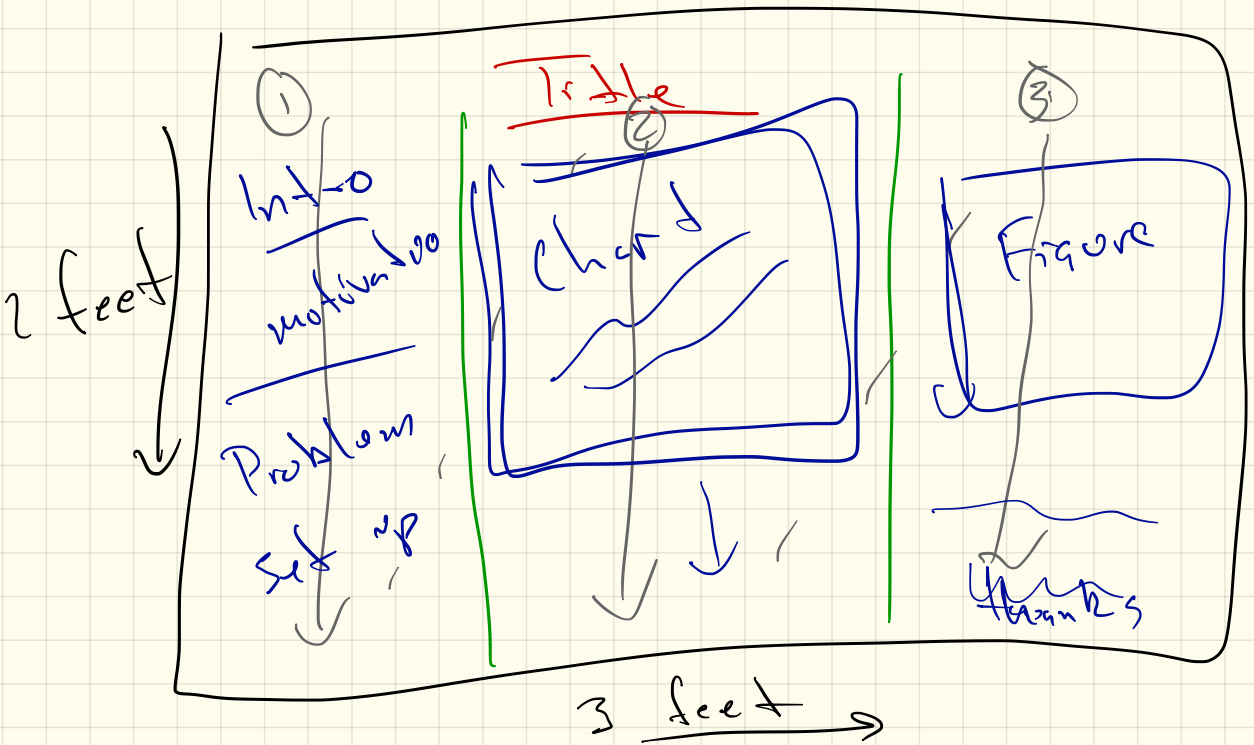
L24 : Graphs → Communities

April 15, 2020

Jeff M. Phillips

Posters

turn in pdf in canvas

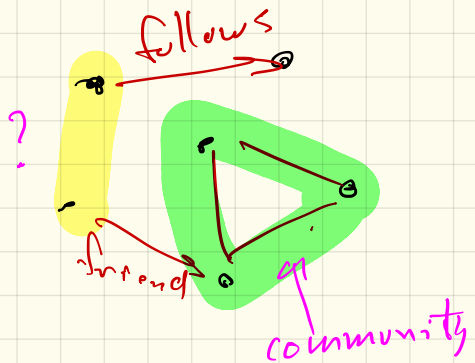


Communities in Graphs

Graph $G = (V, E)$

V: people / entities

E: ?



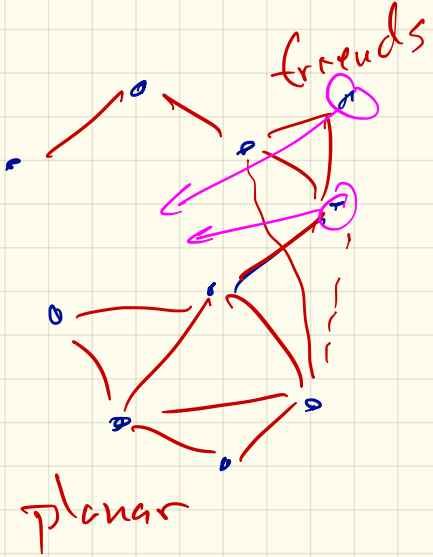
How edges are defined

modeling choice.

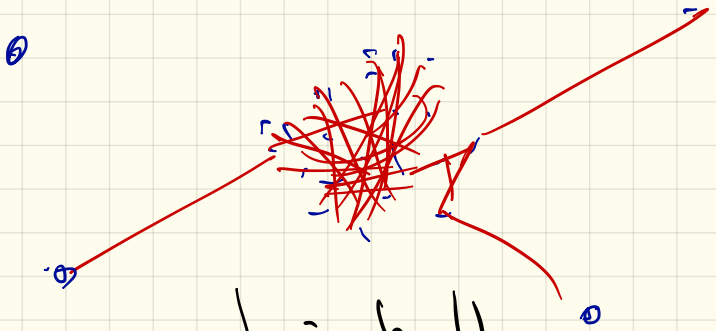
Early web business (.com - boom)
(circa 90s)

Sociology "social network"
manually collect
|V| ~ [100, 700]

seniors in high school



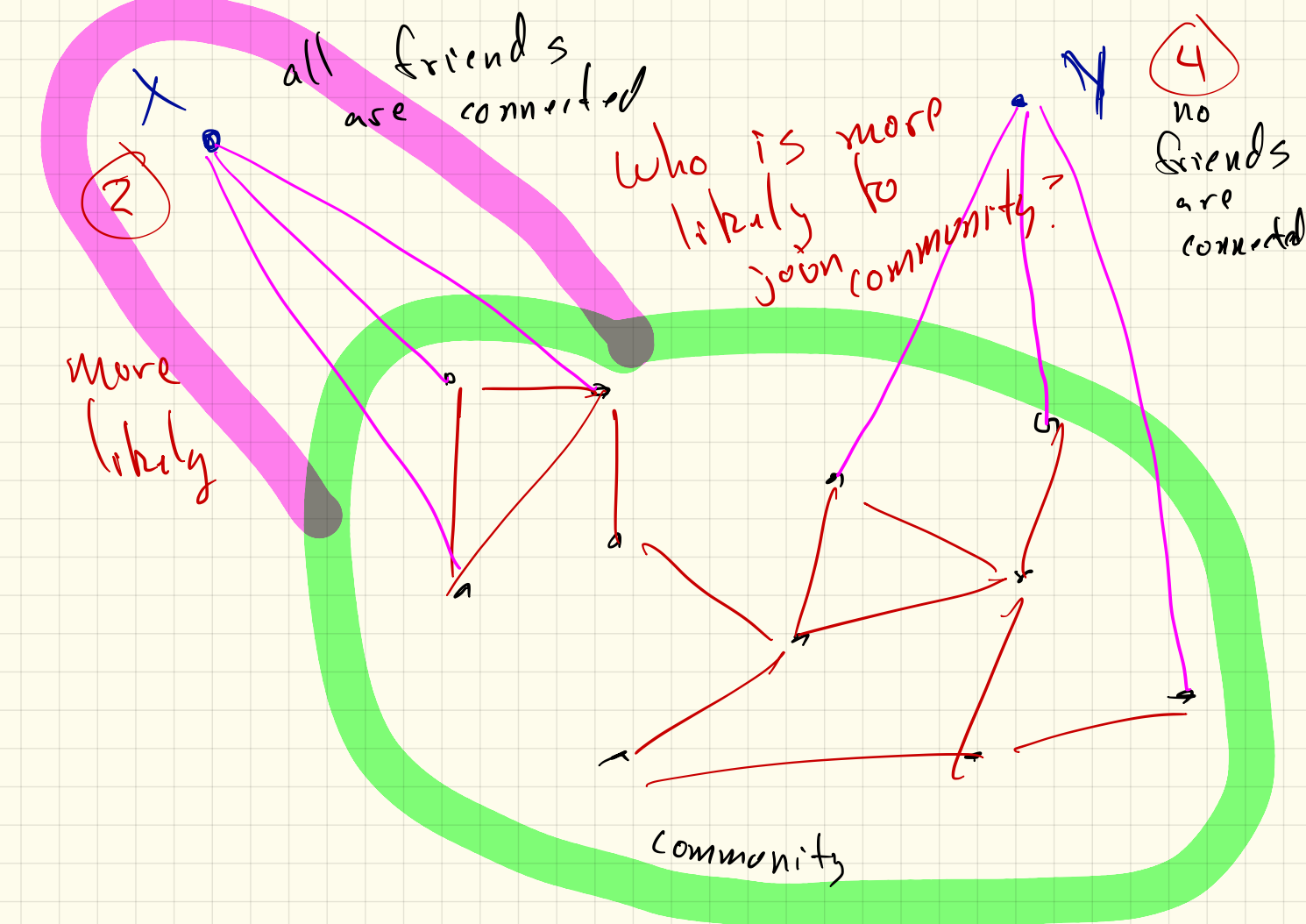
$|V| \geq 1000$



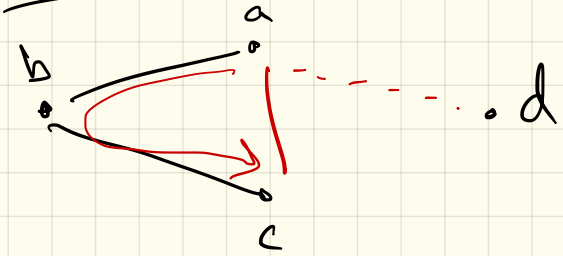
hairball

Mathematical Models of large graphs.

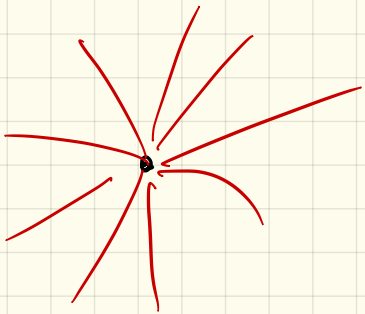
Erdös-Renyi: Pick $n = |V|$ → every edge exist independently w.p. p.
choose $p \in (0,1)$


X
all friends are connected
2
more likely
Who is more likely to join community?
Y
4
no friends are connected
Community

Prefecential Attachment



more likely to form edge (a,c) than (a,d)



average person on twitter followed by $\approx$ 500

most popular (Obama) 130 million followers

How do we find
- communities (Spectral clustering)
- important nodes/edges (pageRank)

Community = tightly connected subgraph $V' \subset V$

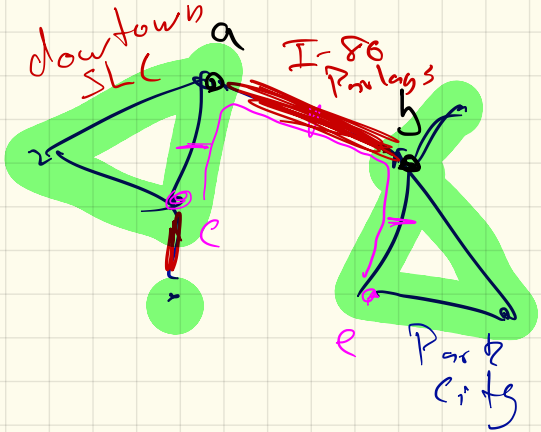
all edges defined by V'

Betweenness: important edges.

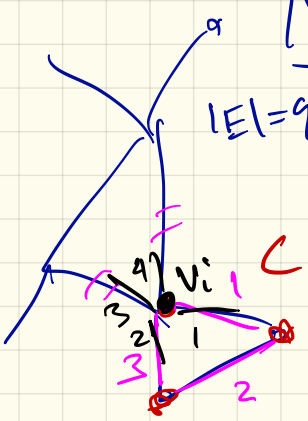
↳ remove important edges
↳ remaining connected components.
↳ communities

betw(a, b) =

fraction of all shortest paths that use edge (a, b)



Modularity $G=(V,E)$



$|E|=9$

score $Q(C)$ $C \subset V$

$\equiv$ (fraction of edges in G_C)

$-\frac{1}{|E|}\sum_{i,j\in C} A_{ij}$

(expected fraction of edges in G_C)

$\frac{1}{|E|}\sum_{i,j\in C} E_{ij}$

$d_i = 4$

adjacency matrix $A = A_{ij} = 1$ if $e_{ij} \in E$, else 0

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

d_i = degree of v_i

$$Q(C) = \frac{1}{4|E|}\left[\sum_{i,j \in C}\left(A_{ij} - \gamma E_{ij}\right)\right]$$

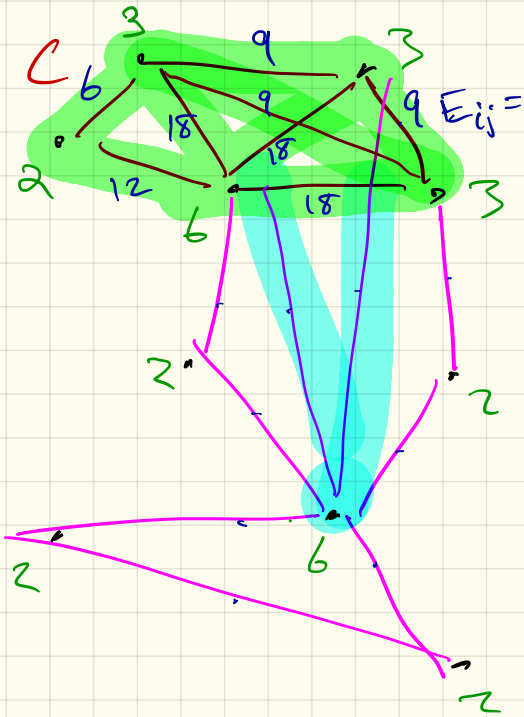
$4|E| = 68$

$Q(C) \in [-1/2, 1]$

$\sum_{i,j \in C} \left(A_{ij} - \gamma E_{ij}\right) = 8$

$\gamma = 3$

$\sum_{ij} E_{ij} = \frac{99}{34}$



$E_{ij} = \frac{3 \cdot 3}{2 \cdot 17}$

$$\frac{1}{68}\left[8 - \frac{99}{34}\right] \approx \frac{5}{68} \approx \frac{1}{14}$$

Find communities $C \subset V$ with largest $Q(C)$

Cliques

is set of vertices

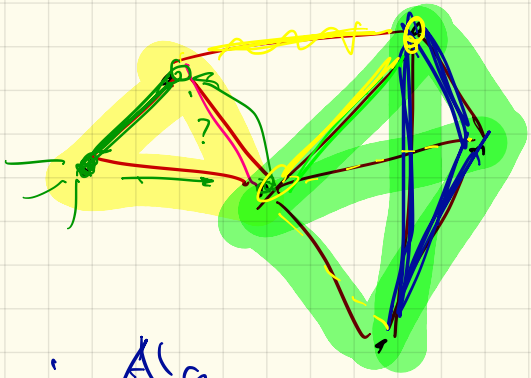
$C \subset V$

where all edges

$i, j \in C$

have edge $(i,j) \in E$

NP-hard



A priori Alg.

2-cliques ~~Start~~ w/ all edges

3-cliques Find all triples of edges which have same 3 vertices

4-cliques - use 3-cliques, expand greedily