# Bridging the Gap Between Physical Location and Online Social Networks

## Cranshaw et al.

**Michael Molignano**
mikem@wpi.edu
**CS 525w – 3/1/2011**

# Overview

- Examines location of 489 users
- Introduces location-based features for analyzing geographic areas
- Provide model for predicting friends
- Relation between entropy of visited locations and number of friends
- Discuss potential benefits offline mobility has for online networks

# Motivation (p1)

- **Heard distinction of *online* and *offline* social networks**

- **"online social networks are contributing to the isolation of people in the physical world"**
  - **Deresieicz**
- **"online social networks have a positive impact on social relations in the physical world"**
  - **Pew Internet and American Life**

# Motivation (p2)

- Location-enabled smartphones everywhere
  - Foursquare, Gowalla, etc.
- Location makes physical behaviors easier to analyze
- Challenge inferring social behavior from locations
  - Especially location tracking alone

# Their Contributions

- **Evaluate on two main tasks**
  - **Predicting whether two co-located users are friends on Facebook**
  - **Predicting number of friends a user has**
- **Contributions:**
  - **1. Establish model of friendship by co-location**
  - **2. Find relationship between mobility pattern and number of friends**
  - **3. Show diversity of location can be used to analyze the context of social interactions**

# Related Work

- Mobility patterns to find statistical models
- Examined features of mobility
  - Proximity at work, Saturday night, etc.
  - Tracked phone conversations
  - Number of unique locations
  - Self report of important factors

- Most work relied solely on co-location without digging further

# METHODS

# Locaccino (p1)

- **Web-application for Facebook**
  - **Developed by Mobile Commerce Lab at CMU**
- **Allows users to share location**
  - **Facebook controlled privacy rules**

- ***Web Application* –** Query friends' locations
- ***Locator Software* –** Updates user location
  - **Runs on laptops and mobile phones**

# Locaccino (p2)

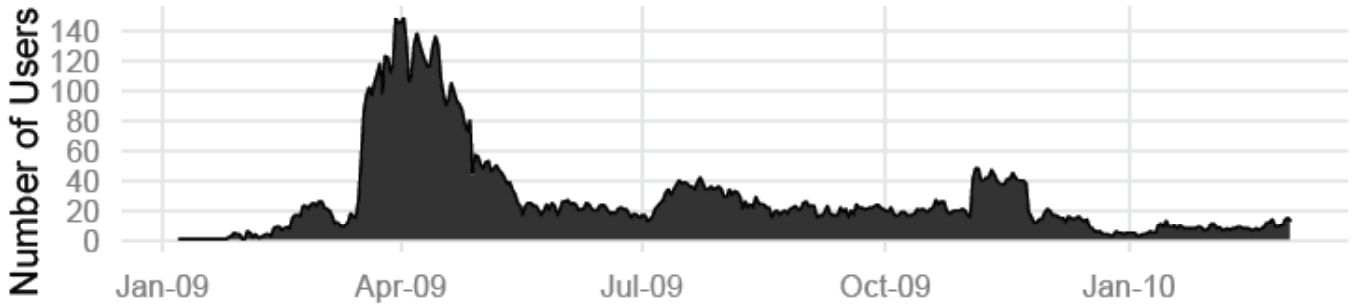
- Runs in background of device
- Updates every 10 minutes

- Uses combination of:
  - GPS (~10m-15m)
  - WiFi (~10m-20m)
  - IP (city or neighborhood)

- Sends time, latitude and longitude

# Demographics

- 489 users from 7 days to several months
- Mostly from university campus



# Data Collection

- 3 million location observations
  - 2 million in Pittsburgh
  - 20 years of human observational data

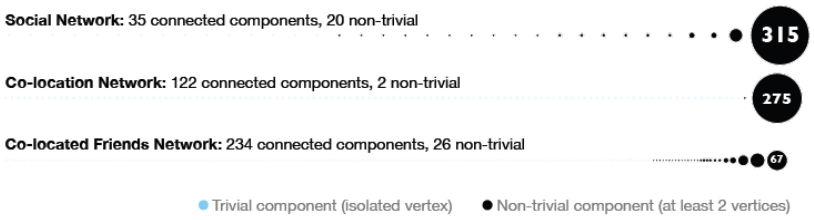
- Divide lat. and lon. into 30m x 30m grid
- Use 10 min. interval for time coordinate

- Co-location = same grid + same time

# The Networks…

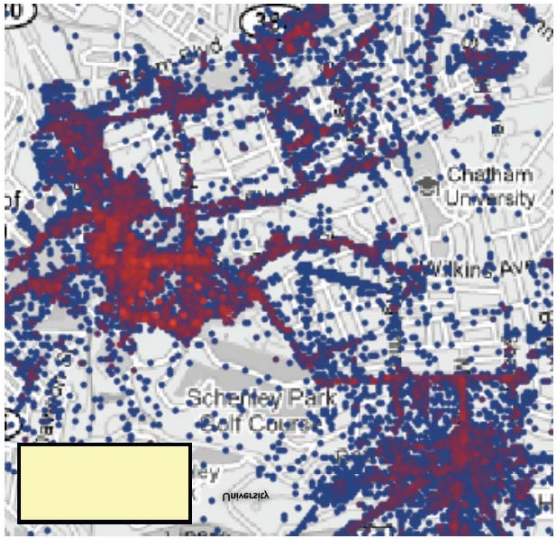
- **Social Network (S)** – Friends in Facebook
- **Co-location Network (C)** – Co-located at least once
- **Co-located Friends Network (S ∩ C)** – Friends and co-located

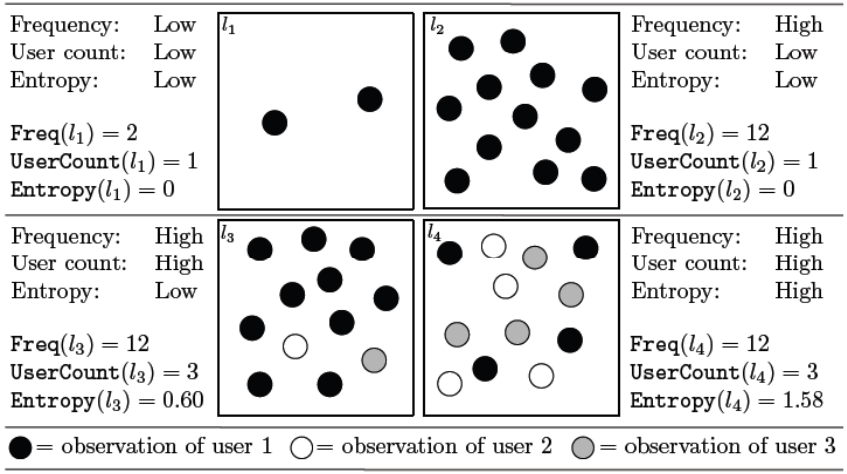
| Graph Structural Properties | $\mathcal{S}$ | $\mathcal{C}$ | $\mathcal{S} \cap \mathcal{C}$ |
|---|---|---|---|
| Number of vertices | 397 | 397 | 397 |
| Number of isolated vertices | 15 | 120 | 206 |
| **Number of edges** | **1063** | **3636** | **307** |
| Num connected components | 106 | 108 | 234 |
| Largest component size | 315 | 275 | 67 |
| Density | 0.014 | 0.046 | 0.004 |
| Connectedness | 0.63 | 0.48 | 0.04 |
| Degree centralization | 0.06 | 0.22 | 0.03 |
| Eigenvector centralization | 0.42 | 0.21 | 0.50 |



# Location Diversity

- **Frequency** – Raw count of observations
- **User Count** – Total unique visitors
- **Entropy** – Number of users and proportions of their observations



# Measured Features

- **Intensity and Duration** – Intensity of and range of user's use of system
- **Location Diversity** – Frequency, user count and entropy
- **Mobility Regularity** – Size and entropy of user schedule
- **Specificity** – How specific a location is to given co-location
- **Structural Properties** – Measures the strength of a relationship

# RESULTS

# Classifiers

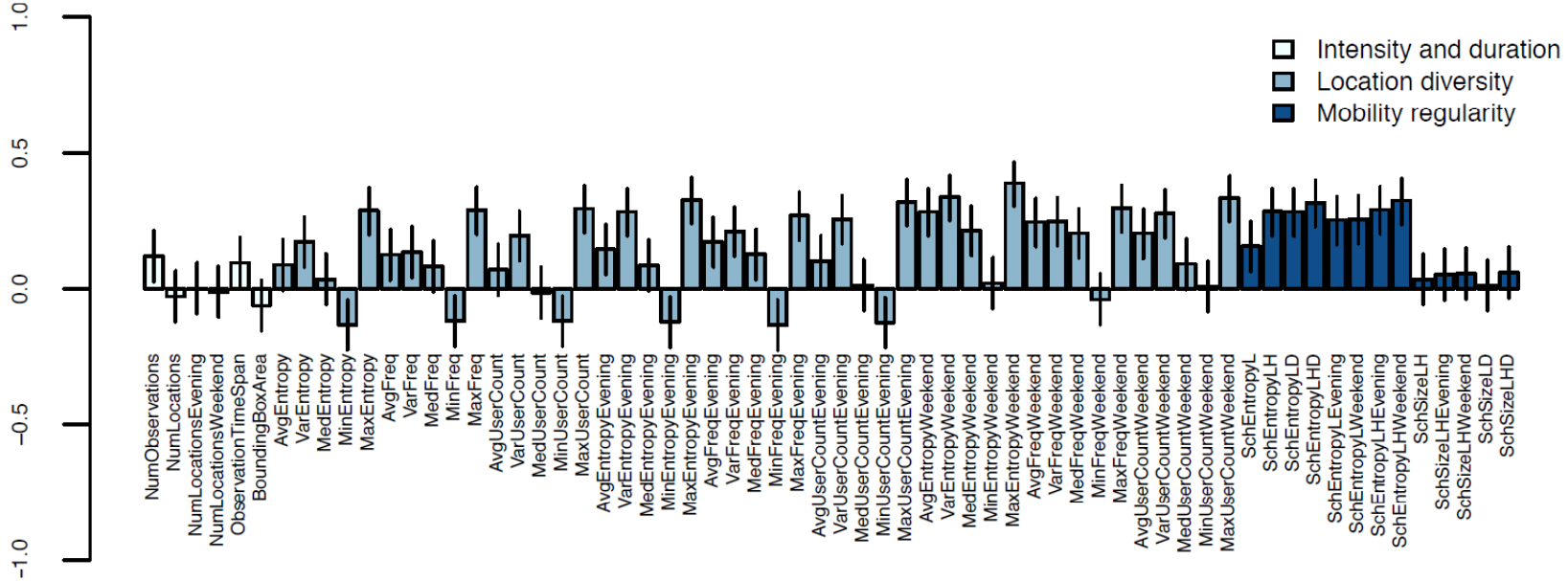
- 50-fold cross validation
- SVM performed the worst
- AdaBoost the best
  - However is skewed to guess better on non-friendships

| Classifier | Prec. | Recall |
|---|---|---|
| RandomForests (10 vars. per node) | 0.62 | 0.22 |
| RandomForests (18 vars. per node) | 0.61 | 0.22 |
| AdaBoost (dec. stumps, exp. loss) | 0.68 | 0.24 |
| AdaBoost (dec. stumps, lgstc. loss) | 0.60 | 0.28 |
| SVM (deg 2 polynomial kernel) | 0.40 | 0.31 |
| SVM (deg 3 polynomial kernel) | 0.26 | 0.37 |

# Inferring Number of Friends

- Look to relate number of Facebook friends to mobility patterns
- Expectations:
  - Users who have used the system longer have more friends
  - Users who visit “high diversity” locations have more friends
  - Users with irregular schedules may have more friends (require help from Locaccino)

# Pearson Correlation of Features




- **Intensity and duration weakest**
- **MaxEntropy, MaxUserCount, MaxFreq best**
  - **Average performs decently**

# Number of Friends (Cont.)

- Location and diversity numbers based on global properties of location
  - Not each users' individual instance at location
- Location information highly important to number of friends
- Schedule irregularity shows more ties in social network
- Number of friends not tied to heavy system use

# CONCLUSIONS

# Conclusions (p1)

- **Found the co-location network 3x larger than social network (edge-wise)**
  - **Social network better connected**
- **Properties of location are crucial**
  - **Especially Entropy**
  - **Difference between high and low entropy**
  - **Help define both relationships and number of friends**

# Conclusions (p2)

- **Created set of features to help classify social network friends**
  - **Better than by simple co-location observations**
- **Found interesting patterns**
  - **Co-location without friends**
  - **Friends without co-location**

# Future Work (p1)

- Use classifiers for social network friend recommendation system
  - Augment and expand current friend-link system in place
- Could help provide insight into strength of relationship
  - Still requires more research and validation
  - Develop system for segregating and categorizing friends
  - Help with privacy rules

# Future Work (p2)

- **Build off relationship between online and offline social behavior**
  - **Using things such as entropy of a location**
- **Use of location patterns of users**
  - **Suggest similar locations to friends**
  - **Suggest similar locations to non-friends with similar behavior**