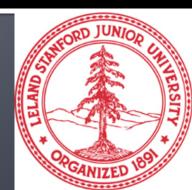
Analysis of Large Graphs: Detecting clusters

Mining of Massive Datasets Leskovec, Rajaraman, and Ullman Stanford University



New Topic: Graph Data!

High dim.

Locality sensitive hashing

Clustering

Dimensional ity reduction

Graph data

PageRank, SimRank

Community Detection

Spam Detection

Infinite data

Filtering data streams

Web advertising

Queries on streams

Machine learning

SVM

Decision Trees

Perceptron, kNN

Apps

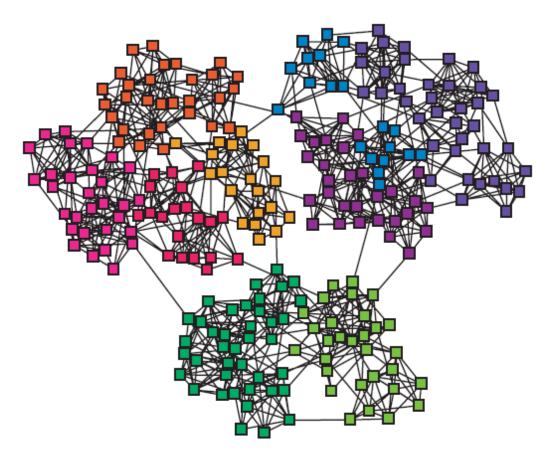
Recommen der systems

Association Rules

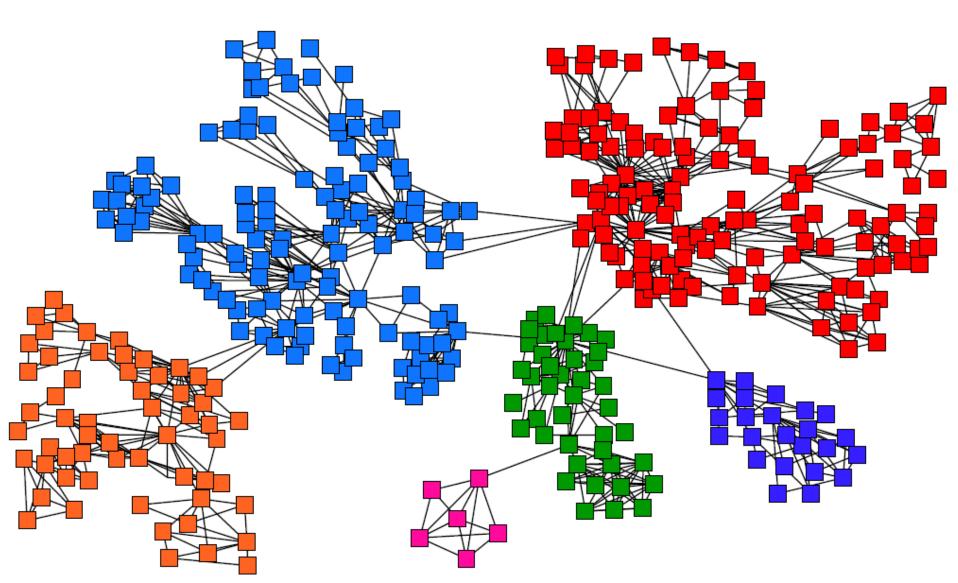
Duplicate document detection

Networks & Communities

 We often think of networks being organized into modules, cluster, communities:

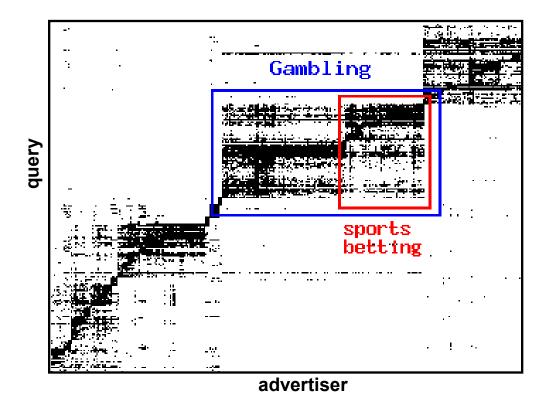


Goal: Find Densely Linked Clusters



Micro-Markets in Sponsored Search

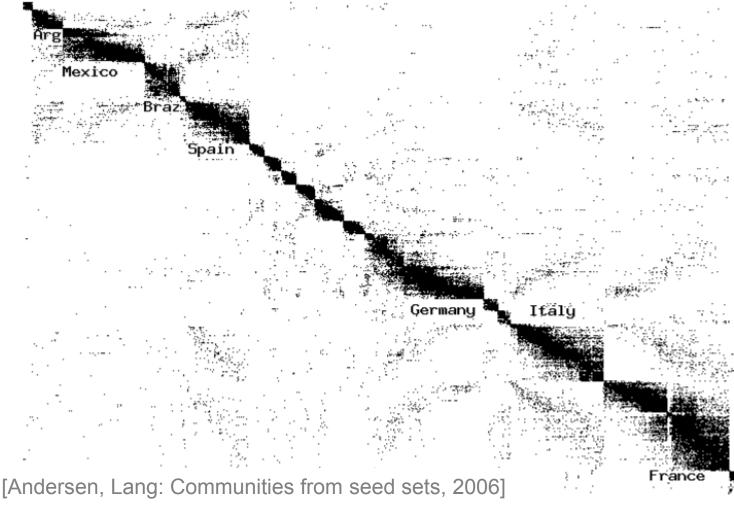
Find micro-markets by partitioning the query-to-advertiser graph:



[Andersen, Lang: Communities from seed sets, 2006]

Movies and Actors

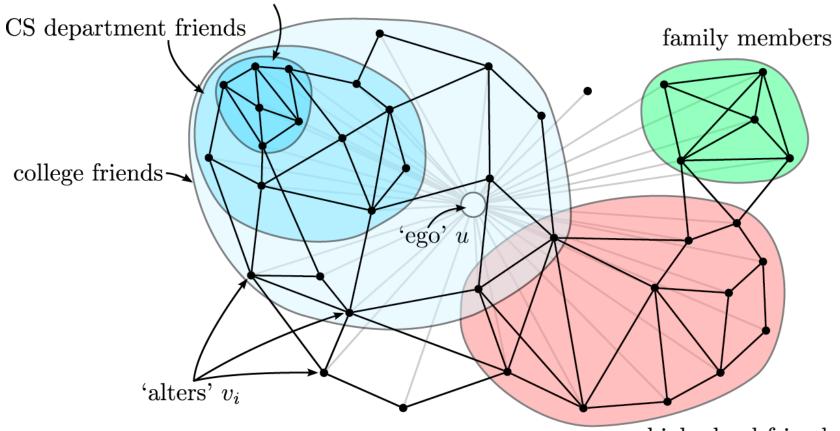
Clusters in Movies-to-Actors graph:



Twitter & Facebook

Discovering social circles, circles of trust:

friends under the same advisor



highschool friends