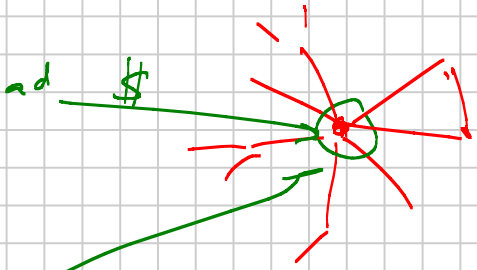
Community Detection

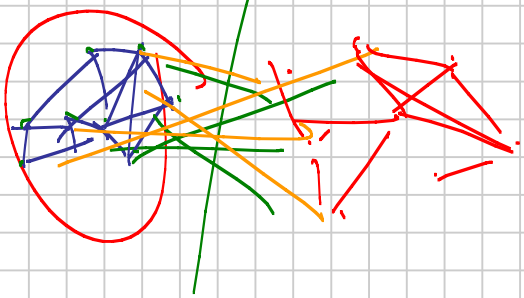
Note Title 4/18/2016

$G = (V, E)$

- social network
  - → Facebook, Twitter, Instagram
  - → Road Networks, Railroad
    Telephone
  - Connections ⟹ useful for targeted ads



SCV arteries

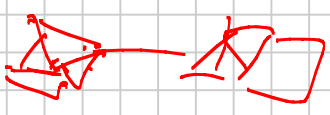


- spectral clustering (disjoint / complete)
- pagerank (on edges)

→ some in multiple communities
→ some (many) in no tight community

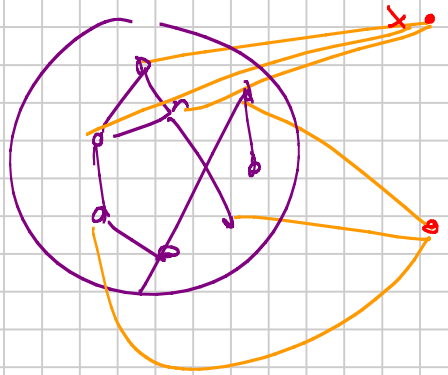
History of Graph Analysis

- 50 years ago | $|V| = 200$
- usually draw



- Math (Random Graphs)
  Erdos-Renyi: n vertices, m edges ← random

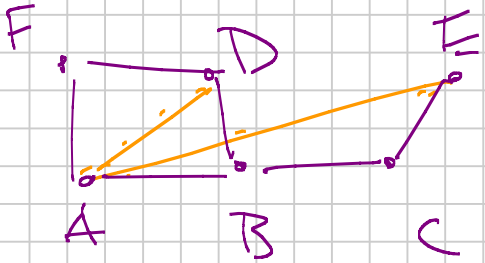
Why do people join groups?



Who more likely to join?

- x close community
- y has diverse support

Preferential Attachment



edge (A, D) more likely than (A, E).

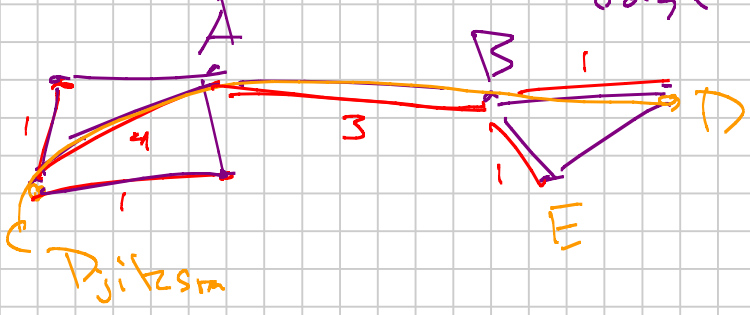
Likely to form triangles

Kronicker Product Graphs

Betweenness

- Find important edges

Betw(A, B) = fraction of all shortest paths that use this edge



$O(|V| \cdot |E|)$

- Run Djikstra for all vertices
- count backwards
- Divide by $\binom{n}{2}$

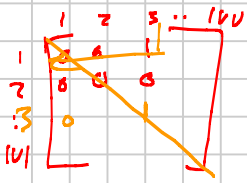
Modularity

SCV

$Q(S) =$ (fraction of edges in S) − (Expected # edges in S)

$G = (V, E)$

↳ adjacy matrix A

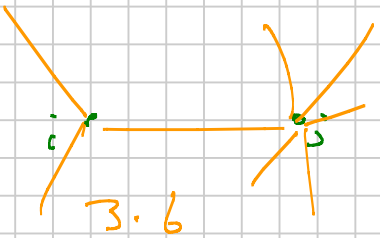


$\#$ edges in S $\quad \sum_{i \in S} \sum_{j \in S} A_{ij}$ $\Rightarrow$ frac edges in S $\quad \sum_{i \in S} \sum_{j \in S} \frac{A_{ij}}{2|E|}$

$\#$ edges $= 3$

Expected # edges

$$E_{ij} = \frac{d_i \cdot d_j}{2|E|}$$



$\frac{3 \cdot 6}{2 \cdot |100|}$

degree $d_j = 6$ = # edges w/ j

Random Graph Model

Input $V, \{(v_1, d_1), (v_2, d_2) \ldots, (v_n, d_n)\}$

Kathy Perry 86 million 159

Justin Bieber 79 mill 261 K

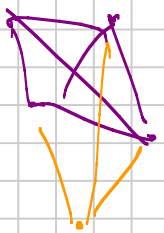
$$Q(S) = \frac{1}{2|E|}\left[\sum_{i \in S} \sum_{j \in S} (A_{ij} - E_{ij})\right]$$

$\in [-1/2, 1]$ typically $Q \in [0.3, 0.7]$

↳ community

no parameter.

Clique



4-clique

NP-hard

Build up

- Find 2-cliques (edges)
- Find all 3-cliques (triangles)
- Find all 4-cliques (or, 3-cliques)

largest in quantity

Apriori