visualizer, the GIS visualizer and Loom
- We talked about how to find similar trails in Loom
- We looked at how we can get new, interested networks out of our trail data

CASOS

June 2020 CASOS Summer Institute 2020 36

