L26: Communities in Graphs

Dec 1, 2025
Data Mining
Jeff Phillips

Graph to Model Social Network

$G = (V, E)$

V = people (or business, account)

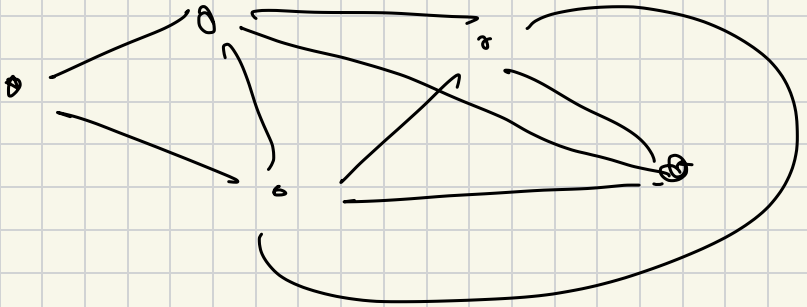
E = interactivity = "follows"

not so clear what defines edge.

Adjacency ✗

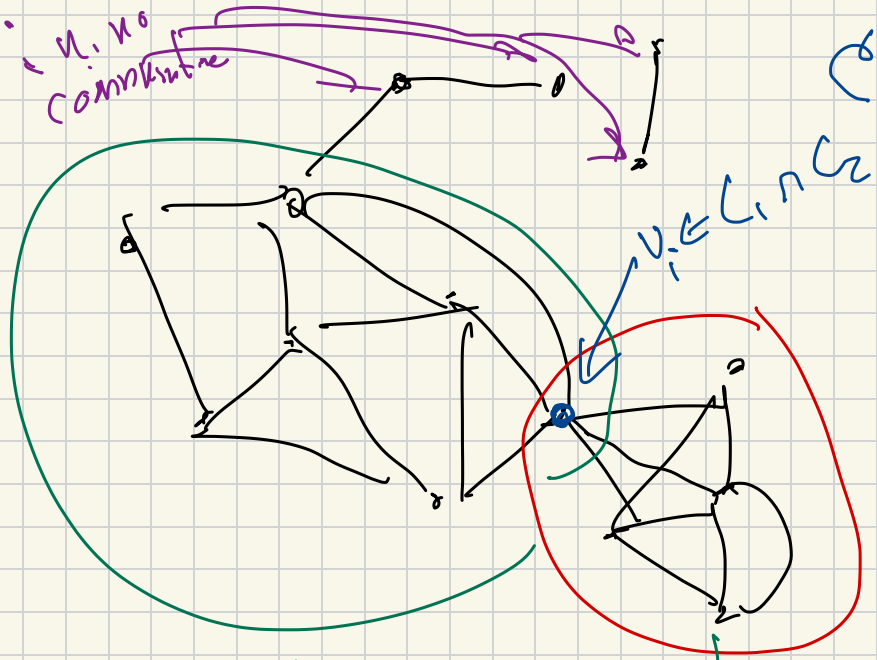
$A_{ij} = 1$ if edge $e_{ij} \in E$

$= 0$ o.w.



Communities ≠ clusters

$C \subset V$ set of communities



$C = \{C_1, C_2, \dots C_n\}$

$C_j \subset V$

$\bigcup_j C_j \neq V$

OR if $C_j \cap C_j \neq \emptyset$

→ dense part of graph.

Each company had 1 graph

$n = 100,000$
1 Million

~ 5-10 public graphs released.

what are lasting social networks.

~2000 (and before) Erdös-Renyi ($p \in (0,1)$)

each edge $e_{ij} = 1$ w.p. p independent.
0 otherwise.

real graphs ~2000

$n = 200$

hair ball graph.

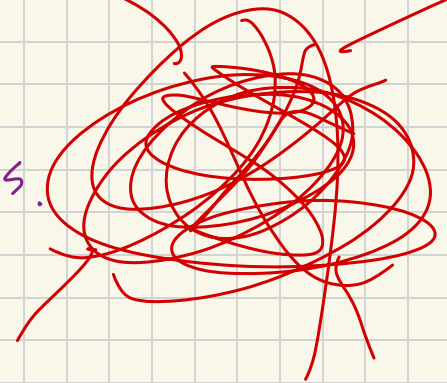
Erdös-Renyi

$deg(v_i) \sim$ uniform $\quad \mathbb{E}[deg(v_i)] = n \cdot p$

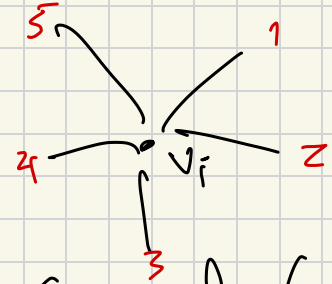
$\Rightarrow \mathbb{E}[\#tri] \approx n^3 \cdot p^3$

Real Graphs

decay degrees.



Degree distribution



degree v_i = number edges

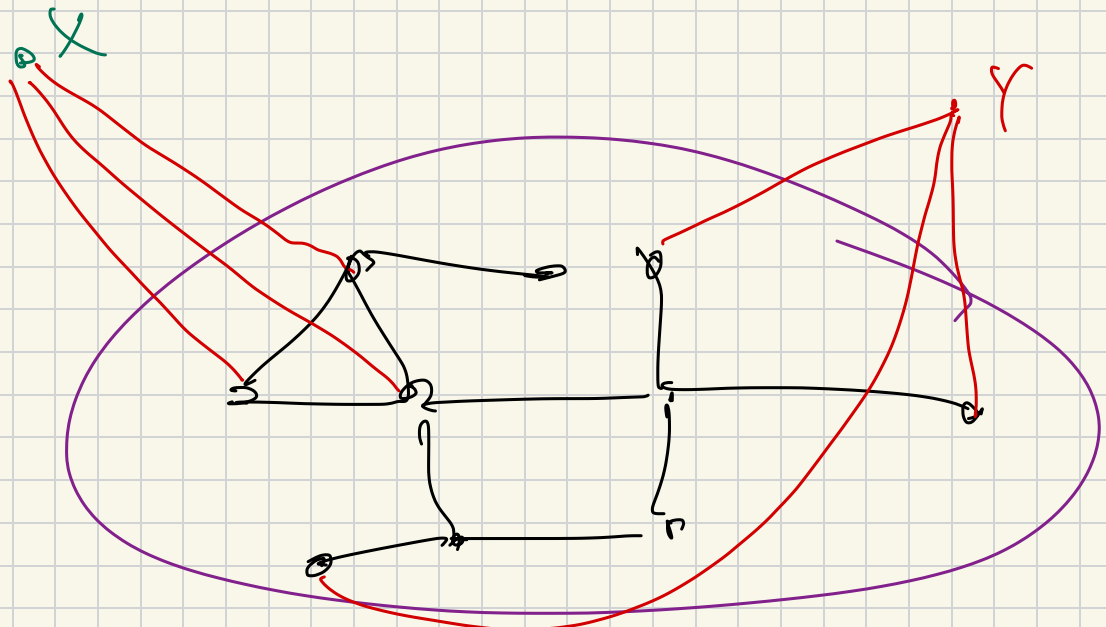
$e_{ij} \in E$

$deg(v_i) = 5$

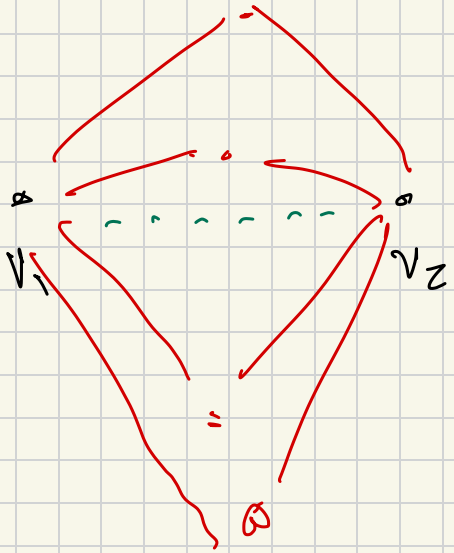
Why do people join communities ~~networks~~ ?

X : 3 friend all connected

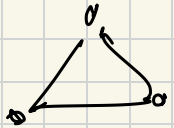
Y : 3 friend, disjoint.



Preferential Attachment



new edges like to create triangle



→ Graphs w/ lots of triangles → more stable.

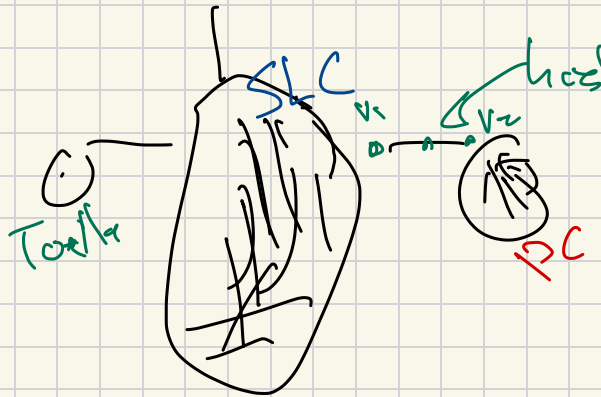
Probability of new edge v_1, v_2 proportional to number of shared friends

Communities Subsets Graphs CCV

with many edges (many triangles)

Betweenness

$btw(v_1, v_2)$ = fraction of shortest paths that use edge v_1, v_2



high btw (score)

↳ Remove edges w/ high btw.

Remaining connected components → communities

Modularity of a group $C \subset V$

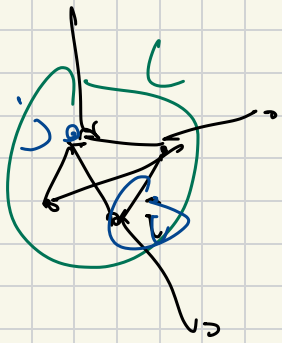
$$Q(C) = \left(\text{fraction of edges in group}\right) - \left(\text{exp. fraction of edges}\right)$$

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

A_{ij}: adjacency matrix — true # edges

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

$d_i = \text{degree}(v_i)$



communities = subgroups C w/ large $Q(C)$

Small Core Communities

Cliques = fully connected $C \subset V$

$\forall_{i,j \in C} \quad e_{ij} = 1$