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

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
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

Static Data
Panel Data

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

What problems do we run into here?

![Diagram showing interactions between Bill, Ted, Phil, and times 1 and 2]
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

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?
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?

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
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
  - …

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
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?

Looking at trails and generating networks from them

**LOOM**
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

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
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
The Data

- Our trail:
  - Locations are our nodes
  - Agents are what is moving between them

- Let's explore the data
  - In ORA proper
  - Networks over time visualizer
  - Geospatial Visualizer

ORA Proper
Networks Over Time Visualizer
Geospatial Visualizer

Choose a Network

Please select a single network to view over time

Agent's Location

OK Cancel
Geospatial Visualizer

Loom
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
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

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.
Feature vector representation using PFSA

\[ \text{State Probability Vector} \quad \text{State Transition matrix} \]

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

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.
Refining the Coarse Clustering

Depth = 1

Depth = 2

Depth = 3

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.
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

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
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
Colocation

Transition
Visitation Frequency Matrix Location

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