Mentor
James Abello
jabello@dimacs.rutgers.edu

Neel Parikh
nparikh@gmail.com

Ahmed Al-Asadi
ahmedal010@gmail.com
CHALLENGES

Identify salient relationships and trends
- Handling massive, streaming data sets
- Making predictions

Graphs provide a logical representation
- Relationships between discrete data entities
- Mechanism for analysis
- How to capture dynamic aspect?

Visualizing the graph
- Laying out vertices to display clearly
- Preserving mental map
- Deciphering semantic interpretation
Preserving the User’s View

- Vertices and edges are changing
- Maintaining clear layout
- Where to position graph elements?
- Movement must not be too drastic

Example of Problematic Graph

- Open Office software development history
- 3 months of commits to trunk of SVN repository
- Visualization on next slide
CHALLENGE: Preserving Mental Map

Produced with Gource 0.38
http://youtube.com/watch?v=a-gAoYapM8U
EXAMPLE: Shipping Manifest Data

Containerized Cargo Shipments • Foreign Ports to U.S Ports

Thirty days of data: Jan 30, 2009 - February 28, 2009
One bipartite graph for each day of data

- Vertices are port pairs and content categories
- Edges weighted by quantity of a good shipped between port pair

Created 30 new graphs using *discrepancy weight*

- Statistical measure of how “out of the ordinary”
- Cumulative - reflects information from time 1 to $t$
- Chooses most salient edges up to that time

Computed the maximum spanning forest

- Simplifies and unclutters visualization
- Preserves the most important information
- Edges selected as they are streamed to forest
Two types of vertices
- Green nodes are port pairs
- Blue nodes are specific content categories

Measures used
- Edge Firing Rate (frequency/time) heat map
  - Low
  - High
- Edge Discrepancy Weight
  - Edge Thickness
- More “Notorious” Vertices
  - Those surrounded by hotter and thicker edges
Contents Vertex
• Hardware-Plumbing-TrapStrainers

Port Pair Vertices
• 62: YantianChina  →  NEWYORK-NY
• 37: ShanghaiChina  →  PORTSMOUTH-VA
• 14: PusanKorRep  →  NEWYORK-NY
• 228: LudaChina  →  SANPEDRO-CA
• 1656: TientsinChina  →  SANPEDRO-CA
• 1000: YingkouChina  →  TACOMA-WA
• 1723: YingkouChina  →  NEWARK-NJ
• 2501: XiamenChina  →  CHARLESTON-SC
Visualizing Social Networks

- Goal is general purpose visualization technique

- Use social network data to extend existing analysis method

- Twitter is a rich source of streaming data
Tweet Meta-Data

Not Just 140 Characters!

- Creation Date
- User Name
- Geo-Tag
- Location Type
- Number of Followers
- Language Preference
- App. That Sent Tweet
Natural Disasters

• Test effectiveness of discrepancy detection
• Hurricane Irene data set
• Over 3,000,000 Tweets

Example: 2011 Japan Earthquake

• 500% increase in Tweets from Japan
• @replies hour before and after earthquake
• Replies into Japan are pink, out of Japan are yellow
Tweeting During Japan Earthquake

Written in Processing.js
http://www.youtube.com/watch?v=716mJnFnY7s
Epidemiology

• Predict an outbreak of disease
• Determine “hot zones”
• Prevent spread

Example: Global Movements Trends

• Tweets with phrases like “just landed in” and “arrived”
• Destination compared with user’s home location
• System plots voyages over time
• Could be used to track spread of flu virus
• Shows what is possible for graph visualization
Global Movement Tweeting

Written in Processing (prog. language)
http://www.youtube.com/watch?v=rUuPBfEkiJs

