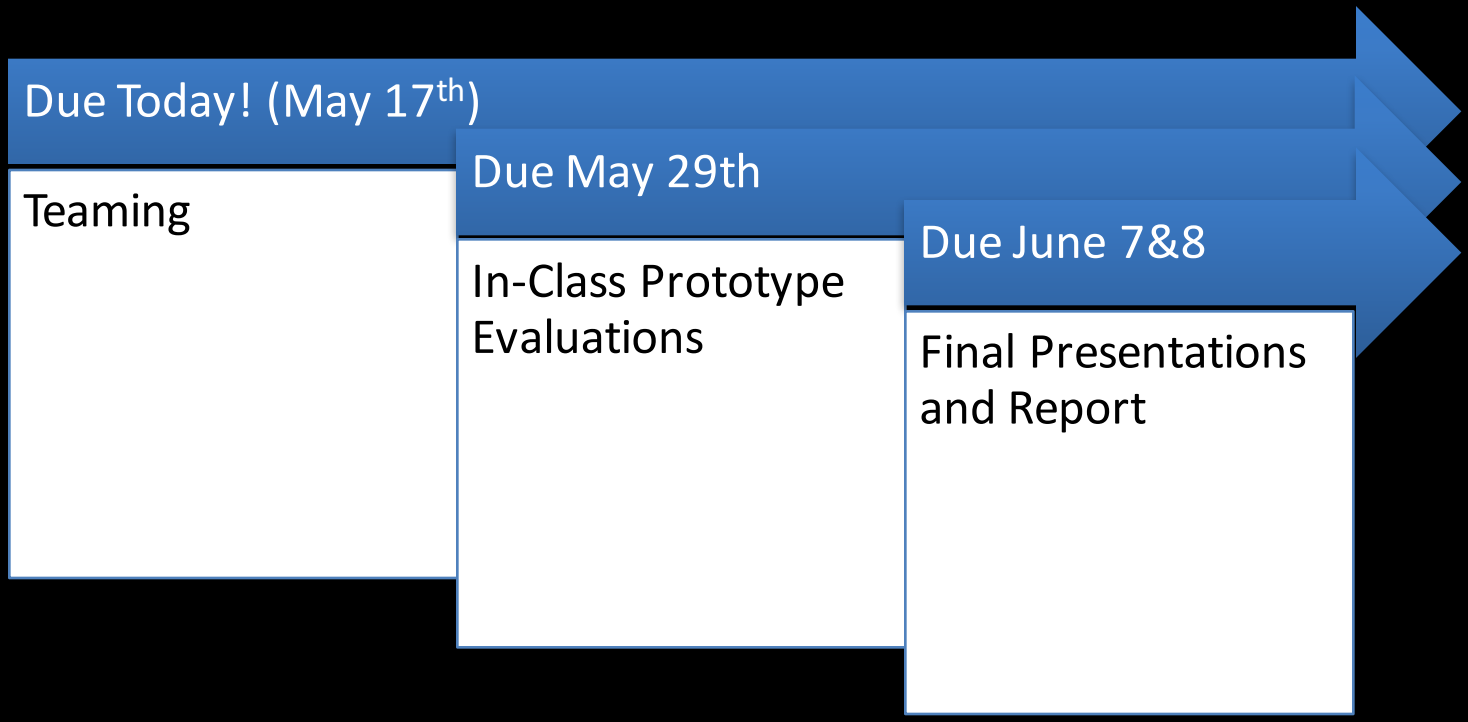
CS185 Projects

John O'Donovan
Computer Science Dept.

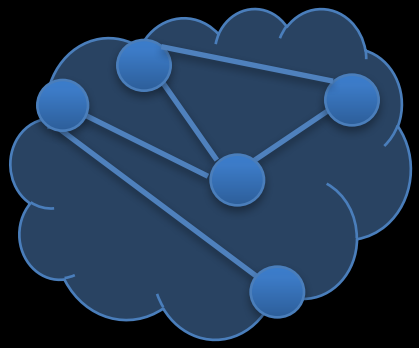
Office: Bldg. 935 room 101
Hours: T/Th 2-3pm, or by email appt.
Email: jod@cs.ucsb.edu

Project Timeline



Abstract Project Outline

Research Questions and **Initial Hypothesis**



Network data
"in the wild"

Model:
Organized representation of data

Algorithm:
Processes that run on the data

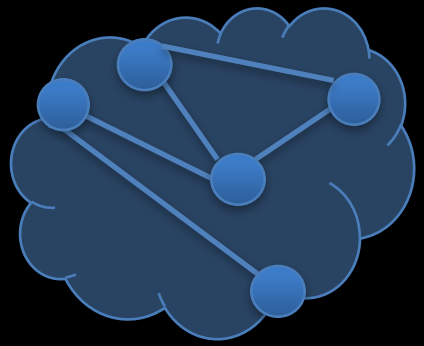
Interface:
Representation of parts of the model & algorithms

Interaction:
Discovery and insight through interaction with the UI

Refined Hypothesis
based on iterated data analysis

->Proven or not.

Examples



Network

- Twitter
- Facebook
- Reddit
- New York Times API
- Wikipedia Articles
- Protein Interactions
- Epidemics
- Fads
- Classmates
- …

Model

- XML / JSON
- SQL
- Memory-based
- Document Collections
- MongoDB
- Similarity Matrices
- Correlations
- …

Algorithm

- Search
- Recommendation
- Pruning
- Filtering
- Personalization
- Content-matching
- Neighborhood formation
- …

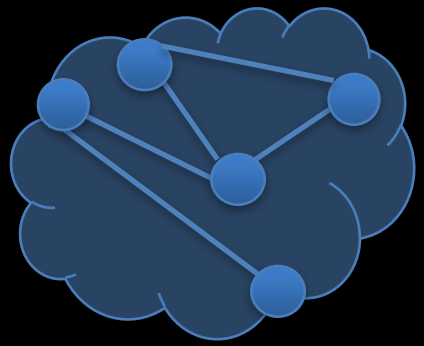
Interface

- Command line
- GUI
- Tabular views
- List views
- Node-link views
- Visual design
- …

Interaction

- Search queries
- Graph interactions
- Clicks
- Drags
- …

Things to consider



Network

- Scale
- Rate Limitations
- Data Availability
- Network Dynamics
- …

Model

- Will model scale to data
- Flexibility
- Portability
- Stability
- Speed
- Queries?
- …

Algorithm

- Suitability
- Speed
- Scale
- Implementations
- …

Interface

- End users
- Functionality
- Complexity
- …

Interaction

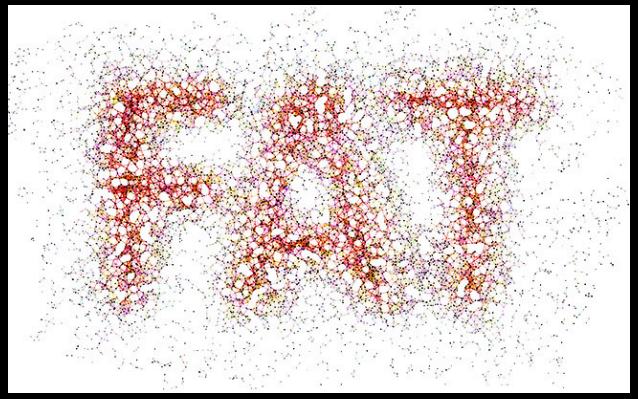
- Cognitive Issues
- HCI issues
- …

Example Project: Exploring Effects of Different Edge Formations in Christakis Obesity Study.

Research Questions:

1: In Christakis' study of obesity patterns, how can we best factor in reasons for friend connections to the prediction model?

2: How will a prediction model with refined edge connections (friendship reasons) compare to the prediction results in Christakis' 2007 paper (Nature) http://www.nejm.org/doi/full/10.1056/NEJMsa066082



Breakdown

Network

Framingham heart study data set from http://www.nejm.org/doi/full/10.1056/NEJMsa066082 Or, is there other data you could use? Twitter wasn't that big in 2007

Model

JSON representation of the network stored in a text file. Lists of <nodeID, nodeID, FType> representing "friend" networks and a classification of friend types.

Ftype = {"Street", "School", "Club", "Business"}

Algorithm

Random Forest prediction model on training data.

Graph layout algorithm (Fruchterman Reingold)

Interface

Python notebook UI

Python Django web page with interactive node-link graph

Gephi (http://gephi.github.io/)

Dynamic statistics plots in Plot.ly (http://plot.ly)

Interaction

Graph interactions

Statistical analysis of graph in Gephi.

Examples from previous years

Projects

- FitBit Friends
- Twitter Topic Modeling
- TweetVisualization
- OpenPayments
- TheRedditProject
- Spectral Domain of Social Networks (link on request)
- http://hwright-ucsb.github.io/CS290D-Final-Project/

Dimensions for thinking about proposals and projects (in addition to project proposal document)

1. What is the HCI component?
   -what is new/interesting about it.
2. Who are the people who will benefit from the project?
   -How / When / Why will they benefit?
3. What are the data requirements?
   -Any privacy / security issues?
4. Timeline for project?
5. What are key technical challenges in the project?
6. What tools and technologies are needed?
7. Milestones, Goals, Deliverables…
8. How is success measured?