HACK MACHINE LEARNING

Hacking Your Facebook Data

Tuesday 02

6:30 PM
1010 DOW

Bernease Herman
@bernease
github.com/bernease

What is Machine Learning?

Machine Learning vs. Data Mining

- Machine learning focuses on prediction, based on *known* properties learned from the training data.

- Data mining (which is the analysis step of Knowledge Discovery in Databases) focuses on the discovery of (previously) *unknown* properties on the data.

"Machine learning" Wikipedia, accessed April 1, 2013

## Machine Learning vs. Big Data

- Information architecture, parallel computing, different kind of clustering

- **Myth:** "It's not the size of the data, but the way you use it."

Defining the Problem

- **Feature set:** known characteristics/facts about event/item, independent variables
- **Label:** description of event/item, dependent variable

- **Supervised Learning:** training set with labeled data
- **Semi-supervised Learning:** small amount of labeled data, mixture of supervised and unsupervised
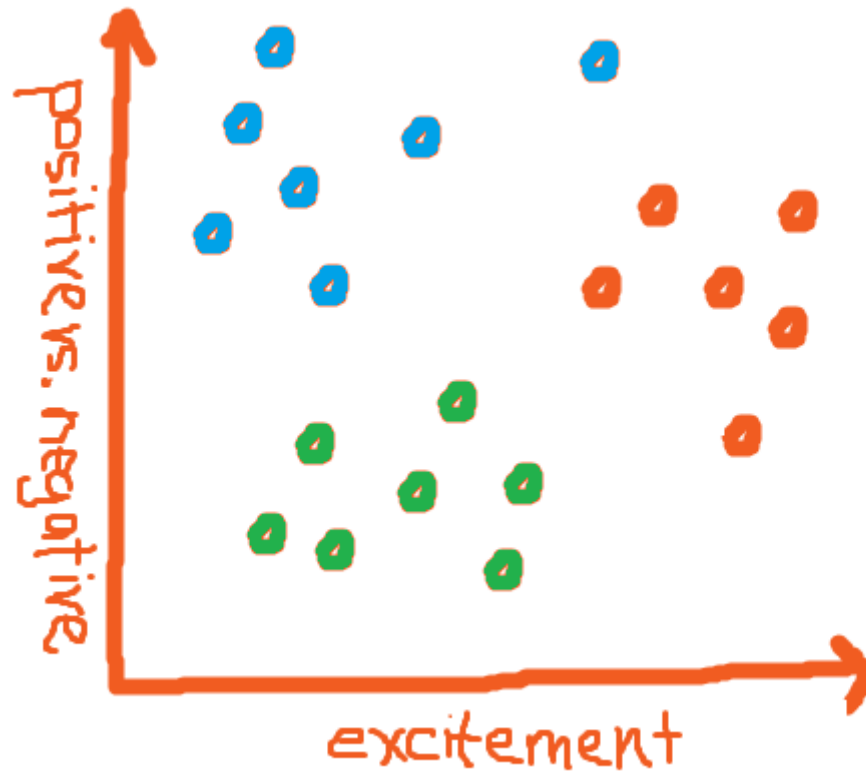- **Unsupervised Learning:** no labeled data, e.g. clustering
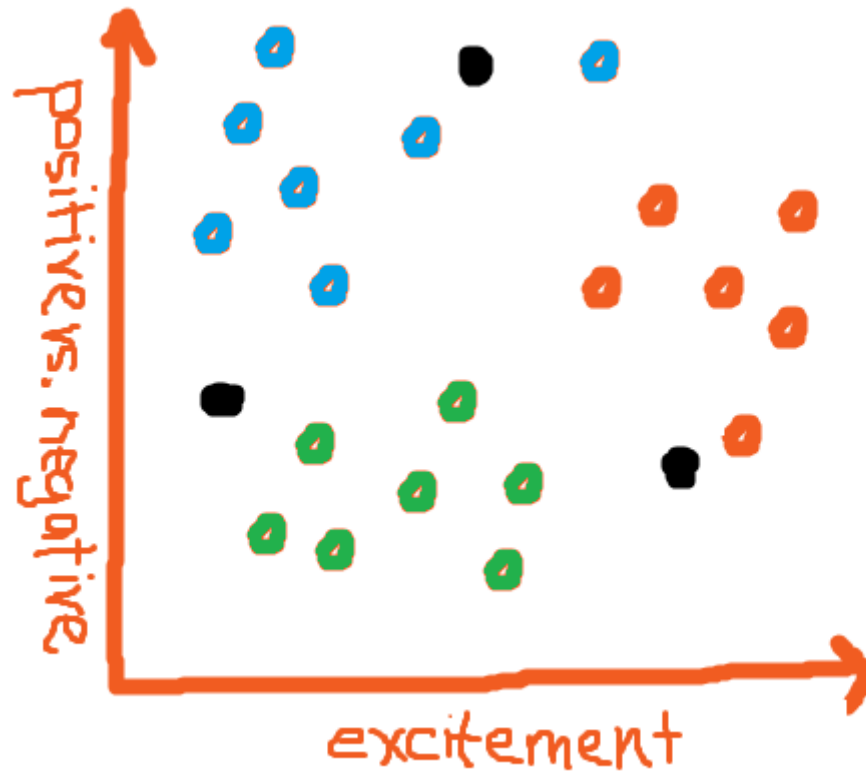
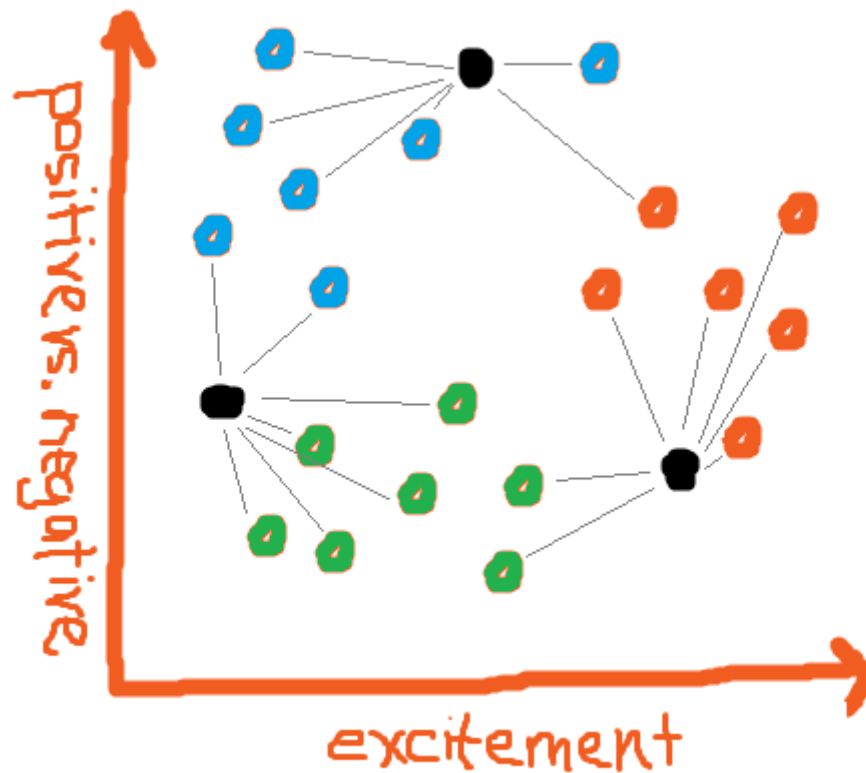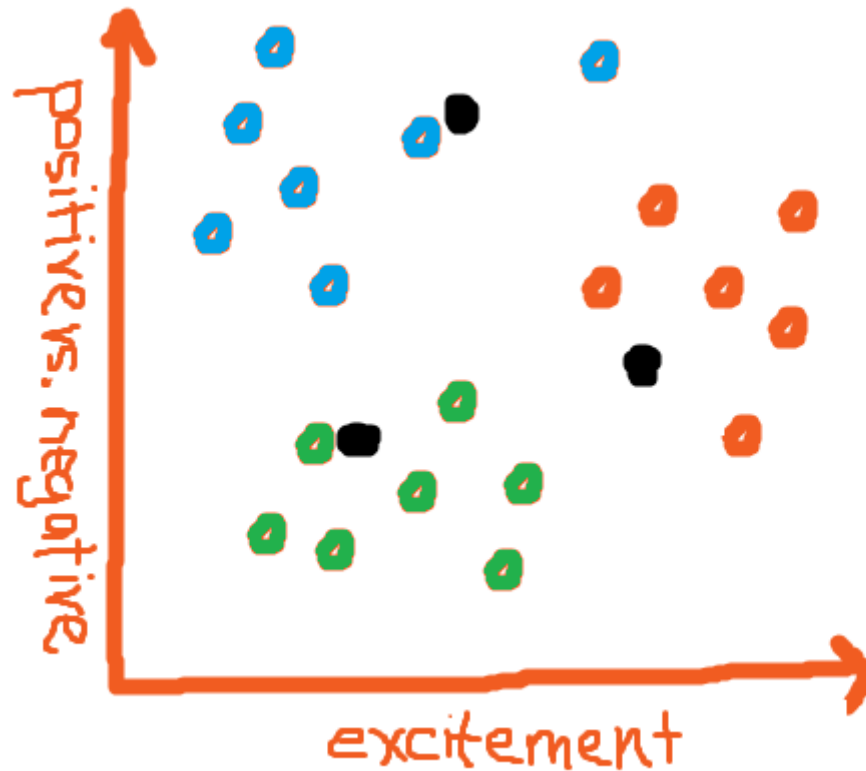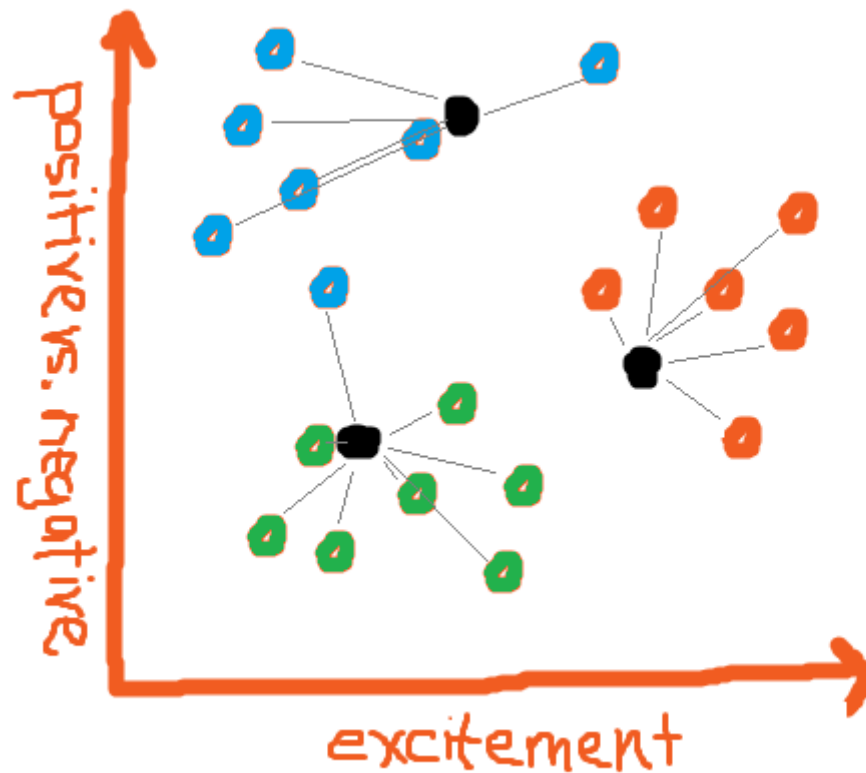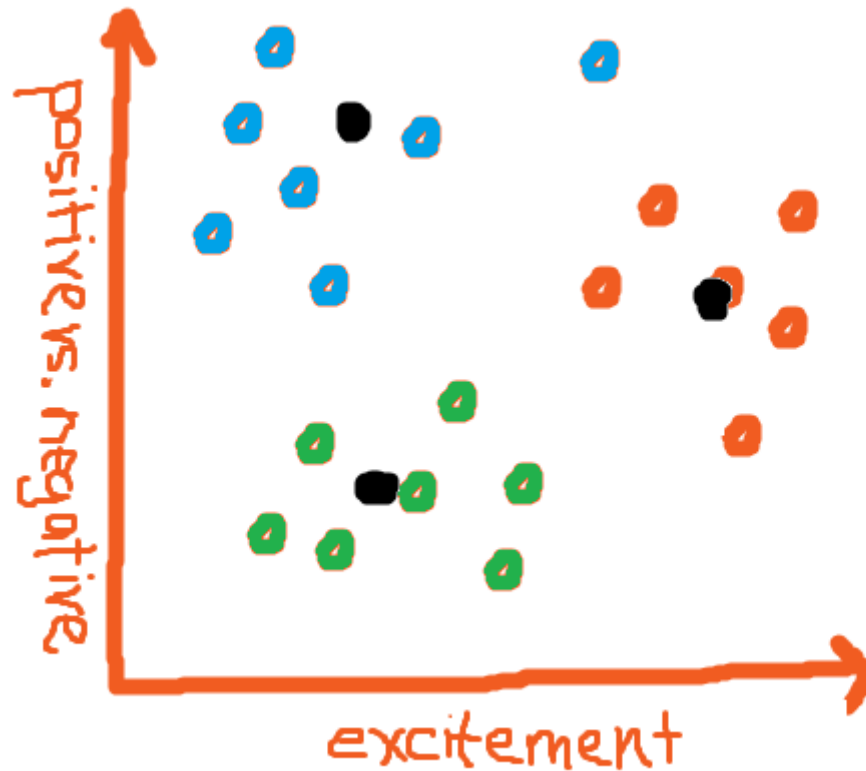# Clustering Method
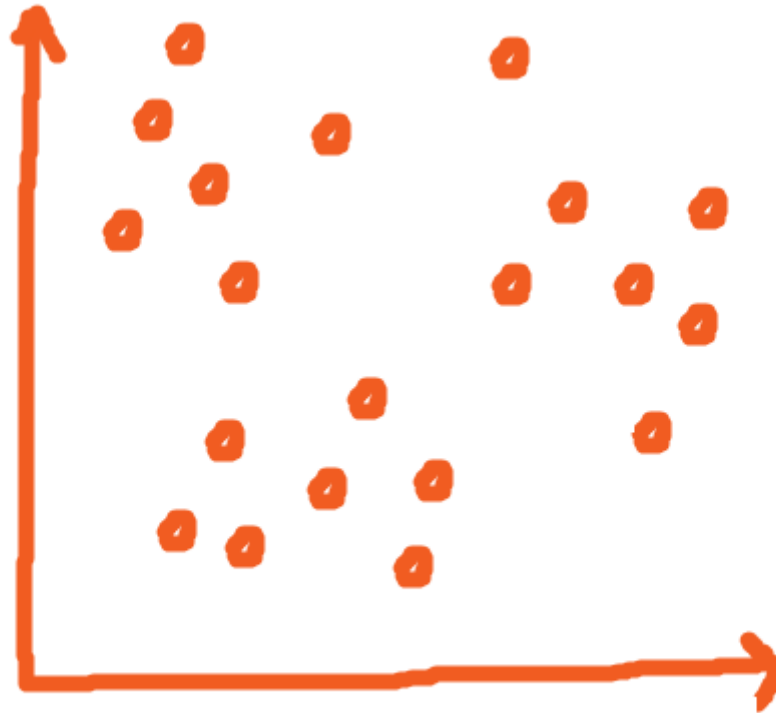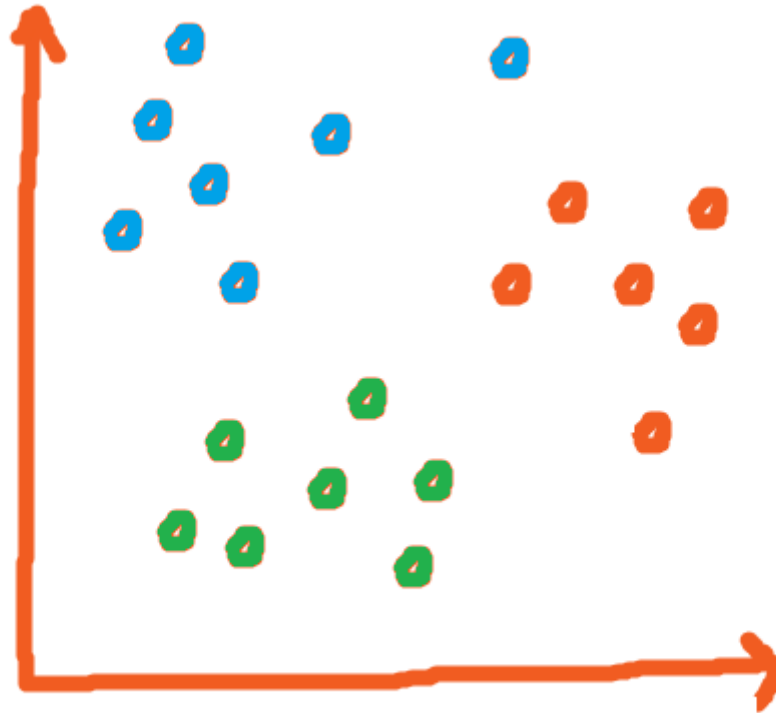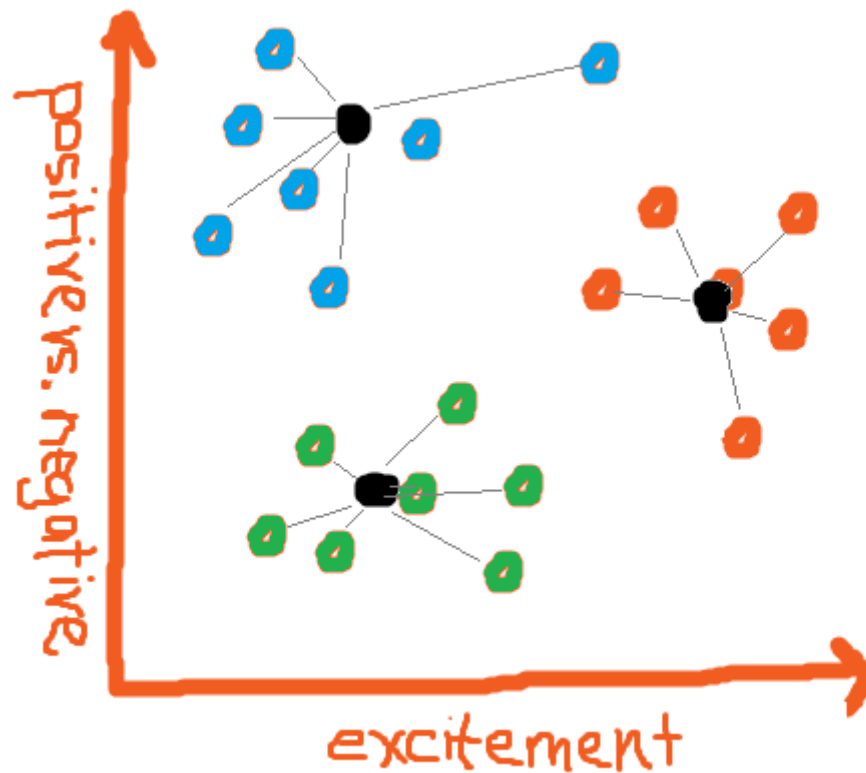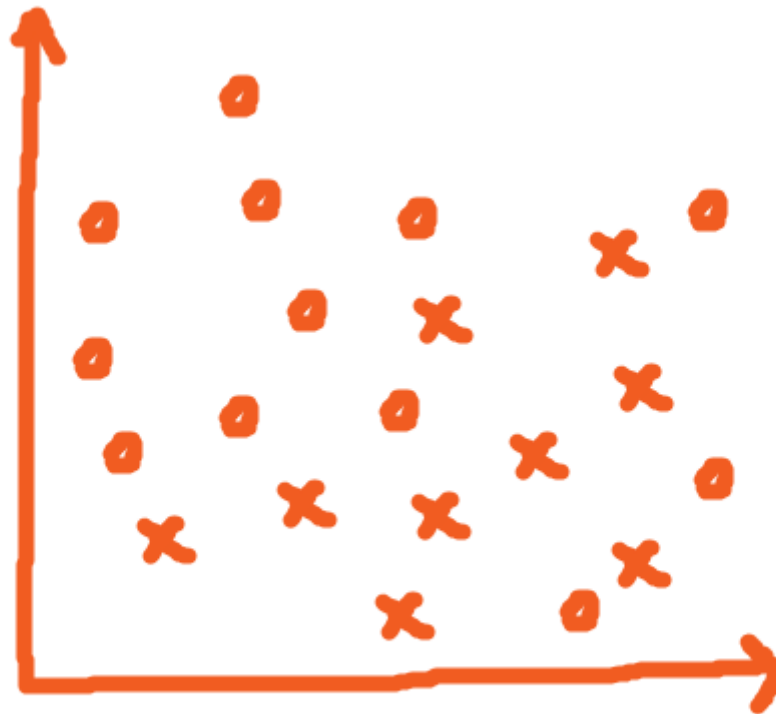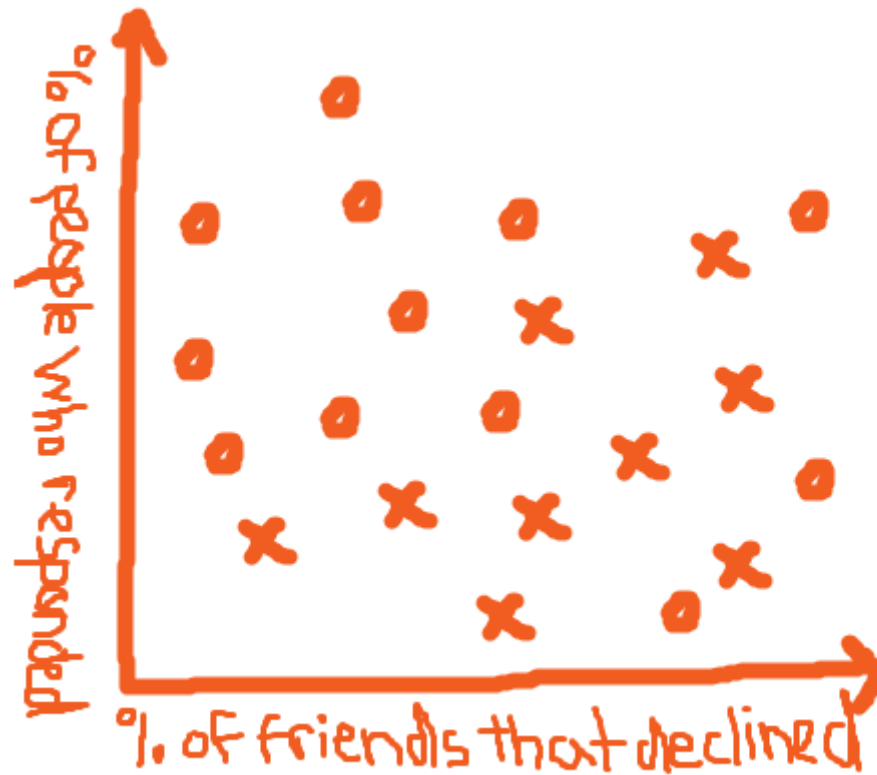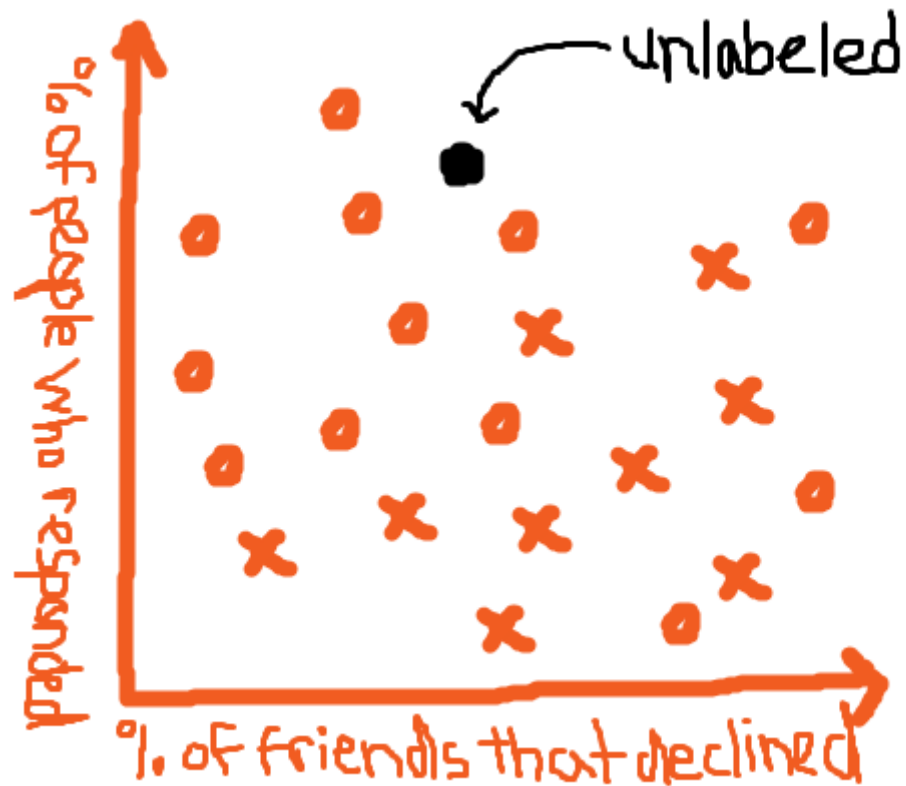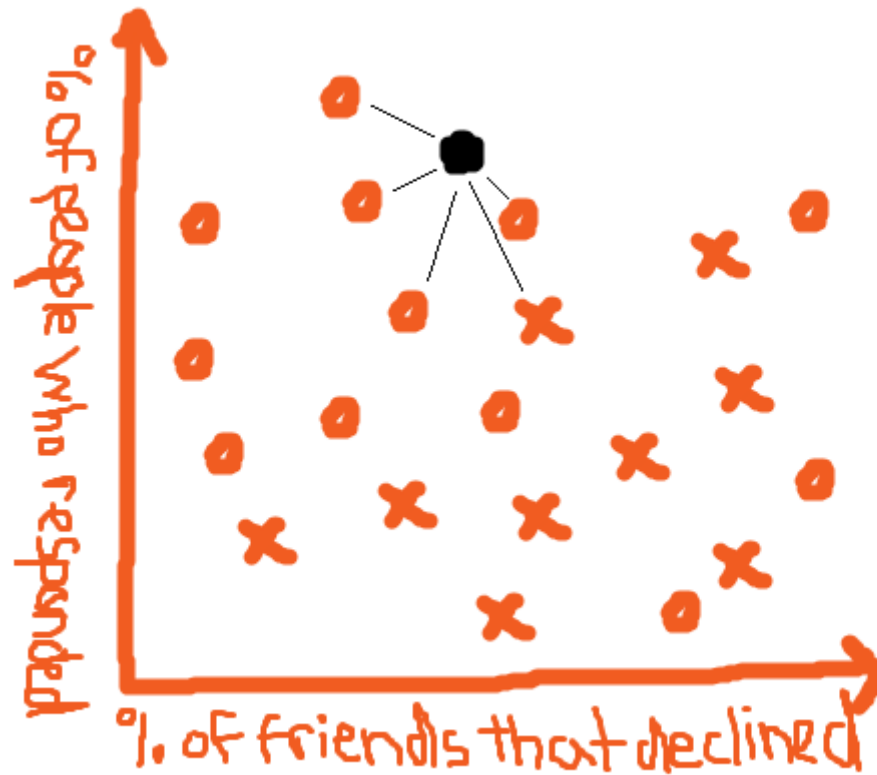
Clustering Method

Clustering Method

Clustering Method

Clustering Method

Clustering Method

Clustering Method

Clustering Method

Clustering Method

Classification Method

Classification Method

Classification Method

Classification Method

Classification Method
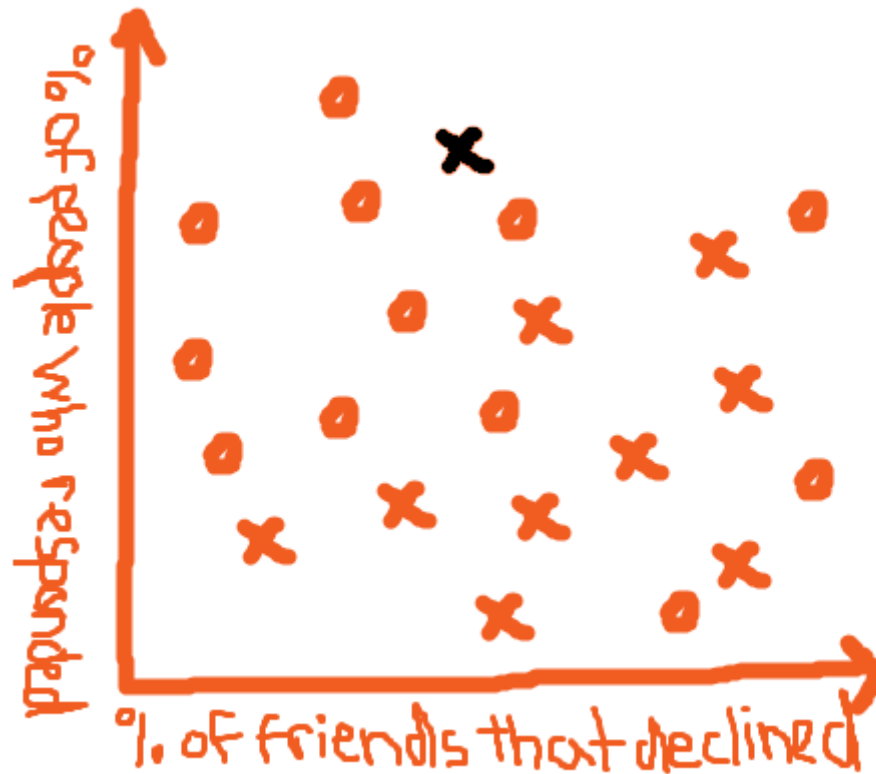
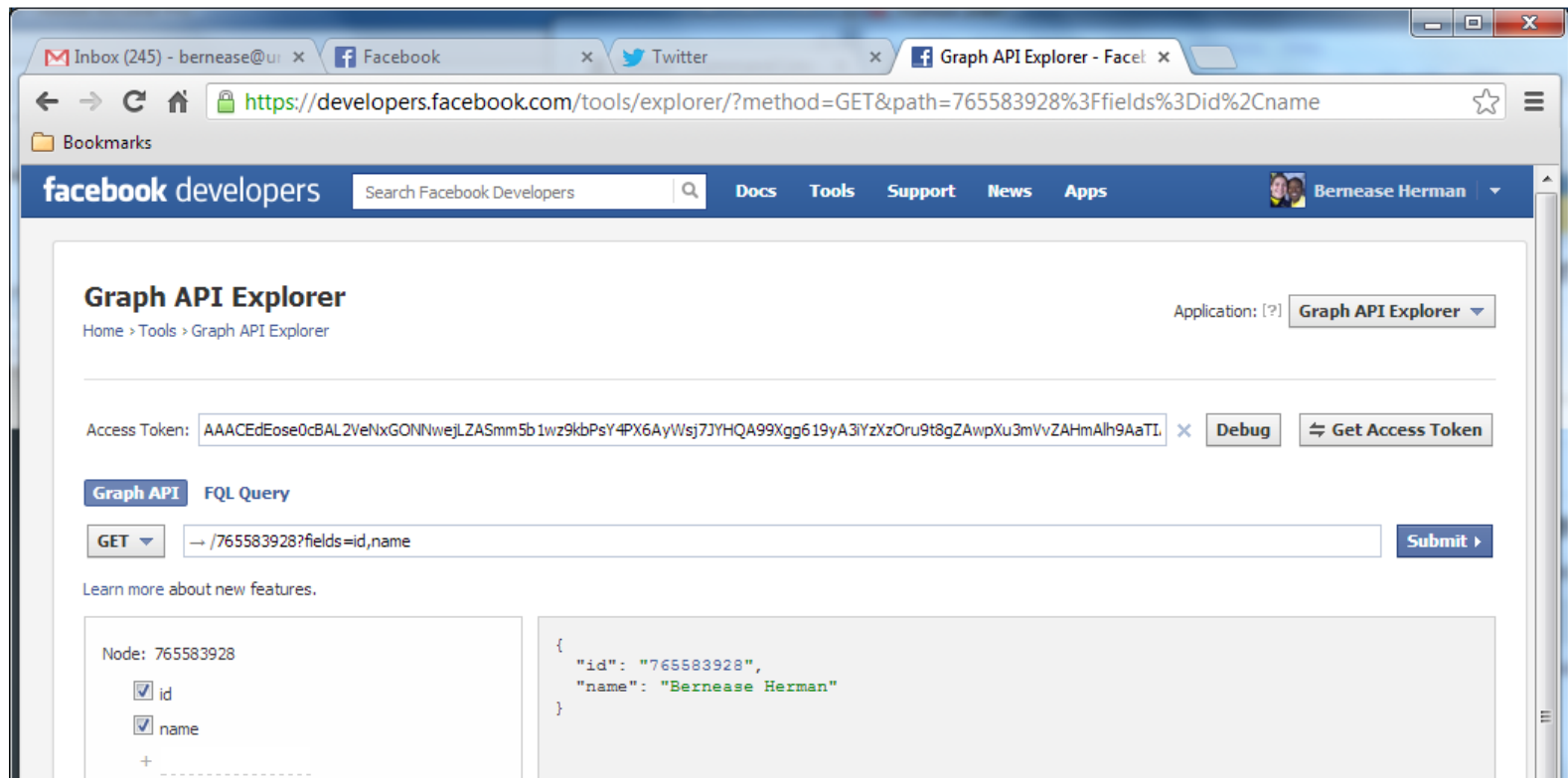# Collaborative Filtering Method

Collaborative Filtering Method
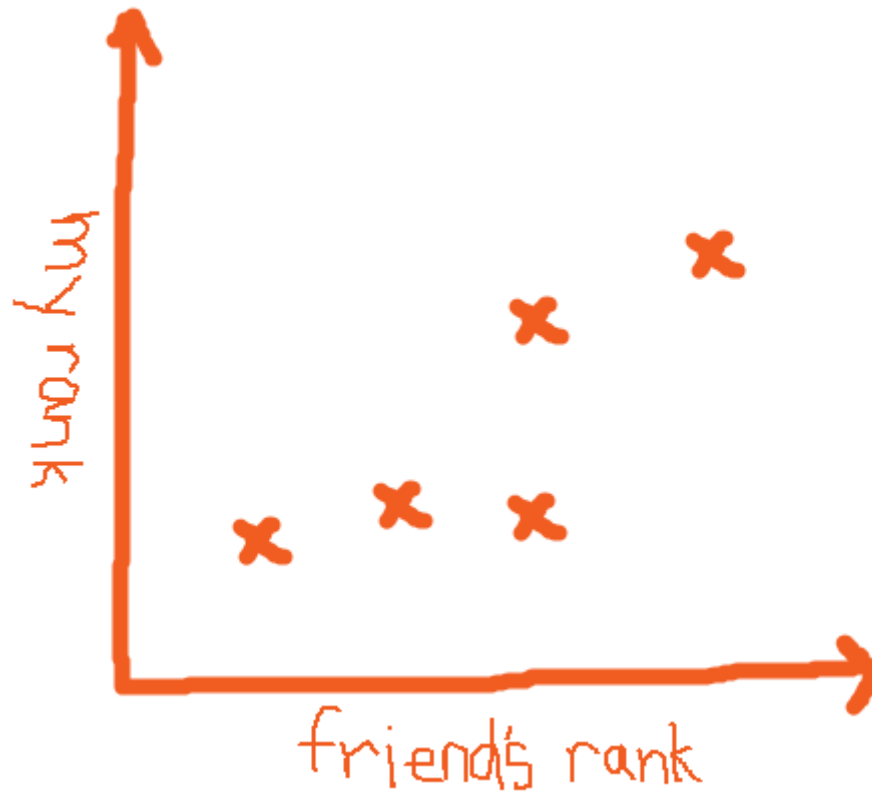
Collaborative Filtering Method

Getting Your Data

https://developers.facebook.com/tools/explorer/

Getting Your Data

https://developers.facebook.com/tools/explorer/

Getting Your Data

https://developers.facebook.com/tools/explorer/

Getting Your Data

https://developers.facebook.com/tools/explorer/

Getting Your Data