



## Trails and Networks: Loom; Going from Trails to Networks and Networks to Trails

Kenny Joseph  
Aparna Gullapalli

[kjoseph@cs.cmu.edu](mailto:kjoseph@cs.cmu.edu)

[aparna@andrew.cmu.edu](mailto:aparna@andrew.cmu.edu)



**CarnegieMellon**

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



## Overview

- An initial detour – a characterization of network data
- What is a trail?
- How do we get trail data?
- Trails and Loom
  - Visualization
  - Networks from trails
  - Finding similar trails



6.15.2016

© 2012 CASOS, Director Kathleen M. Carley

2



Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

## Types of network data

- Three broad flavors
  - Static – data on relations from a single time period
  - Panel – data on the same set of people collected at intermittent points in time
  - Relational event data – timestamped data that relates two things

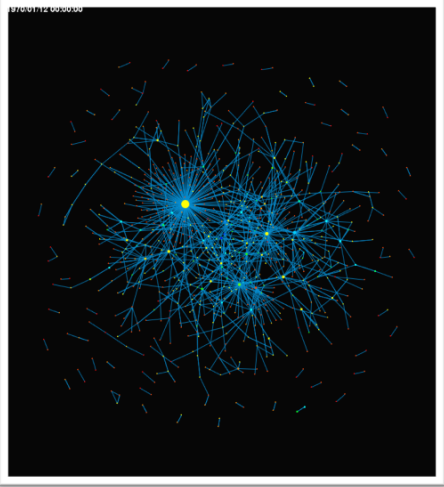
D  
y  
n  
a  
m  
i  
c

CASOS

6.15.2016 © 2012 CASOS, Director Kathleen M. Carley 3

Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

## Static Data



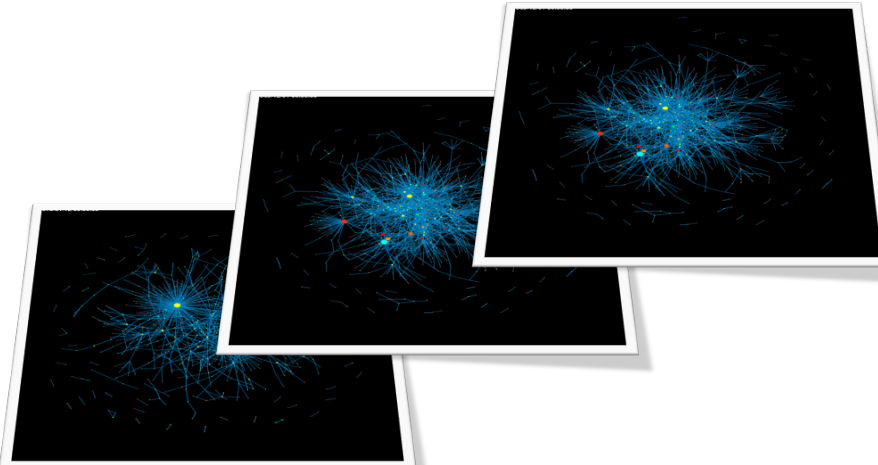
CASOS

6.15.2016 © 2012 CASOS, Director Kathleen M. Carley 4



Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

## Panel Data



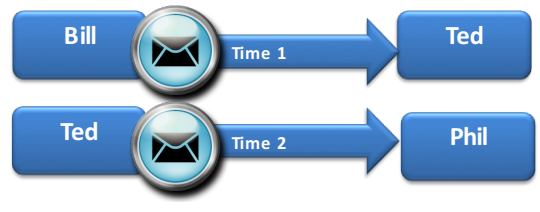
The figure displays three network graphs arranged in a staggered, overlapping sequence from bottom-left to top-right. Each graph shows a central node with several other nodes connected to it by edges, representing a network structure that evolves over time. The nodes are represented by small blue dots, and the edges are thin blue lines. The background of each graph is black.

CASOS

6.15.2016 © 2012 CASOS, Director Kathleen M. Carley 5

Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

## Event Data



The diagram shows two sequential events. The first event is labeled 'Time 1' and shows a blue arrow pointing from a box labeled 'Bill' to a box labeled 'Ted'. A circular icon containing an envelope symbol is positioned between the two boxes. The second event is labeled 'Time 2' and shows a blue arrow pointing from a box labeled 'Ted' to a box labeled 'Phil'. A circular icon containing an envelope symbol is positioned between the two boxes.

Event data allows us to think about an **unknown underlying network** which may change over time

CASOS

6.15.2016 © 2012 CASOS, Director Kathleen M. Carley 6



Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

## Event Data (cont.)

- Sometimes we want a static picture of the network (or a panel)... how can we get that from event data?
- **Aggregate**
- \*\*Aggregation is necessary to obtain a *static* network from event data, but not to have a network
  - Inference on the event stream itself is possible and useful

CASOS  
6.15.2016 © 2012 CASOS, Director Kathleen M. Carley 7

Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

## What problems do we run into here?

The diagram illustrates a sequence of two events. The first event, labeled 'Time 1', shows a blue box labeled 'Bill' sending an email (represented by an envelope icon) to a blue box labeled 'Ted'. The second event, labeled 'Time 2', shows a blue box labeled 'Ted' sending an email to a blue box labeled 'Phil'. A red arrow points from these two events towards a network graph visualization on the right, which shows a dense, star-like network of blue lines on a black background, with a central yellow node.

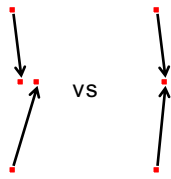
CASOS  
6.15.2016 © 2012 CASOS, Director Kathleen M. Carley 8



Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

## Spatial Review

- Granularity of observations
  - Aggregation issues
    - Disconnect between space and networks
      - Locations are continuous
      - Networks defined between discrete entities
      - Same location versus 10 meters apart
  - Resolution issues
    - Mismatch between problem and data
      - Observe people in cities
      - Want to know most central country
  - Scale issues
    - Analysis at certain scales becomes intractable as the amount of data increases



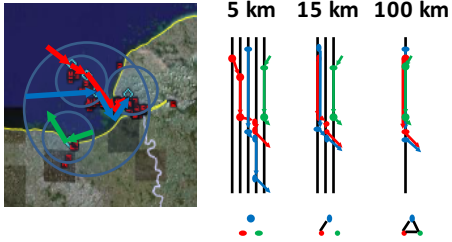
CASOS

6.15.2016 © 2012 CASOS, Director Kathleen M. Carley 9

Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

## Getting a network from spatial events

- Aggregation assumes that at some level
  - Space is discrete at some level
  - All events happen more or less at the same place



5 km 15 km 100 km

What do we miss here if we aggregate by 100 km?

CASOS

6.15.2016 © 2012 CASOS, Director Kathleen M. Carley 10



Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

## Getting a network from relational events

- Aggregation assumes that at some level
  - order doesn't matter
  - All events happen more or less at the same time

What do we miss here if we aggregate by three time steps?

CASOS  
6.15.2016 © 2012 CASOS, Director Kathleen M. Carley 11

Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

## Uses of Aggregation

- Examples of when you **would** want to aggregate:
  - It doesn't matter when I exercised today, the point is that I did it!
- Examples of when you **wouldn't** want to aggregate
  - There is a big difference between when I work out before work versus after

CASOS  
6.15.2016 © 2012 CASOS, Director Kathleen M. Carley 12



Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

## So – What’s a Trail?

- A **trail** is a **trace** of the **movement** of something **over time**
- Thus, 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
  - Twitter hashtags
  - ...

CASOS  
6.15.2016 © 2012 CASOS, Director Kathleen M. Carley 13

Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

## Trail data from network data

- So, we have networks - how do we get trail data?
  - Static - Simulation
  - Panel - From simulation and from the data, though time between observations is inferred
  - “Event data” - from simulation and **directly from the data**

D  
y  
n  
a  
m  
i  
c

CONSTRUCT

Bill → Ted

Ted → Phil

CASOS  
6.15.2016 © 2012 CASOS, Director Kathleen M. Carley 14



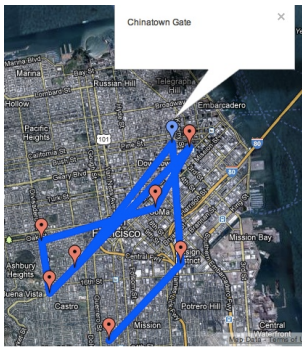
Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

## Event Data and trails

- In a series of relational email events, *information* may flow
- Today – look at geospatial trails: agents travelling to different locations
- What’s the network?
- What the object that is moving?

CASOS  
6.15.2016 © 2012 CASOS, Director Kathleen M. Carley 15

Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH



Looking at trails and generating networks from them

## LOOM

CASOS  
6.15.2016 © 2012 CASOS, Director Kathleen M. Carley 16






Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

## Trails and Loom


- **Visualization** of things 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

 6.15.2016 © 2012 CASOS, Director Kathleen M. Carley 17

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

 6.15.2016 © 2012 CASOS, Director Kathleen M. Carley 18



Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

## Import a dynamic meta-network

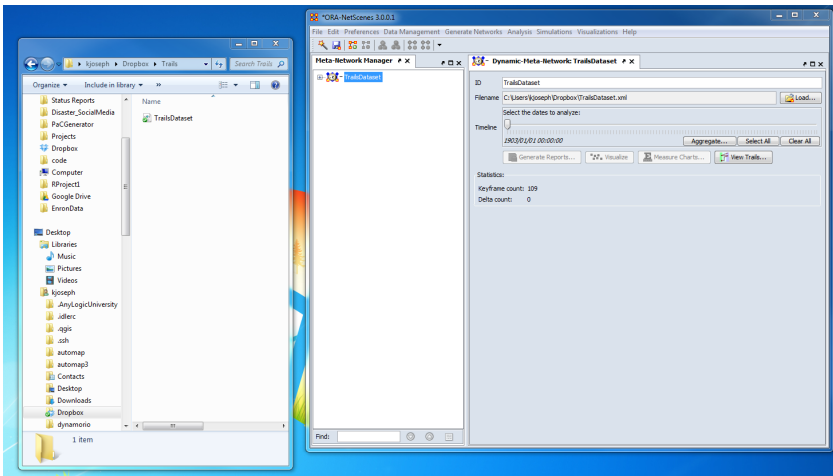
- Same as importing a regular meta-network
  - Drag-and-drop
  - File->Open Meta Network
- Import TrailsDataset.xml
- NOTE: This is actually panel data, but we'll think of it as event data

CASOS

6.15.2016 © 2012 CASOS, Director Kathleen M. Carley 19

Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

## Importing



NetScout 3.0.0.1

Dynamic-Meta-Network-TrailsDataset

ID: TrailsDataset

Filename: C:\Users\joseph\Dropbox\TrailsDataset.xml

Timefile: 20130202\_010000

Statistics:  
Keyframe count: 309  
Delta count: 0

CASOS

6.15.2016 © 2012 CASOS, Director Kathleen M. Carley 20



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 proper
  - Networks over time visualizer
  - Geospatial Visualizer

CASOS  
6.15.2016 © 2012 CASOS, Director Kathleen M. Carley 21

Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

## ORA Proper

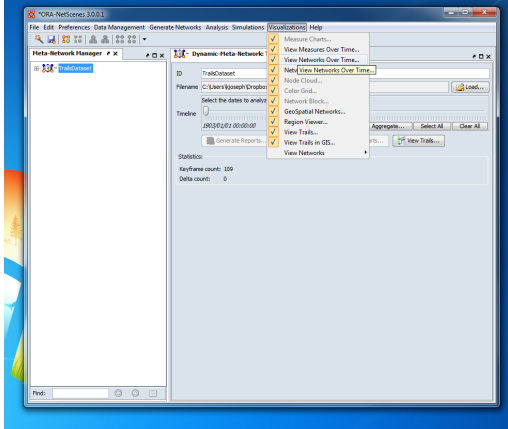
Search nodes	Pittsburgh	Seattle	Montreal
The Commissioner	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Mr. O'Brien	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Vince Lombardi	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Lord Stanley	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

CASOS  
6.15.2016 © 2012 CASOS, Director Kathleen M. Carley 22



Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

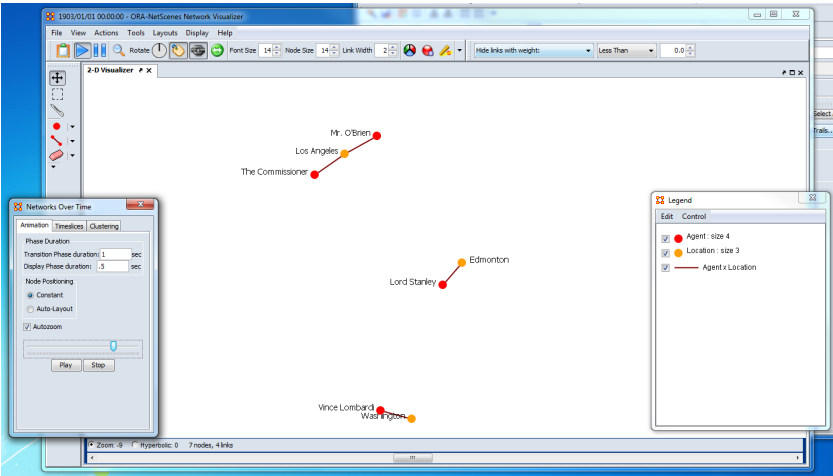
# Networks Over Time Visualizer



6.15.2016 © 2012 CASOS, Director Kathleen M. Carley 23

Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

# Networks Over Time Visualizer



6.15.2016 © 2012 CASOS, Director Kathleen M. Carley 24



Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

# Geospatial Visualizer

6.15.2016 © 2012 CASOS, Director Kathleen M. Carley 25

Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

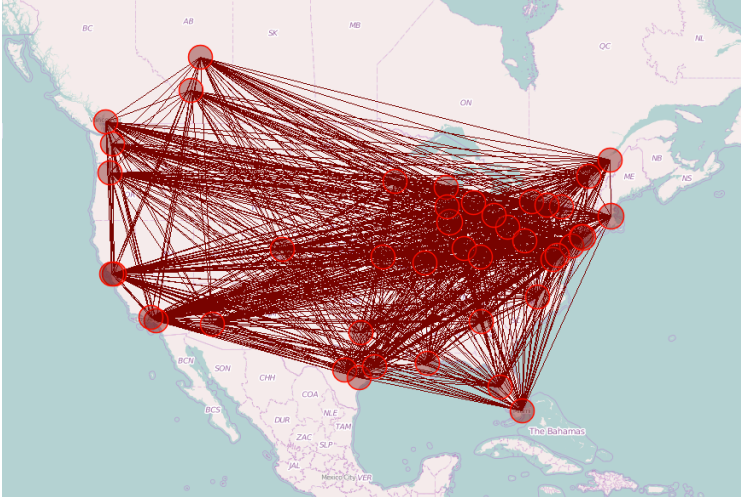
# Geospatial Visualizer

6.15.2016 © 2012 CASOS, Director Kathleen M. Carley 26



Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

## Geospatial Visualizer

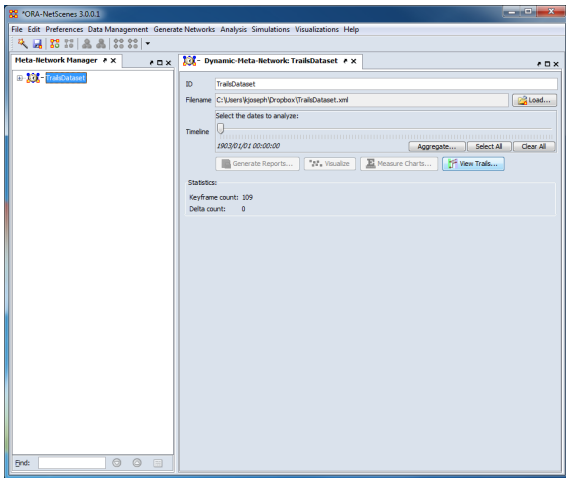


CASOS

6.15.2016 © 2012 CASOS, Director Kathleen M. Carley 27

Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

## Loom



Dynamic: Meta-Network: TrailDataset

Filename: C:\Users\joseph\Dropbox\TrailDataset.xml

Select the dates to analyze:

Timeline: 1/30/2012 00:00:00

Generate Reports... Visualize Measure Charts... View Trails...

Statistics:

Keyframe count: 309

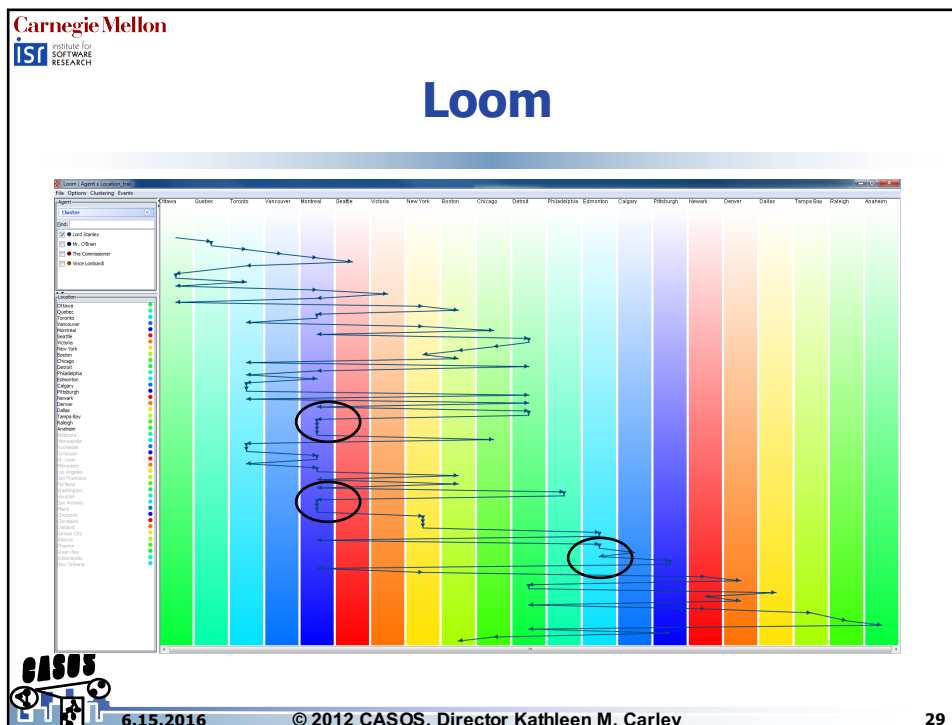
Delta count: 0

End:

CASOS

6.15.2016 © 2012 CASOS, Director Kathleen M. Carley 28





**View features for trail data**

- 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

The slide includes the Carnegie Mellon IST logo at the top left, the title "View features for trail data" in the center, and a list of features and drawbacks for three visualization methods. The bottom of the slide features the CASOS logo, the date "6.15.2016", the copyright notice "© 2012 CASOS, Director Kathleen M. Carley", and the page number "30".



Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

## What we'll do

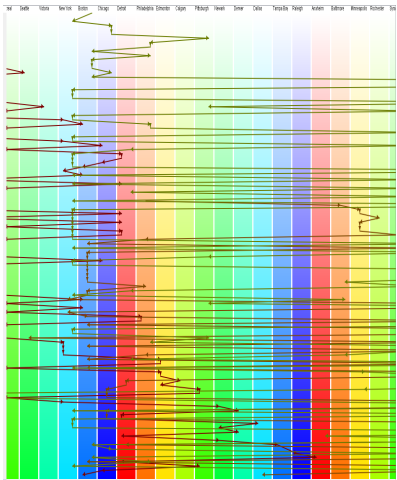
- Import a "DynamicMetaNetwork" with spatial information
- Understand the benefits and drawbacks of different visualizations of trail data
  - ORA Over-time visualizer
  - ORA GIS visualizer
  - Loom
- **Use Loom to cluster similar trails**
  - We'll cover this at a high level - for details, please ask after!
- Obtain networks from trails

CASOS  
6.15.2016 © 2012 CASOS, Director Kathleen M. Carley 31

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  
6.15.2016 © 2012 CASOS, Director Kathleen M. Carley 32





Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

## Feature vector representation using PFSA

βαααβαββααββααββα.....

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

6.15.2016      © 2012 CASOS, Director Kathleen M. Carley      33

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

6.15.2016      © 2012 CASOS, Director Kathleen M. Carley      34



Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

## Refining the Coarse Clustering

Depth = 1

Depth = 2

Depth = 3

CASOS

6.15.2016 © 2012 CASOS, Director Kathleen M. Carley 35

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  
Longest common string

AANA  
Longest common subsequence

New York Boston Pittsburgh Baltimore St. Louis Oakland

CASOS

6.15.2016 © 2012 CASOS, Director Kathleen M. Carley 36



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  
6.15.2016 © 2012 CASOS, Director Kathleen M. Carley 37

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

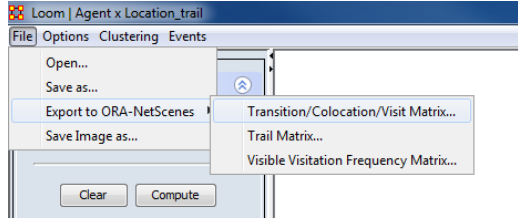
- Generate the networks
- View them in ORA Proper
- Use ORA Network Visualizer

CASOS  
6.15.2016 © 2012 CASOS, Director Kathleen M. Carley 38



Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

## Exporting the Matrices

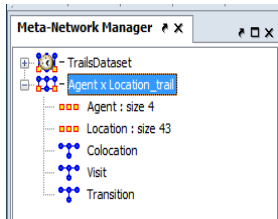


The screenshot shows the Loom application window titled 'Loom | Agent x Location\_trail'. The 'File' menu is open, and the 'Export to ORA-NetScenes' option is selected, which has opened a sub-menu. The sub-menu contains three options: 'Transition/Colocation/Visit Matrix...', 'Trail Matrix...', and 'Visible Visitation Frequency Matrix...'. Below the menu, there are 'Clear' and 'Compute' buttons.

CASOS  
6.15.2016 © 2012 CASOS, Director Kathleen M. Carley 39

Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

## What we now have



The screenshot shows the 'Meta-Network Manager' window. It displays a tree view of network components. The root is 'TrailsDataset', which contains 'Agent x Location\_trail'. Under 'Agent x Location\_trail', there are 'Agent : size 4' and 'Location : size 43'. Below these are 'Colocation', 'Visit', and 'Transition'.

- ORA uses the entire trailset 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

CASOS  
6.15.2016 © 2012 CASOS, Director Kathleen M. Carley 40



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 DRA-NetScenes, CASOS Center @ CMU

**CASOS**  
6.15.2016 © 2012 CASOS, Director Kathleen M. Carley 41

Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

# Transition

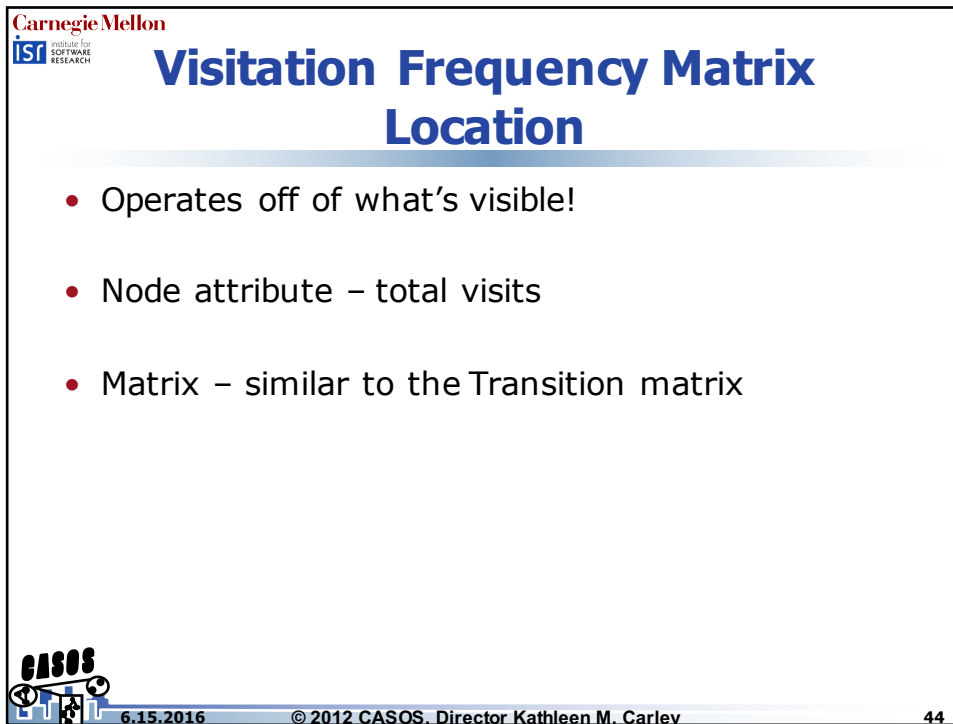
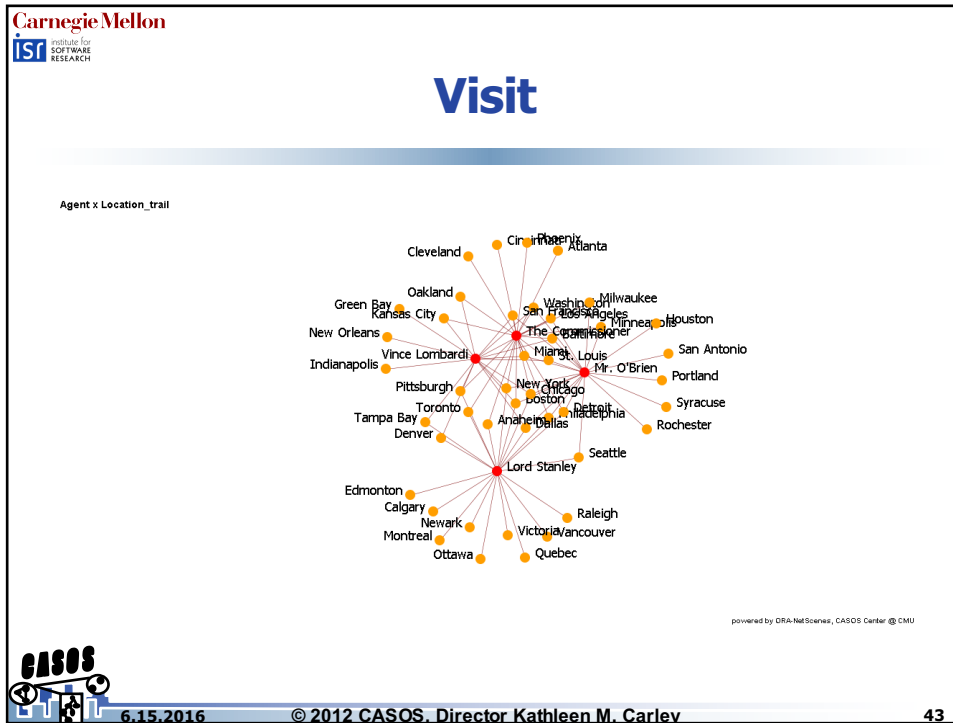
Agent x Location\_trail

```
graph TD; NewOrleans --- GreenBay; GreenBay --- Edmonton; Edmonton --- Calgary; Calgary --- SanFrancisco; SanFrancisco --- Oakland; Oakland --- Indianapolis; Indianapolis --- Baltimore; Baltimore --- Milwaukee; Milwaukee --- Cincinnati; Cincinnati --- Cleveland; Cleveland --- Minneapolis; Minneapolis --- Rochester; Denver --- Dallas; Dallas --- Washington; Washington --- Chicago; Chicago --- Detroit; Detroit --- StLouis; StLouis --- LosAngeles; LosAngeles --- NewYork; NewYork --- Philadelphia; Philadelphia --- KansasCity; KansasCity --- Toronto; Toronto --- Ottawa; Ottawa --- Atlanta; Atlanta --- Phoenix; Phoenix --- Syracuse; Houston --- Newark; Newark --- SanAntonio; SanAntonio --- TampaBay; TampaBay --- Seattle; Seattle --- Victoria; Victoria --- Raleigh; Raleigh --- Miami; Miami --- Montreal; Montreal --- Anaheim; Anaheim --- Vancouver; Vancouver --- Boston;
```

powered by DRA-NetScenes, CASOS Center @ CMU

**CASOS**  
6.15.2016 © 2012 CASOS, Director Kathleen M. Carley 42





Carnegie Mellon  
ISI Institute for SOFTWARE RESEARCH

## Visitation Frequency Matrix Location

Trail Visitation Frequency Matrix - ORA-NetSchemas Network Visualizer

File View Actions Tools Layouts Display Help

Font Size: 18 Node Size: 14 Link Width: 0.5

Hide links with weights: Less Than 0.0

Node Size Selector

Use this window to manipulate nodes by measure value or attribute.

Select an attribute: visitation count

Select a measure: <select>

Expand to control individual values.

Apply Changes Close

Legend

Location: size 40

Transitions

Zoom: -6 Hyperbolic: 0 40 nodes, 141 links

CASOS

6.15.2016 © 2012 CASOS, Director Kathleen M. Carley 45

Carnegie Mellon  
ISI Institute for SOFTWARE RESEARCH

## Summary

- We discussed what a trail was – a trace of the movement of something through a network over time
- We talked about how we can get network data from static, panel and event-level data, and then how we can get trail data from there
- We used an example event-level (actually panel) dataset and looked at trail data three different ways – in the Networks Over Time visualizer, the GIS visualizer and Loom

CASOS

6.15.2016 © 2012 CASOS, Director Kathleen M. Carley 46



## Summary (cont.)

- 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