Advances in Semantically Augmented Flow Data for Dynamic Impact Assessment, Response Selection, and Alert Prioritization

Nik Kinkel*, Harris T Lin, Chris Strasburg

The Ames Laboratory

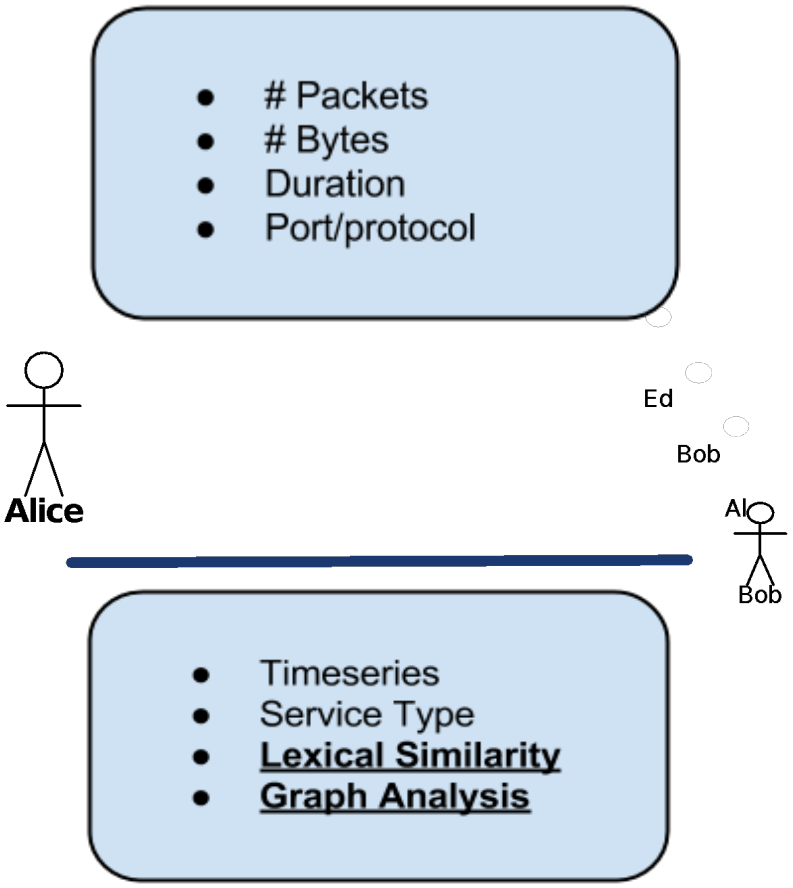
{nskinkel,htlin,cstras}@ameslab.gov

* - Presenting

Outline

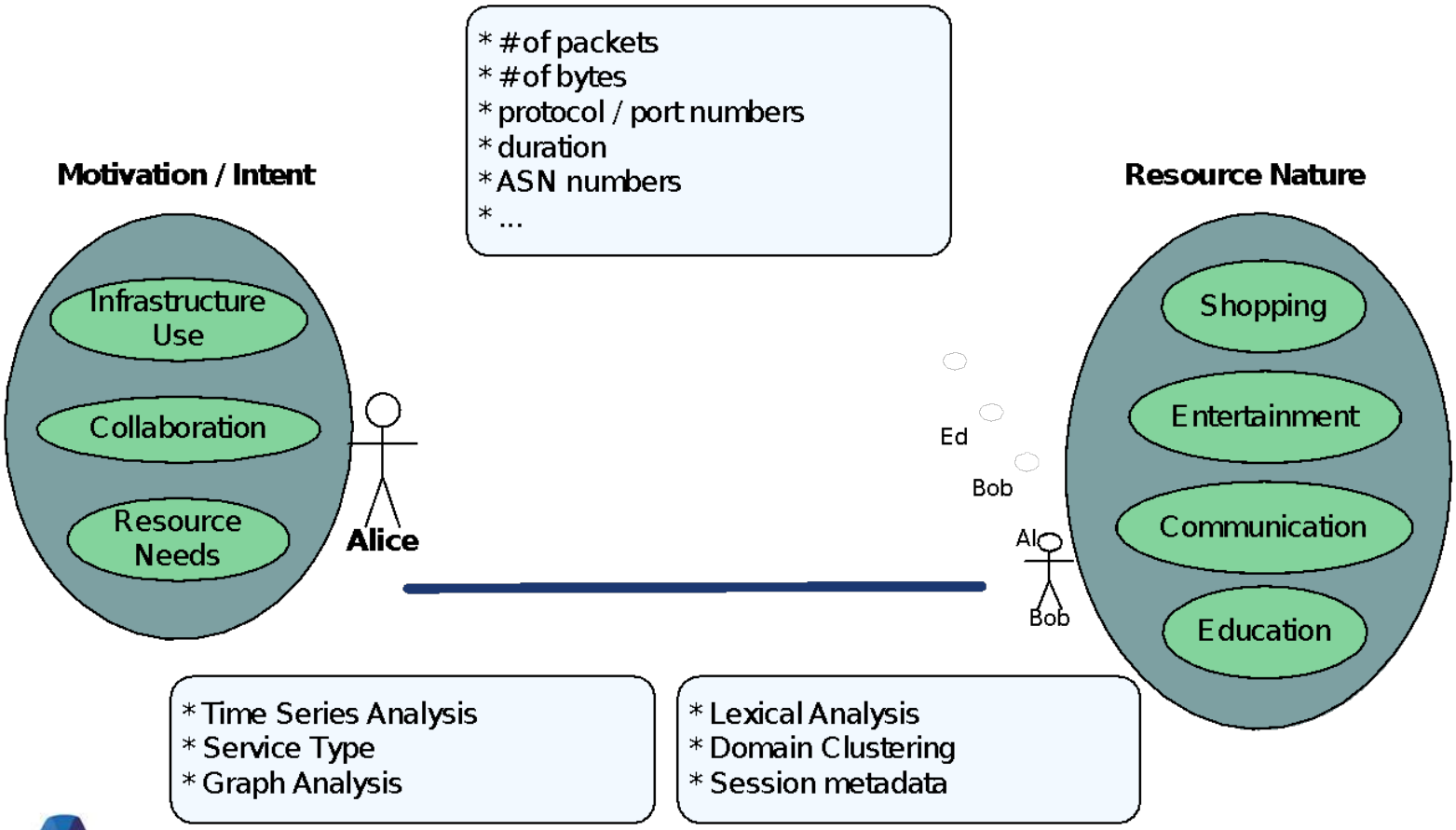
1. Semantic Augmentation Overview
2. Lexical Augmentation
3. Augmentation through Graph Analysis
4. Visualization, Data Exploration, User Interfaces

What is Semantic Flow Augmentation



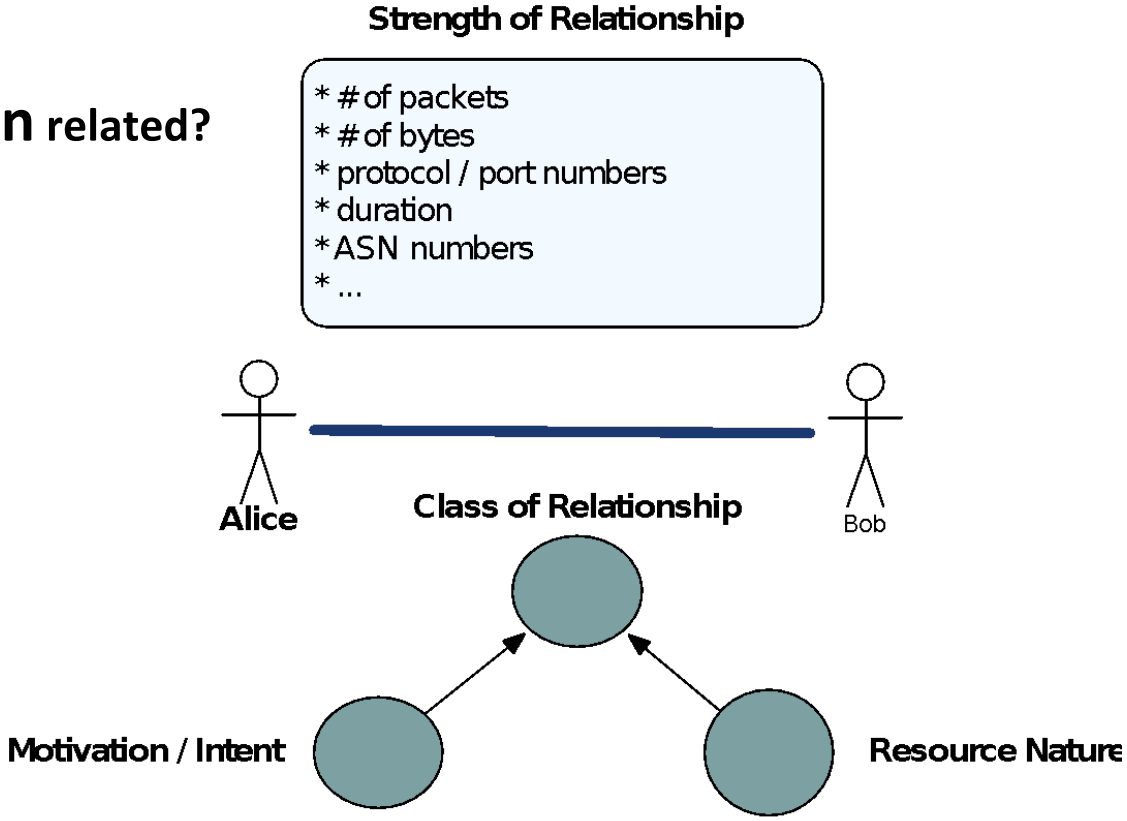
What is Semantic Flow Augmentation

- Semantic – *Of or relating to meaning…*

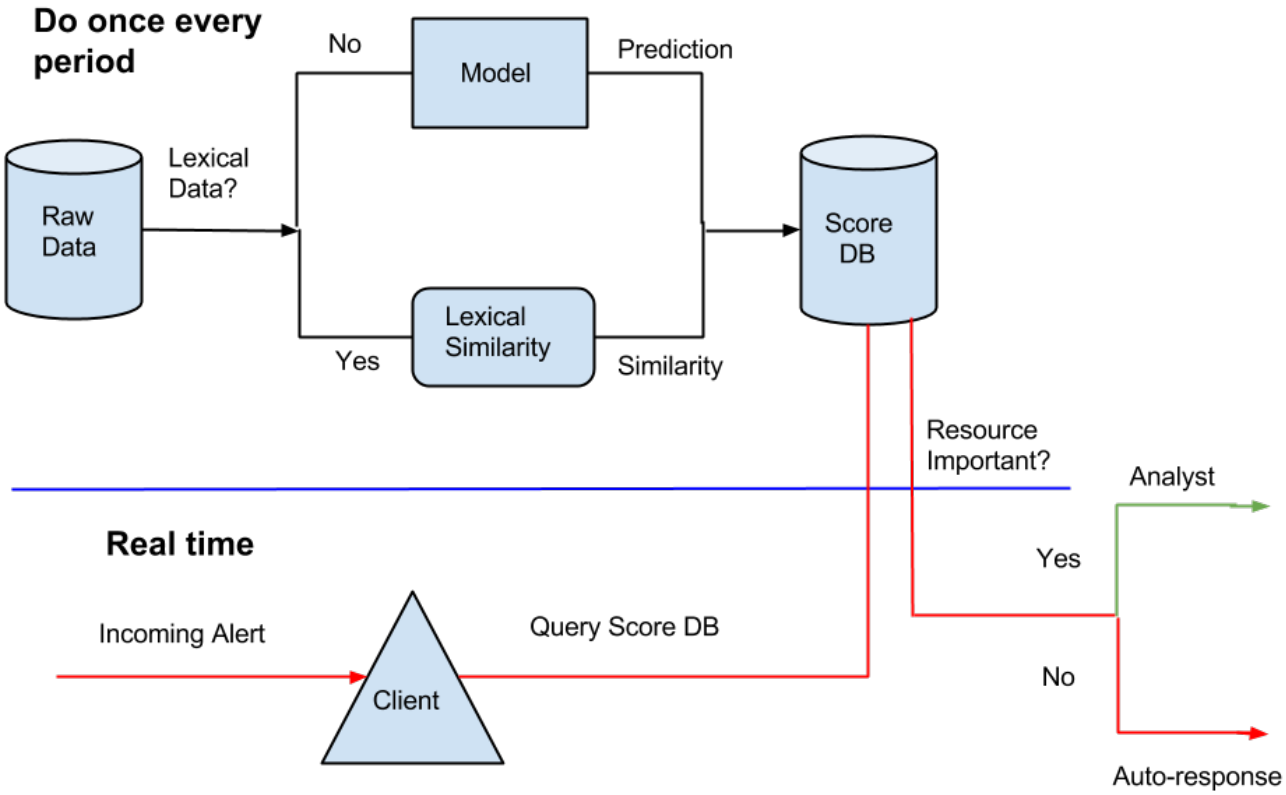


Why Semantic Augmentation

Is it mission related?



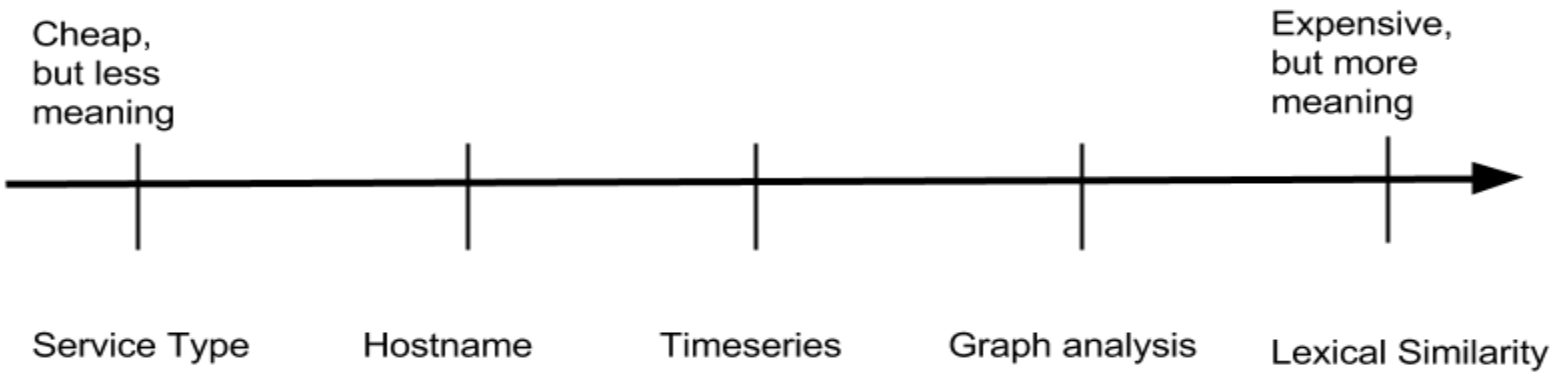
Architecture Overview



Outline

1. Semantic Augmentation Overview
2. Lexical Augmentation
3. Augmentation through Graph Analysis
4. Visualization, Data Exploration, User Interfaces

Semantic Data Sources



Challenges in Lexical Data Collection

- Can be very slow
- Tough to scale
- Noisy, machines may get blocked/blacklisted for scraping
- Complicated bootstrap/software install process

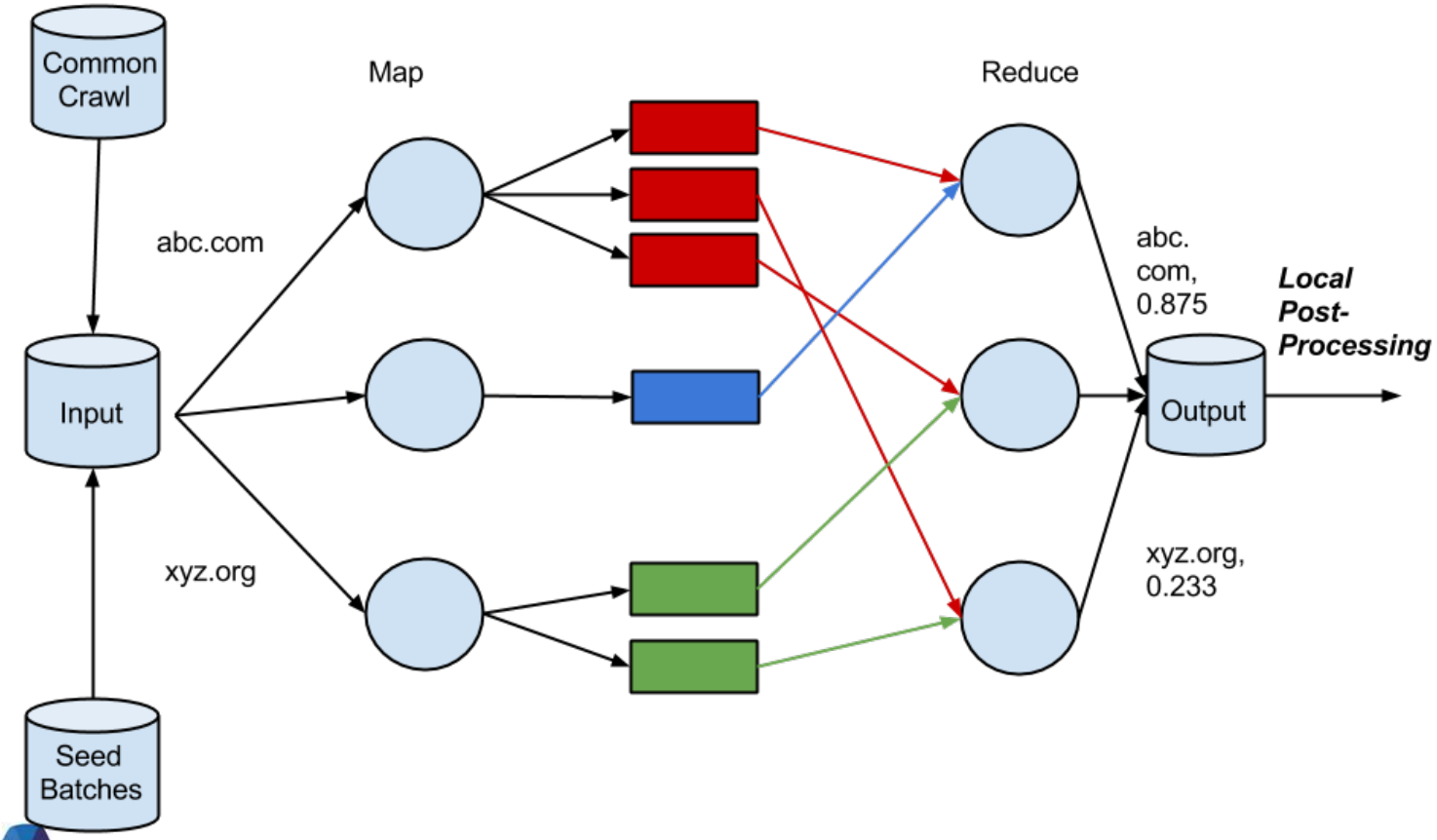
Non-solutions

- Extract data on the wire -- too slow, political concerns
- local 24/7 scraper(s) -- orgs don't want to run scrapers, machines/netblocks get blocked, angry emails
- outsource: scrape remotely -- leak traffic stats to 3rd party

Scalable, Distributed Solution

- Need a large, public data set
- Need to parallelize
- A real solution!

Amazon Elastic MapReduce



The Data

- Characteristics
- Cost
- Coverage

Outline

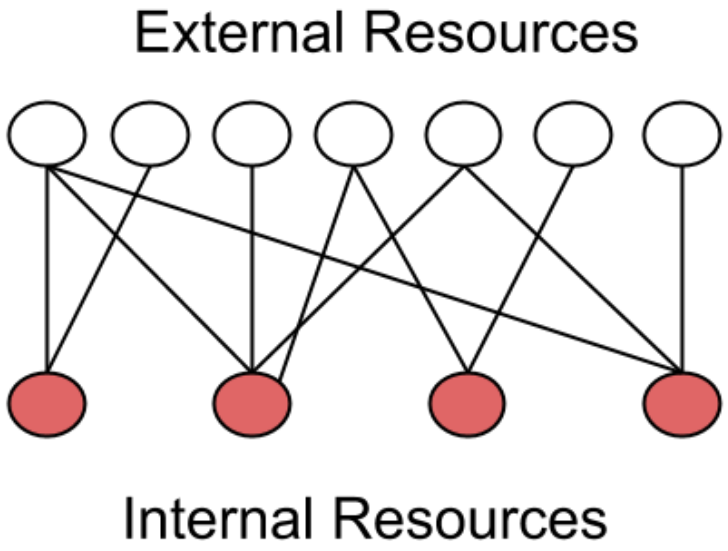
1. Semantic Augmentation Overview
2. Lexical Augmentation
3. Augmentation through Graph Analysis
4. Visualization, Data Exploration, User Interfaces

Graph Analysis: Another Semantic Data Source

- Can we use graph analysis techniques to predict resource importance?
- Can we visualize relations between remote resources based on their interactions with local resources?
- Can we figure out distinct “classes” of resources?

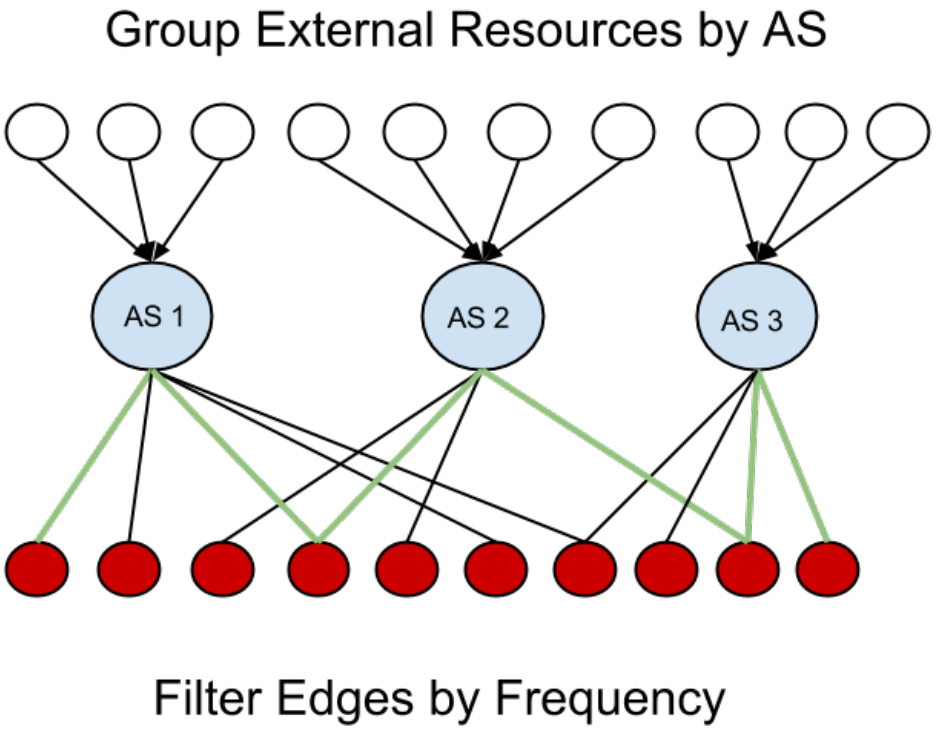
Graph of External and Internal Resources is Bipartite

- Red nodes internal resources
- Black nodes external resources
- Draw an edge if they communicate
- Clustering on this raw data is expensive!



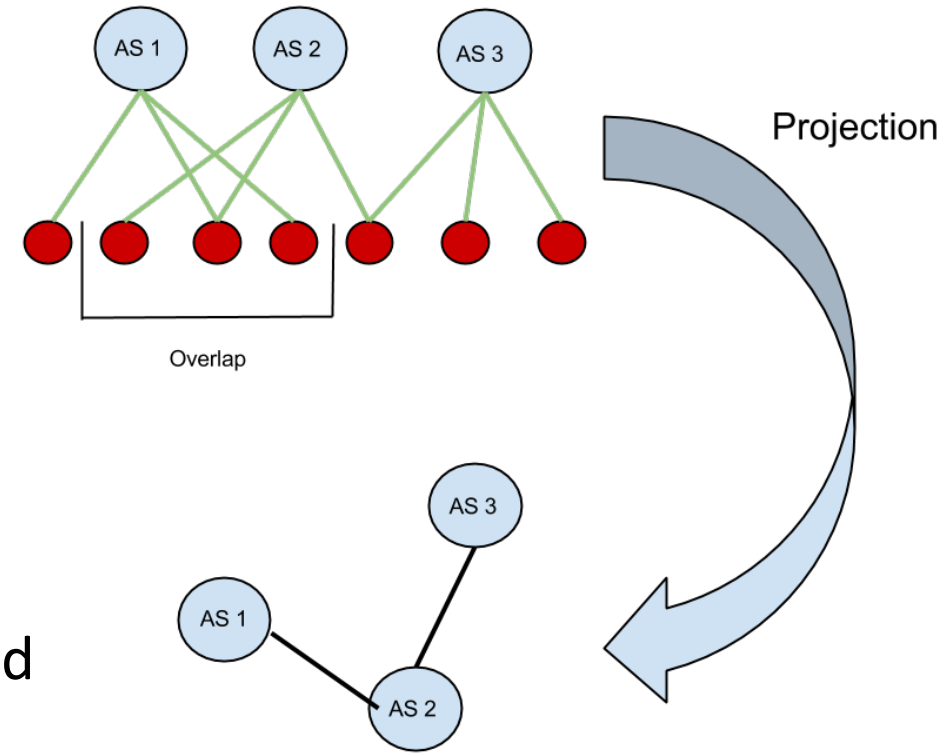
Optimizations

- Preprocess to reduce both nodes and edges
- Nodes: group by Autonomous System
- Edges: only draw an edge if sustained interaction over some time interval



Bipartite Projection

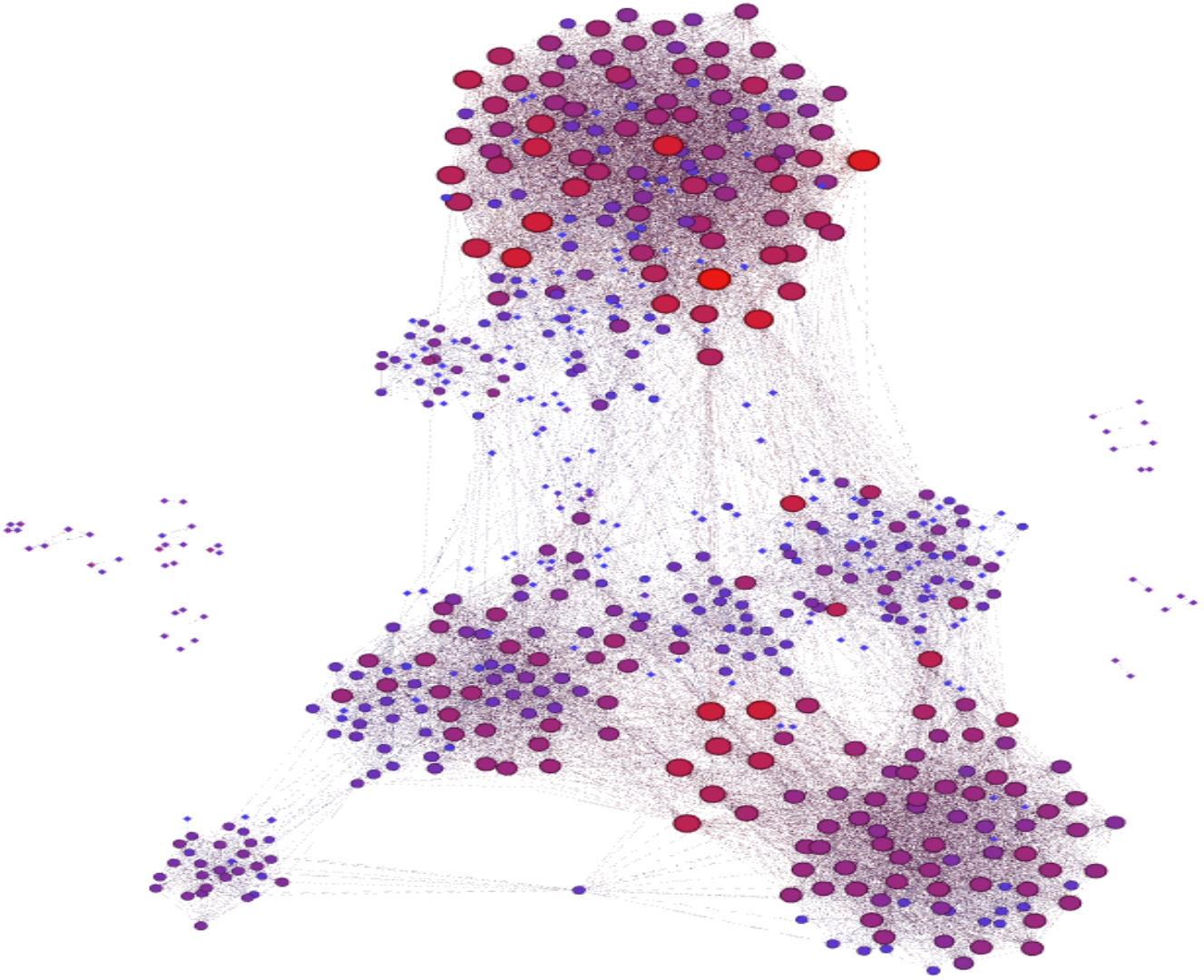
- Remove edge weights
- Use conditioned threshold approach to draw edges between ASs
- 3 threshold approaches:
  - Unconditional
  - Agent-degree conditioned
  - Dual-degree conditioned



Computing Edges Between Autonomous Systems

$$Pr(P_{ij} = x) = \frac{\binom{A}{x}\binom{A-x}{D_i-x}\binom{A-D_i}{D_j-x}}{\binom{A}{D_i}\binom{A}{D_j}} = \frac{\binom{D_i}{x}\binom{A-D_i}{D_j-x}}{\binom{A}{D_j}}$$

Zachary Neal. "The backbone of bipartite projections: Inferring relationships from co-authorship, co-sponsorship, co-attendance and other co-behaviors." Social Networks 39, 2014, 84–97



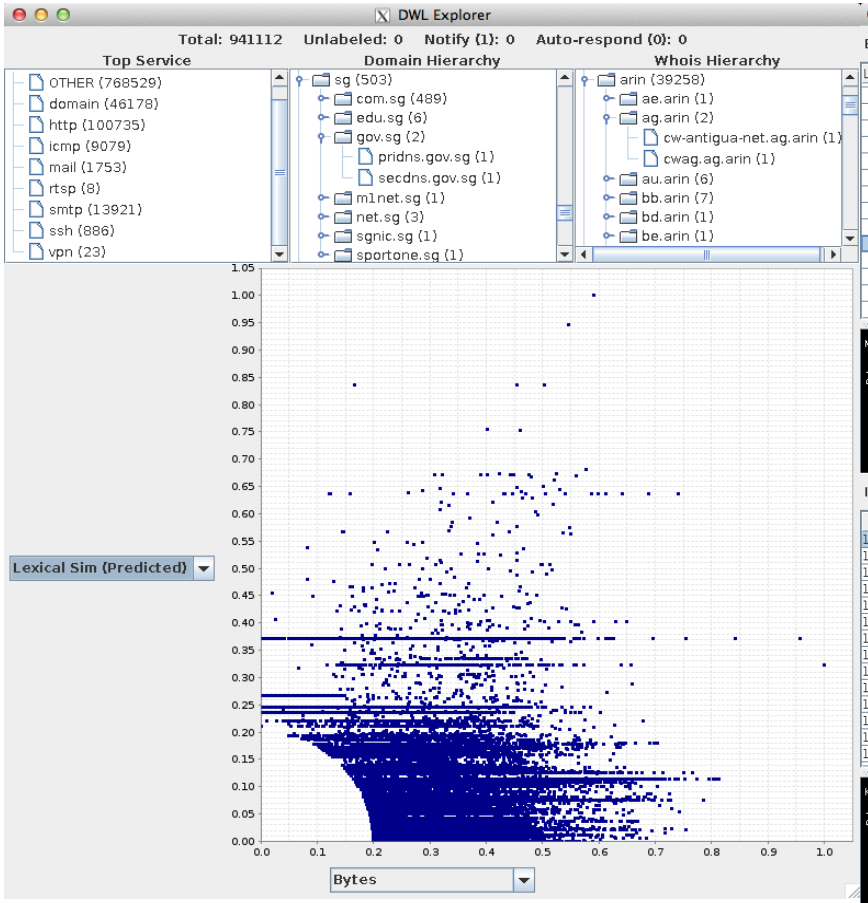
Outline

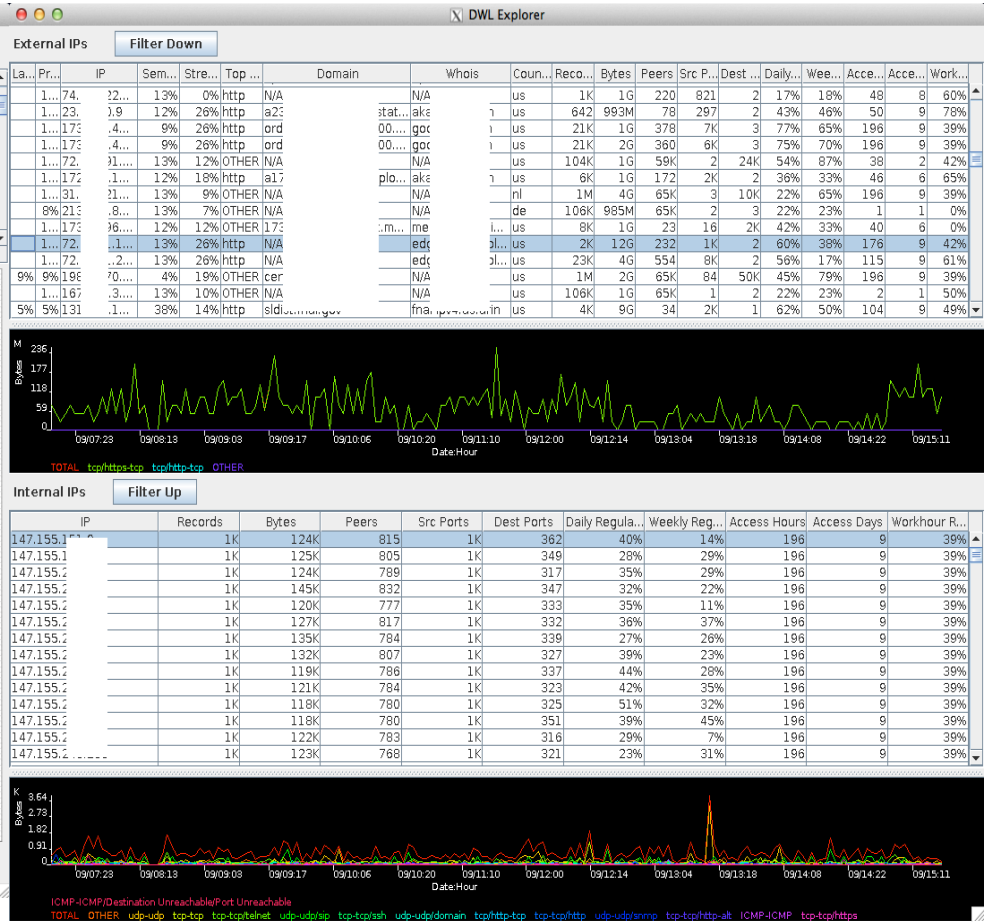
1. Semantic Augmentation Overview
2. Lexical Augmentation
3. Augmentation through Graph Analysis
4. Visualization, Data Exploration, User Interfaces

Encouraging Analyst Exploration

- Help analysts explore relationships between resources
- Discover unexpected relationships/anomalies
- Help us figure out new features
- Better understand our own network

Exploration





Software Release Information

- Software is released under BSD License
- https://github.com/Ames-Laboratory-Cyber-Group/Cydime
- Cydime is provided in the hope it will be useful, but it is primarily a research project. It may not be production ready or user friendly. Please get in touch if you'd like to try it out. We'd love to talk to you.
- cstras@ameslab.gov
- htlin@ameslab.gov
- nskinkel@ameslab.gov

Future Work

- Verify and validate lexical information
- Better results for Common Crawl resource coverage (we need more sites with more variety)
- How much information is gained/lost from lexical to bipartite graph?
- Applications beyond network security?

Special Thanks To

CloudHelix

For sponsoring my student conference scholarship.