



Clustering

Introduction to Data Science Algorithms

Jordan Boyd-Graber and Michael Paul

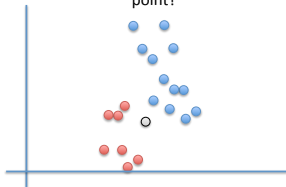
SLIDES ADAPTED FROM DAVE BLEI AND LAUREN HANNAH

Lecture for Today

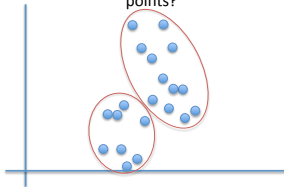
- What is clustering?
- K-Means
- Gaussian Mixture Models

Clustering

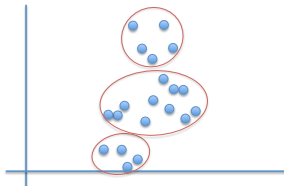
Classification: what is label of new point?



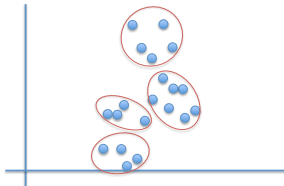
Clustering: how should we group these points?



Clustering: or is this the right grouping?



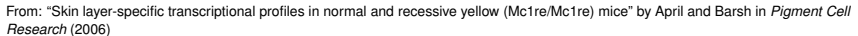
Clustering: what about this?



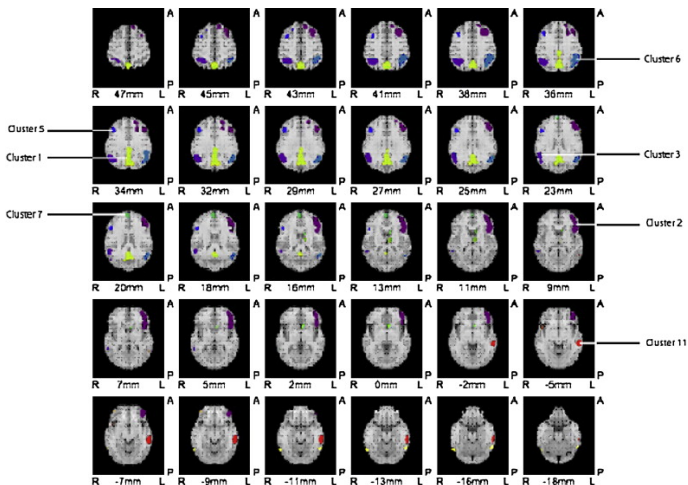
Clustering

Uses:

- genomics
- medical imaging
- social network analysis
- recommender systems
- market segmentation
- voter analysis

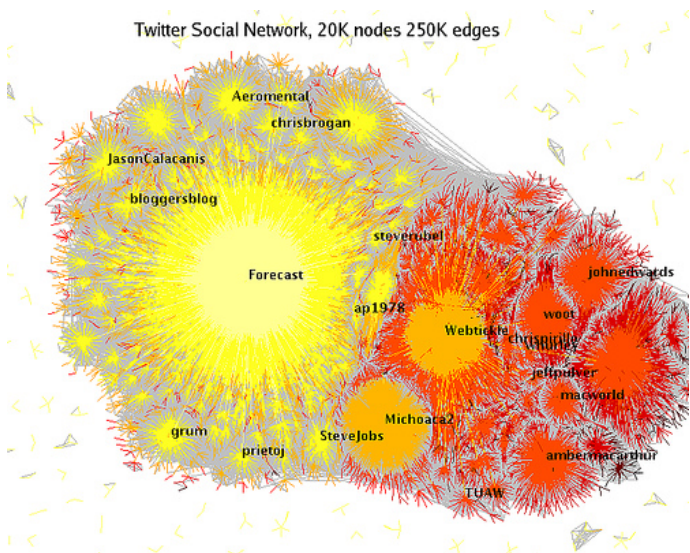


Medical Imaging (MRIs and PET scans)

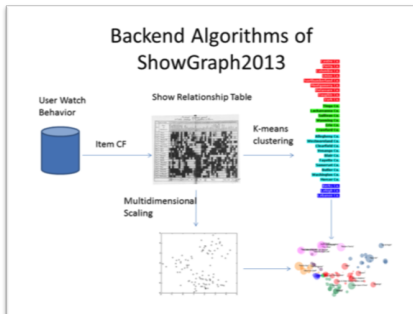


From: "Fluorodeoxyglucose positron emission tomography of mild cognitive impairment with clinical follow-up at 3 years" by Pardo et al. in *Alzheimer's and Dementia* (2010)

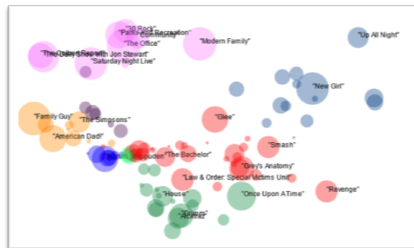
Social Networks



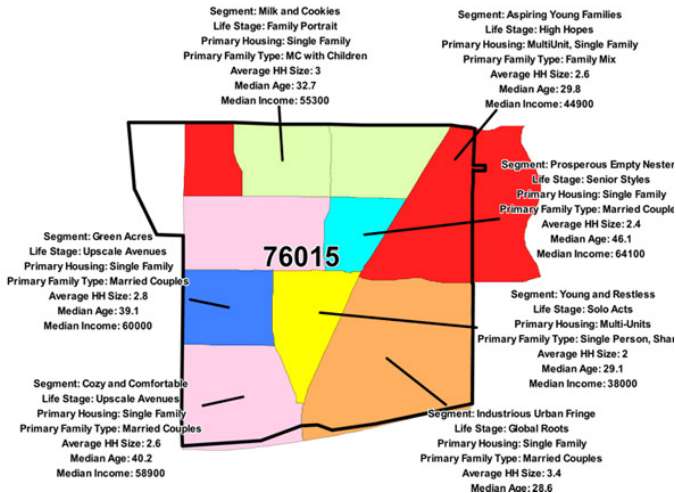
Recommender Systems



From: tech.hulu.com/blog/

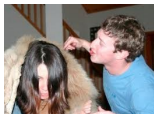
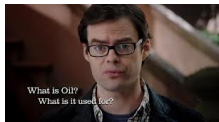


Market Segmentation



From: mappinganalytics.com/map/segmentation-maps/segmentation-map.html

Voter Analysis



- soccer moms (female, middle aged, married, middle income, white, kids, suburban)
- Nascar dads (male, middle aged, married, middle income, white, kids, Southern, suburban or rural)
- security moms (...)
- low information voters
- Ivy League Elites