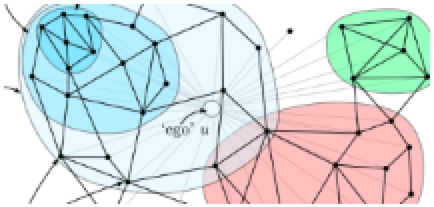


http://www.kaggle.com/c/learning-social-circles

Kaggle Learning Social Circles in Networks

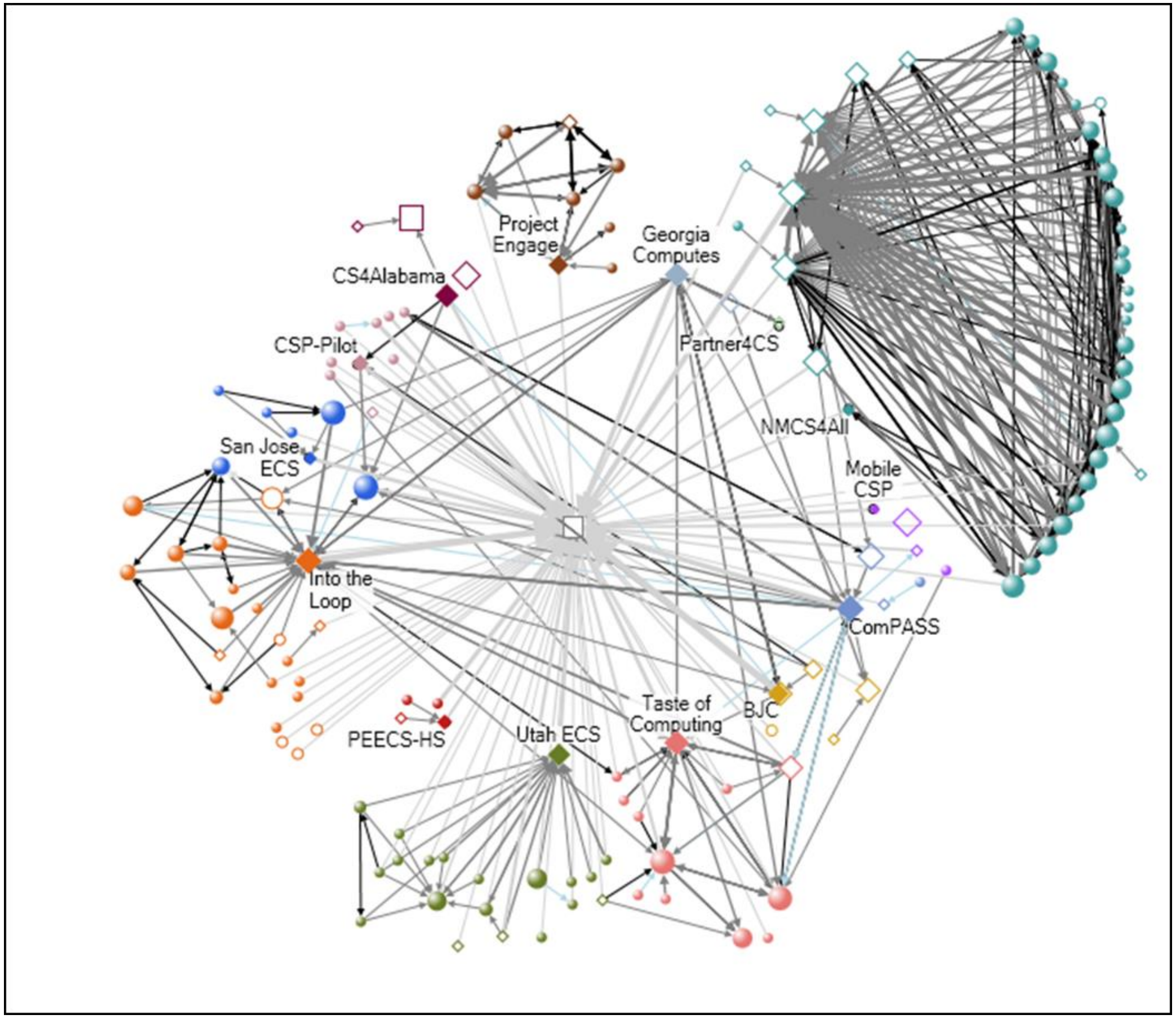
Kathleen Perez-Lopez
GADSDC2
Final Project Status Report
10/6/2014

Background

- Behavioral norms are conveyed through SOCIAL NETWORKS
  - Health – eating, exercising, sleeping, drinking
  - Consumer – what we think we need/must buy
  - Political – attitudes, activities (GOTV, protesting, bombing)
  - Intellectual
    - Whether we sign up for - & stick with – GADSDC
- Kaggle Learning Social Circles
  - Help us organize our FB friends into circle
  - Probably VERY interesting to other entities
- My interest: social network analysis of online communities of practice for educators, .e.g., NSF CS10K

NSF CS10K Community of Computer Science Educators

Posts during 2013



Competition Description

- Goal: Infer the FB users’ circles of friends
- Data
  - Egonets for 110 FB users (include alters, not ego; coded by ego)
  - For **60 training egos**, a file of circles of alters
  - For all 110 , file of anonymized features
- Measuring submission, derived circles for **50 test egos**
  - Minimum edit distance[submission, ground truth]
- Provided
  - Python code to implement edit distance
  - Sample submissions
    - one big circle of all friends: score = 5553
    - connected components submission: score = **5199**

Network as Adjacency Matrix

April 10, 2012 / Mike Bostock

Les Misérables Co-occurrence

Order: by Name

This matrix diagram visualizes character co-occurrences in Victor Hugo's *Les Misérables*.

Each colored cell represents two characters that appeared in the same chapter; darker cells indicate characters that co-occurred more frequently.

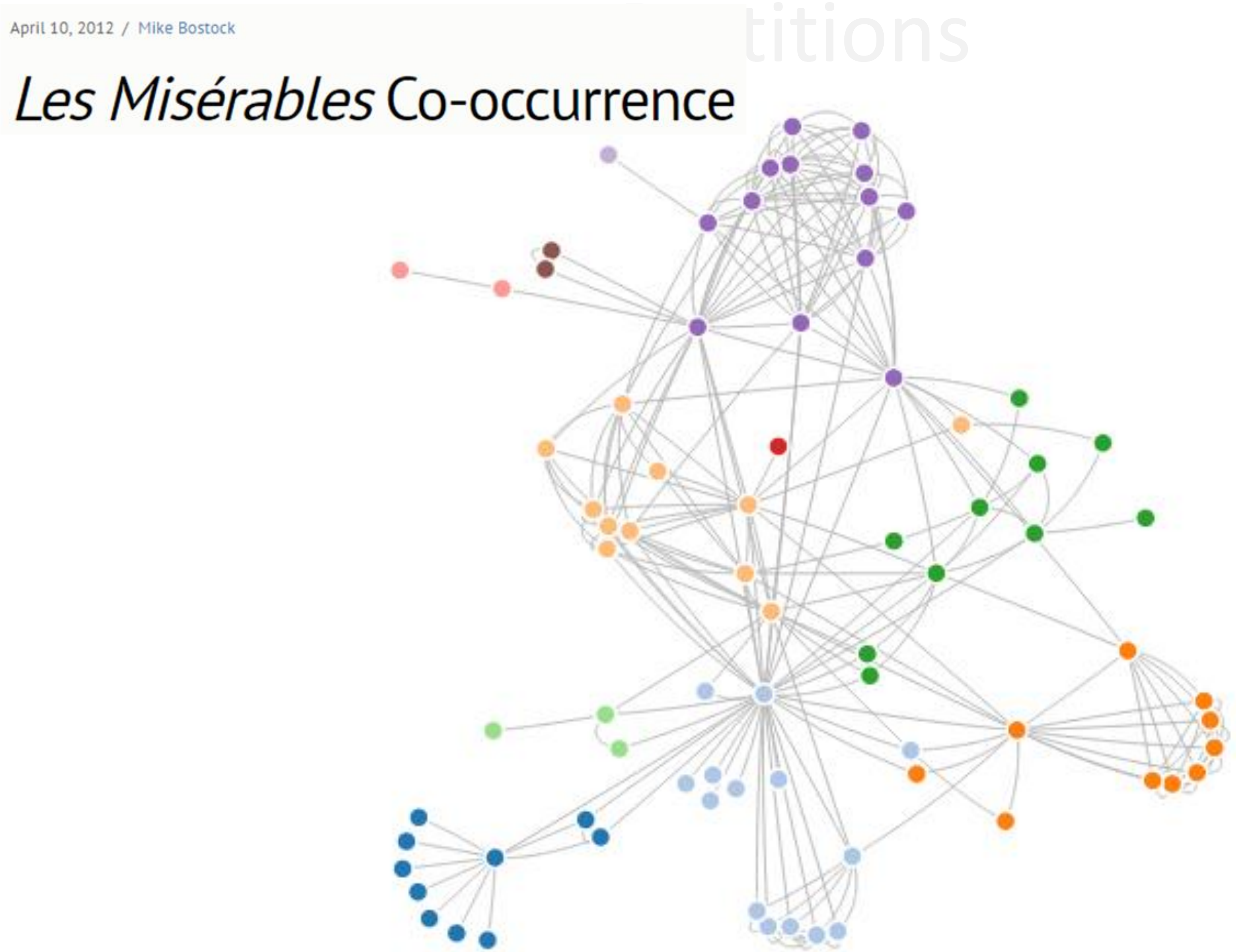
Use the drop-down menu to reorder the matrix and explore the data.

Built with d3.js.

Source: The Stanford GraphBase.

mike/miserables/

Network Clusters as Adjacency Matrix titions

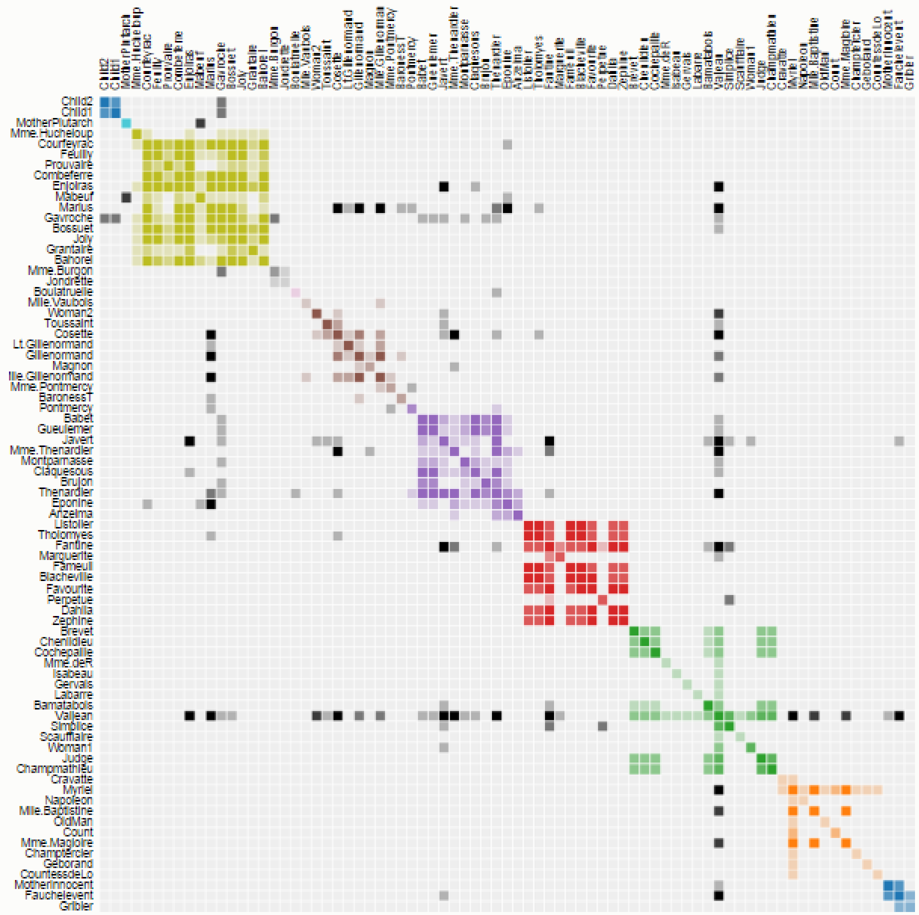


http://bl.ocks.org/mbostock/4600693

Network Clusters as Adjacency Matrix Partitions

April 10, 2012 / Mike Bostock

Les Misérables Co-occurrence



Order: by Cluster

This matrix diagram visualizes character co-occurrences in Victor Hugo's *Les Misérables*.

Each colored cell represents two characters that appeared in the same chapter; darker cells indicate characters that co-occurred more frequently.

Use the drop-down menu to reorder the matrix and explore the data.

Built with d3.js.

Source: The Stanford GraphBase.

http://bost.ocks.org/mike/miserables/

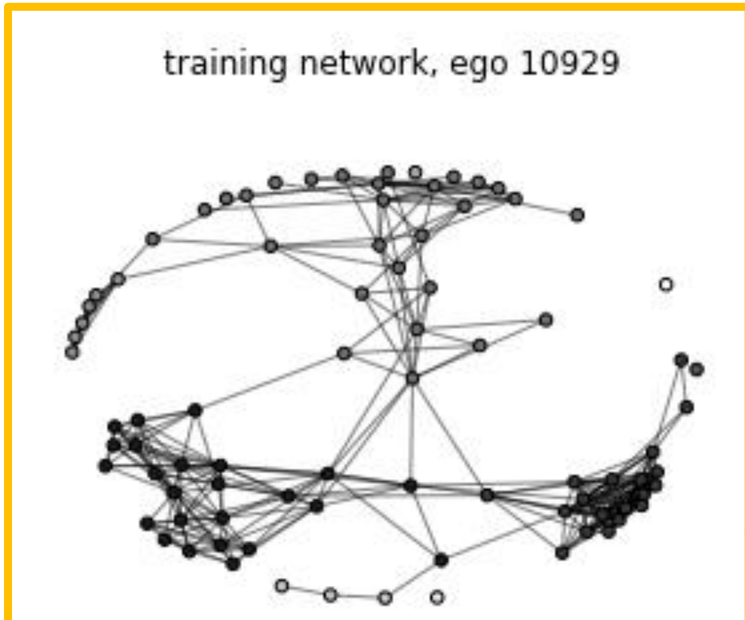
Strategy: Use Egonet Structure

1. Examine and characterize egonets
   a. Compute & visualize graph-level metrics
      Number of nodes, edges, components; density; ave nodes/component; transitivity; cluster_coefs; geodesic; ave shortest path
   b. Visualize egonets
2. Tailor clustering algorithm to egonet
   a. Apply many graph clustering algorithms
   b. Classify each egonet as good/poor wrt each alg
   c. Predict algorithm outcome based on egonet metrics
3. Test: select best clustering algorithm based on metrics
4. Improve using user users' features
5. Submit by OCTOBER 28

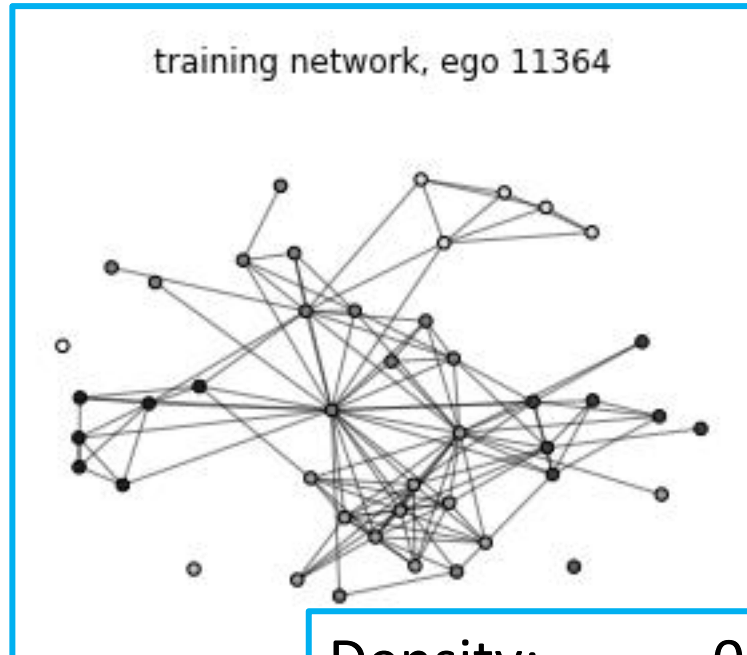
Status: ~40%

1. Examine and characterize egonets
   a. Compute & visualize graph-level metrics
      Number of nodes, edges, components; density; ave nodes/component; transitivity; cluster_coefs; geodesic; ave shortest path
   b. Visualize egonets
2. Tailor clustering algorithm to egonet
   a. Apply some graph clustering algorithms, Louvain Method
   b. Classify each egonet as good/poor wrt each alg
      Edit distance program doesn’t produce benchmark values
   c. Predict algorithm outcome based on egonet metrics
3. Test: select best clustering algorithm based on metrics
4. Improve using user users’ features
5. Submit by OCTOBER 28

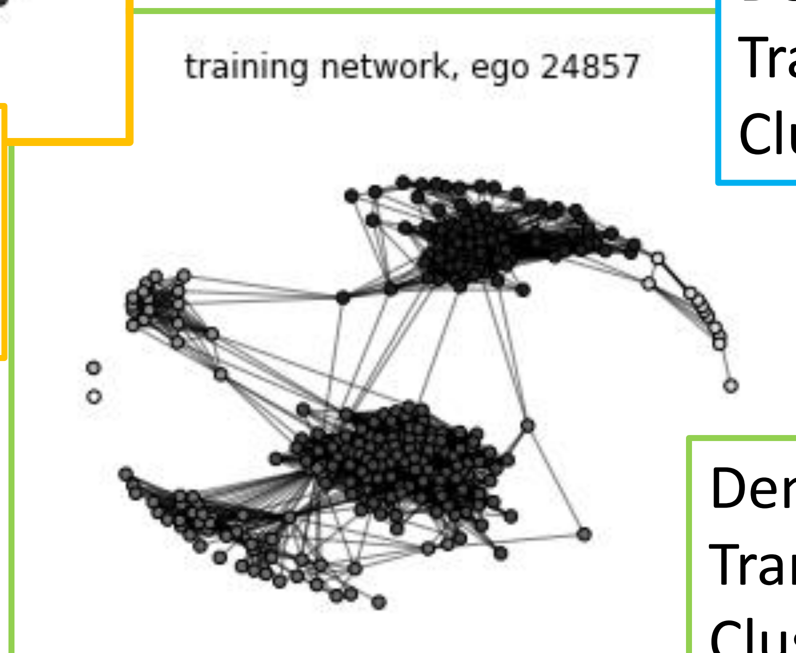
Louvain Partitions



Density: 0.09
Transitivity: 0.70
Clustering: 0.58



Density: 0.13
Transitivity: 0.54
Clustering: 0.63



Density: 0.09
Transitivity: 0.55
Clustering: 0.63

extra

- NetworkX
- Python
- Munkres for edit distance (skeleton provided)

Training Ego	num nodes	num edges	density	num components	ave nodes / component	transitivity	cluster coefs	geodesic	ave shortest path
24857	301	3992	0.088	3	100.333	0.550	0.627	7	3.118
11364	45	128	0.129	4	11.250	0.424	0.537	5	2.263
10929	84	329	0.094	5	16.800	0.698	0.578	9	3.738