<Your Name>

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

Kenny Joseph
Aparna Gullapalli

kjoseph@cs.cmu.edu aparnag@andrew.cmu.edu

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

Carnegie Mellon

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

<Your Name>

Carnegie Mellon
ISR 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

Dynamic

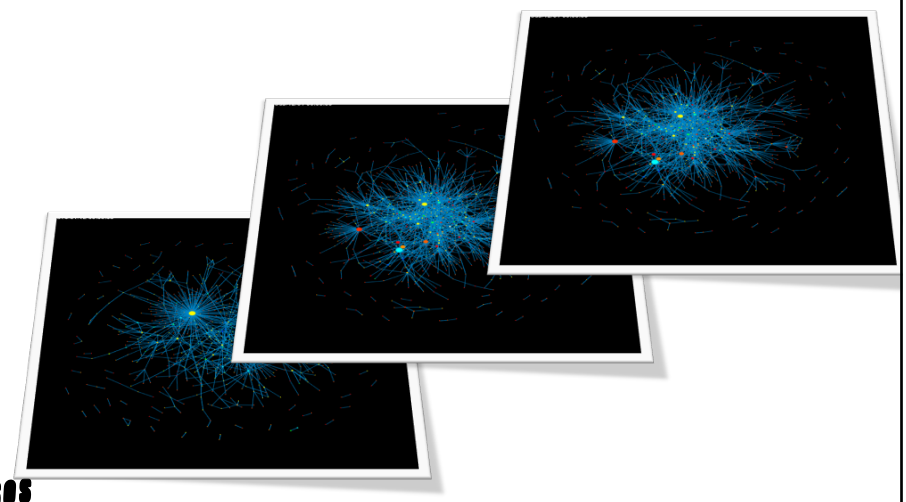
Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

Static Data

<Your Name>

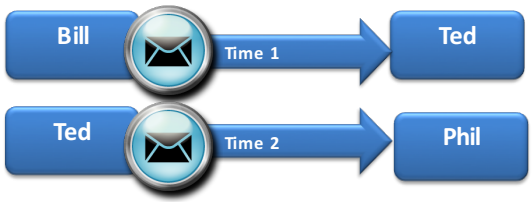
Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

Panel Data



Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

Event Data



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

<Your Name>

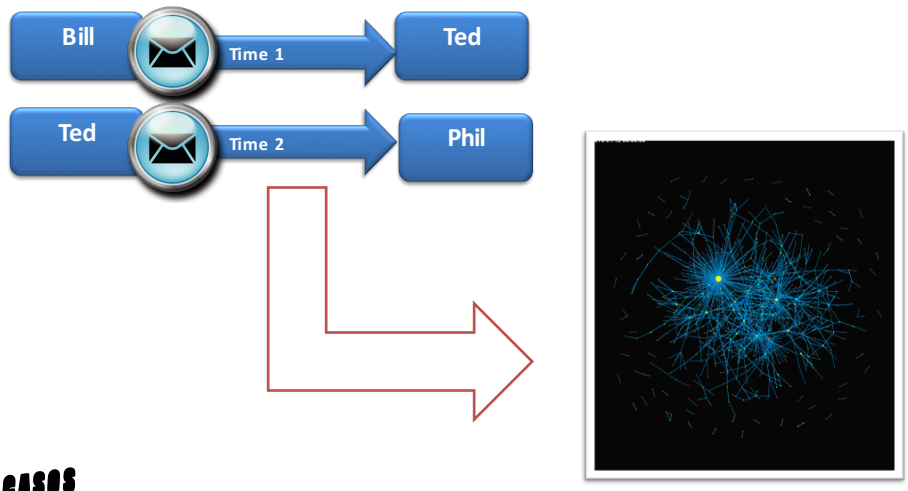
Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

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

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

What problems do we run into here?

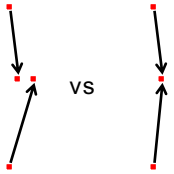


<Your Name>

Carnegie Mellon
ISR 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

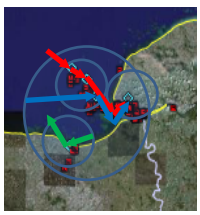


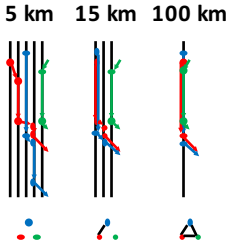
  - Resolution issues
    - Mismatch between problem and data
      - Observe people in cities
      - Want to know most central country
  - Scale issues
    - Analysis at certain scales becomes intractable as the amount of data increases

Carnegie Mellon
ISR 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





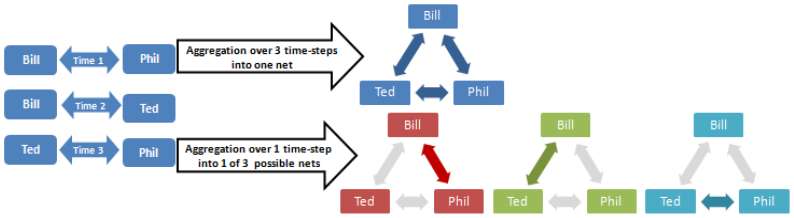
What do we miss here if we aggregate by 100 km?

<Your Name>

Carnegie Mellon
ISR 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?

Carnegie Mellon
ISR 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

<Your Name>

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

So – What's a Trail?

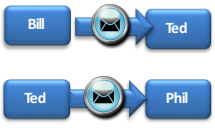
- A **trail** is a **trace** of the **movement** of something **over time**
- Thus, for example, the movement of an attachment through a series of email communications creates a *trail*
- What are some other examples of trails?
  - People moving from place to place
  - Twitter hashtags
  - …

Carnegie Mellon
ISR 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**

Dynamic



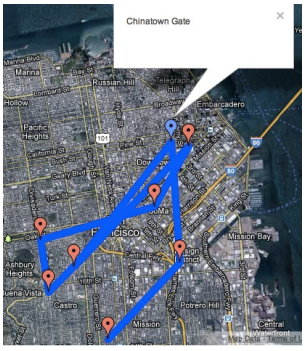
<Your Name>

Carnegie Mellon
ISR 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?

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH



Looking at trails and generating networks from them

LOOM

<Your Name>

Carnegie Mellon
ISR 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

Carnegie Mellon
ISR 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

<Your Name>

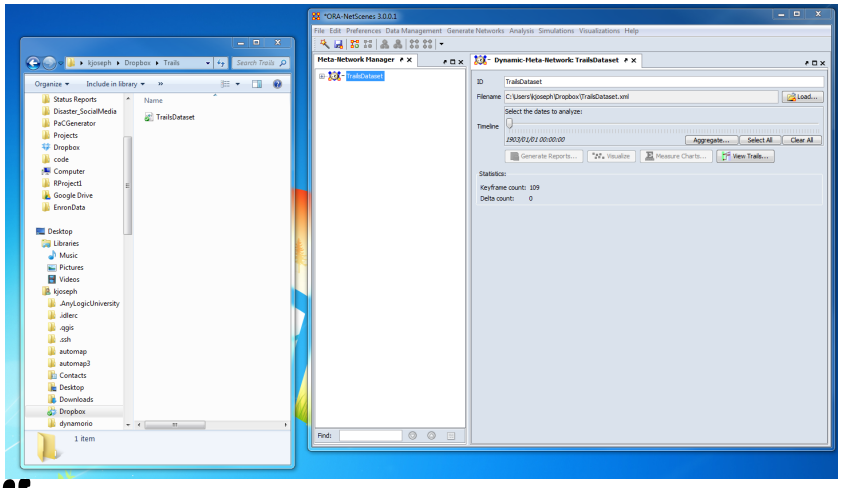
Carnegie Mellon
institute for SOFTWARE RESEARCH

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

Carnegie Mellon
institute for SOFTWARE RESEARCH

Importing



<Your Name>

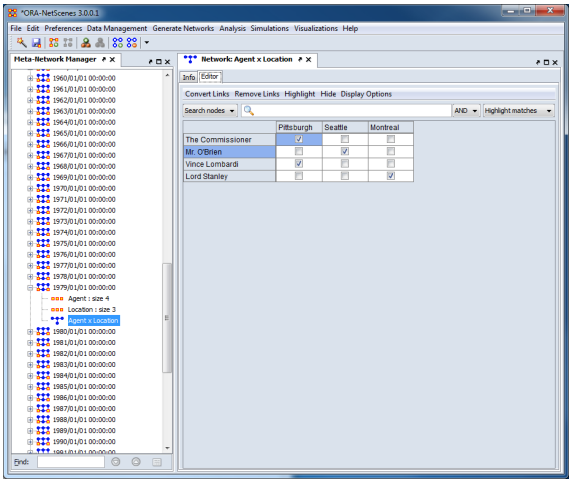
Carnegie Mellon
institute for SOFTWARE RESEARCH

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

Carnegie Mellon
institute for SOFTWARE RESEARCH

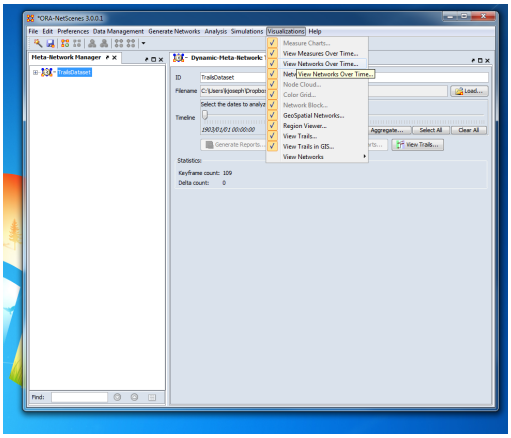
ORA Proper



<Your Name>

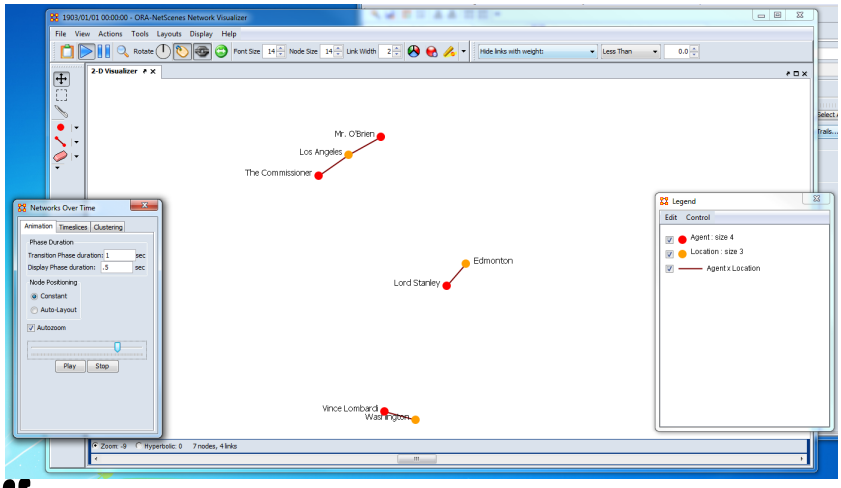
Carnegie Mellon
institute for SOFTWARE RESEARCH

Networks Over Time Visualizer



Carnegie Mellon
institute for SOFTWARE RESEARCH

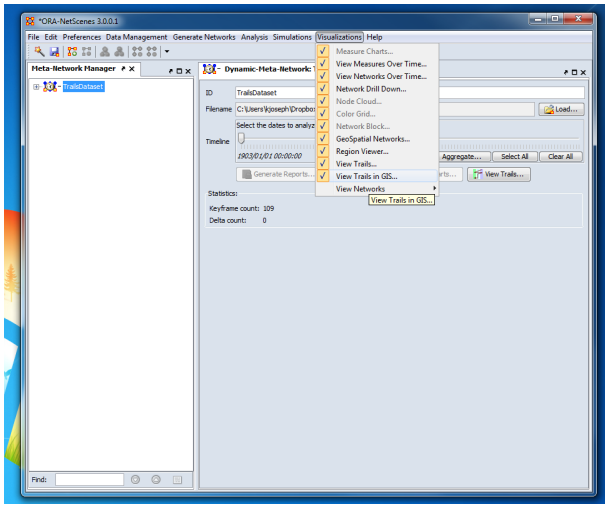
Networks Over Time Visualizer



<Your Name>

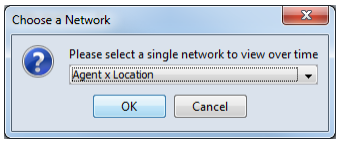
Carnegie Mellon
institute for SOFTWARE RESEARCH

Geospatial Visualizer



Carnegie Mellon
institute for SOFTWARE RESEARCH

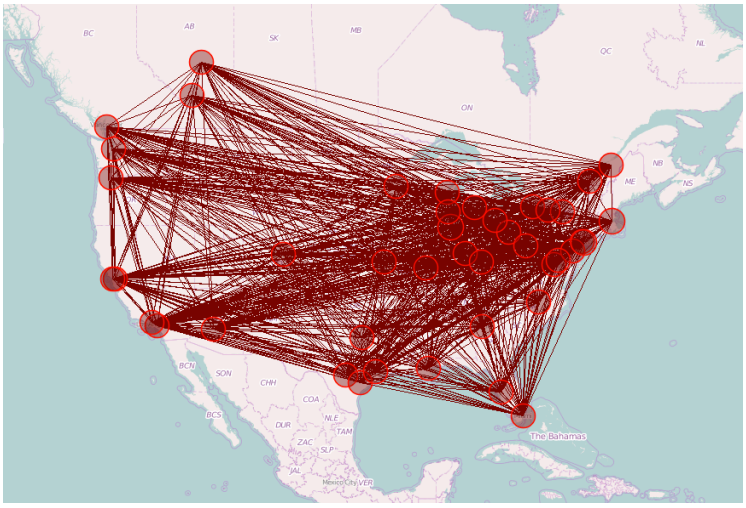
Geospatial Visualizer



<Your Name>

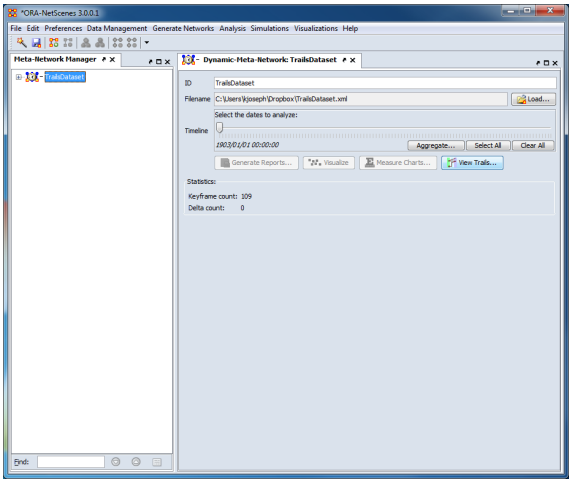
Carnegie Mellon
institute for SOFTWARE RESEARCH

Geospatial Visualizer



Carnegie Mellon
institute for SOFTWARE RESEARCH

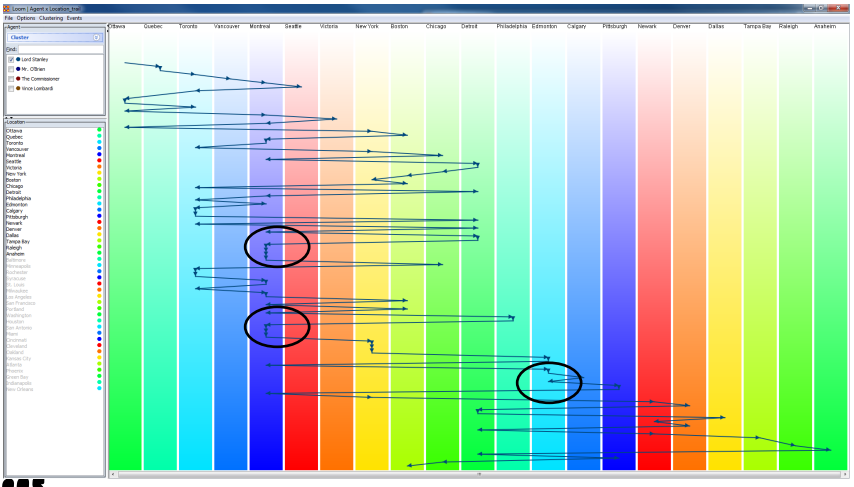
Loom



<Your Name>

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

Loom



Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

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

<Your Name>

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

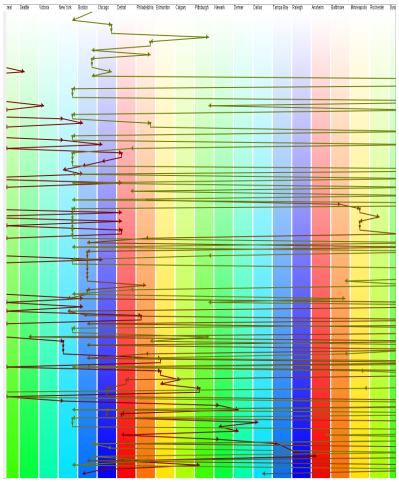
What we'll do

- 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

Carnegie Mellon
ISR 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.

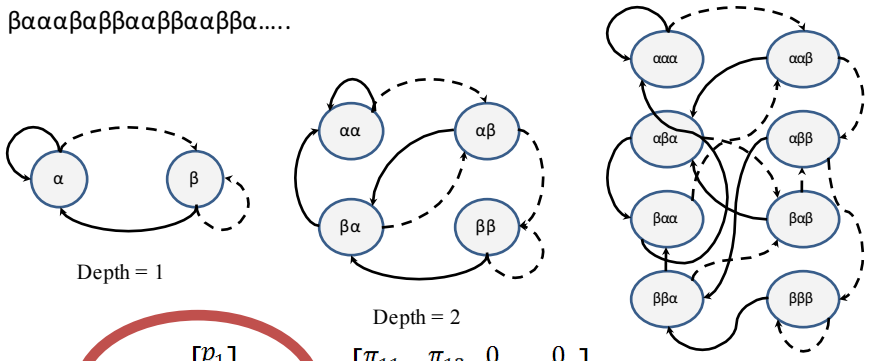


<Your Name>

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

Feature vector representation using PFSA

βααaβββααββαββα…..



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

State Probability Vector *State Transition matrix*

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

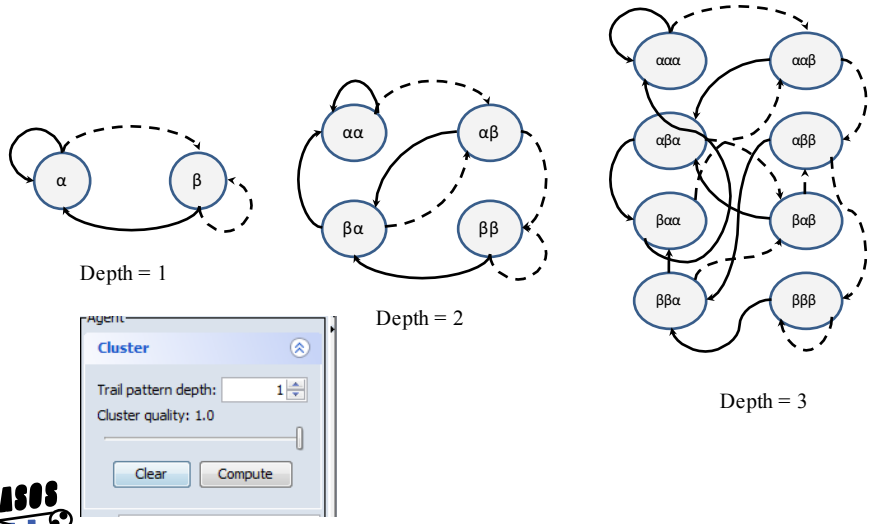
Clustering of Trails using PFSA

- 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.

<Your Name>

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

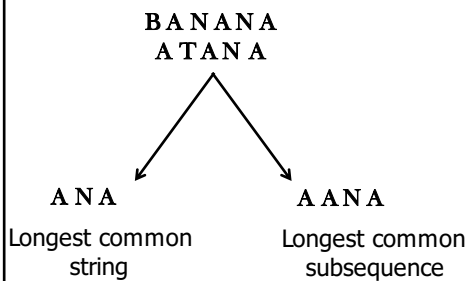
Refining the Coarse Clustering

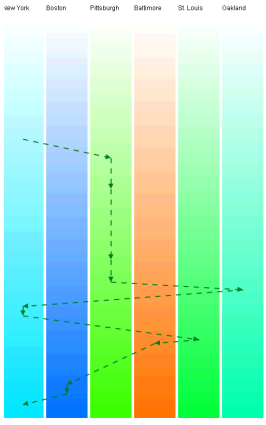


Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

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.





<Your Name>

Carnegie Mellon
ISR 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**

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

Generating Networks from Trails

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

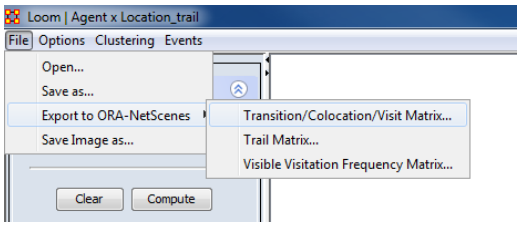
What we'll do

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

<Your Name>

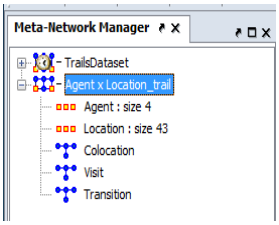
Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

Exporting the Matricies



Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

What we now have



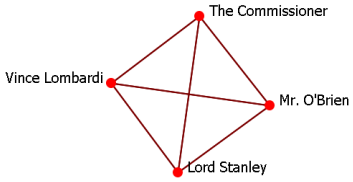
- ORA uses the entire trailset and outputs a single meta-network
  - Colocation – An edge is created between the trophies if they ever existed at the same place at the same time
  - Visit Matrix – An edge is created between city and trophy if the city ever won that trophy
  - Transition – An edge is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

<Your Name>

Carnegie Mellon
institute for SOFTWARE RESEARCH

Colocation

Agent x Location_trail

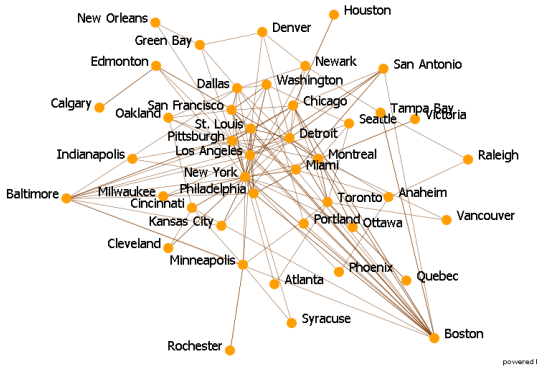


powered by ORA-NetScenes, CASOS Center @ CMU

Carnegie Mellon
institute for SOFTWARE RESEARCH

Transition

Agent x Location_trail



powered by ORA-NetScenes, CASOS Center @ CMU

<Your Name>

Carnegie Mellon
institute for SOFTWARE RESEARCH

Visit

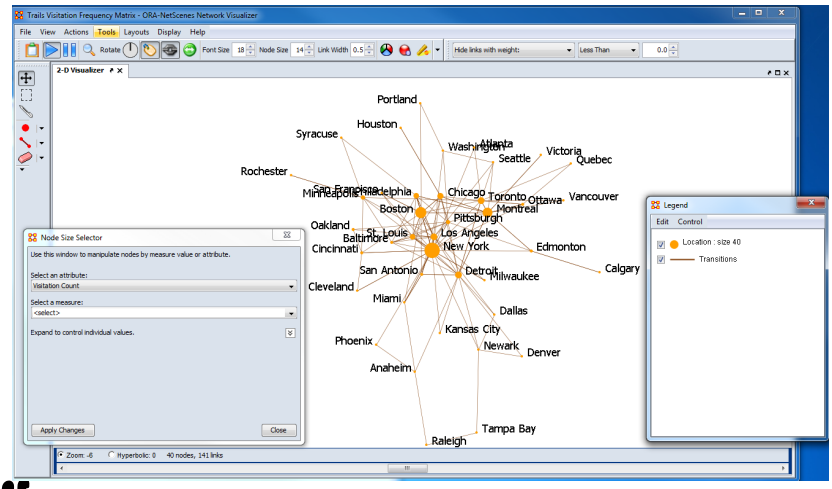
Carnegie Mellon
institute for SOFTWARE RESEARCH

Visitation Frequency Matrix Location

- Operates off of what's visible!
- Node attribute – total visits
- Matrix – similar to the Transition matrix

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

Visitation Frequency Matrix Location



Carnegie Mellon
ISR 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

<Your Name>

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

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