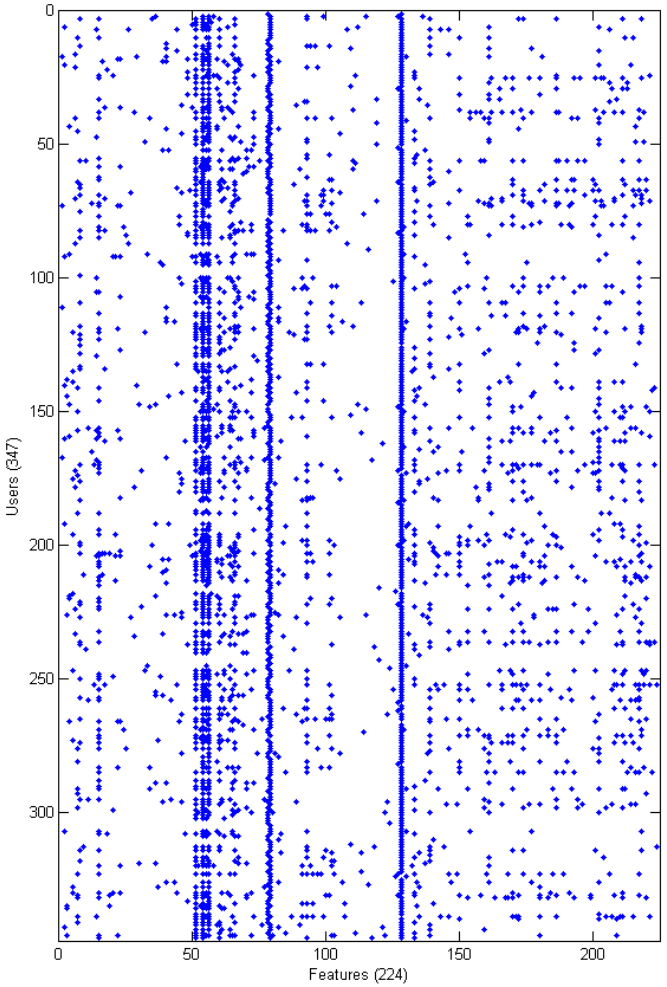
Discovering  riendships

Drew Wicke
Mike Sharov

Dataset

- 347 Facebook users split into friendship circles
- 224 anonymized binary features grouped into categories

```
birthday;
education;classes;id;
education;concentration;id;
education;degree;id;
education;school;id;
education;type;
education;year;id;
first_name;
gender;
hometown;id;
languages;id;
last_name;
locale;
location;id;
work;employer;id;
work;end_date;
work;location;id;
work;start_date;
```



[1] http://snap.stanford.edu/data/egonets-Facebook.html

Objective

- Unsupervised clustering based on feature data alone

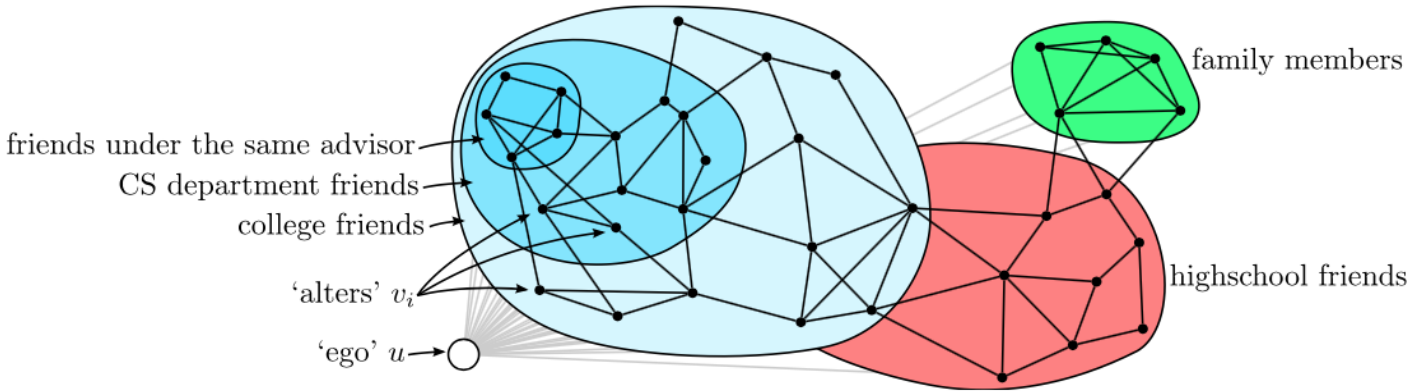


Image from:
Jure Leskovec and Julian J. Mcauley. Learning to Discover Social Circles in Ego Networks. .p. 548--556 2012 http://books.nips.cc/papers/files/nips25/NIPS2012_0272.pdf

Approaches

- K-Means
- DBSCAN
- SOM
- Spectral Clustering

Accuracy Measure

$$u_{ij} = |C_i \cup O_j| \qquad for\ i = 1 \ldots m,\ \ j = 1 \ldots n$$

$$s_{ij} = \frac{u_{ij}}{max(|C_i|, |O_j|)}$$

$$a_i = max(s_{i1},\ s_{i2}, \ldots, s_{in})$$

$$r = \frac{1}{m} \sum_{i=1}^{m} a_i$$

NOTE:

C_i is the *i*'th cluster we identified and O_j is an actual user specified cluster

Selected Results

Conclusions