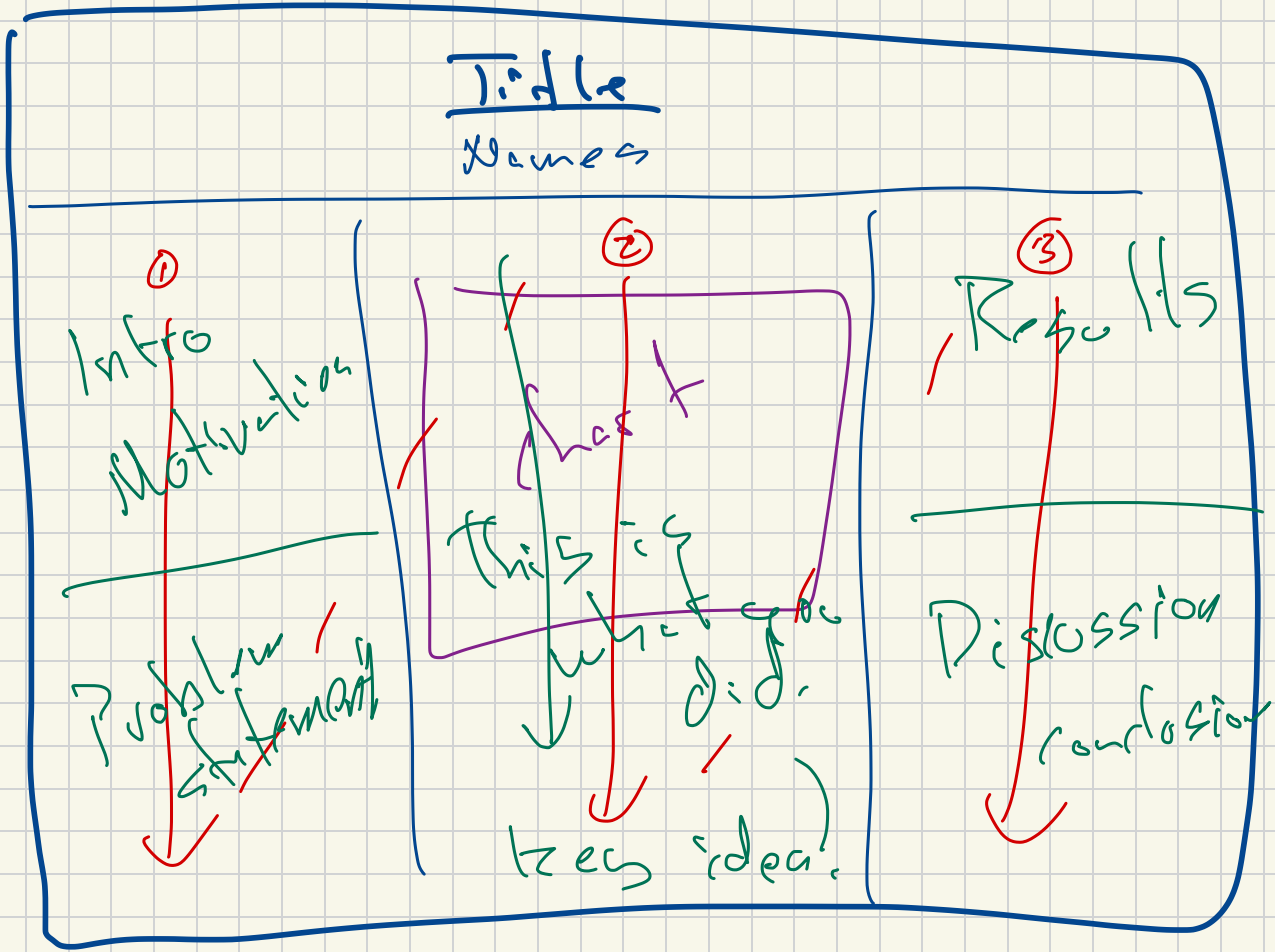
L24: Communities (in Graphs)

Apr 14, 2025
Data Mining

Jeff M. Phillips


Title
Names
1
Intro
Motivation
Problem statement
2
Method
Theory
Implementation
Key idea
3
Results
Discussion
conclusion

Communities in Graph

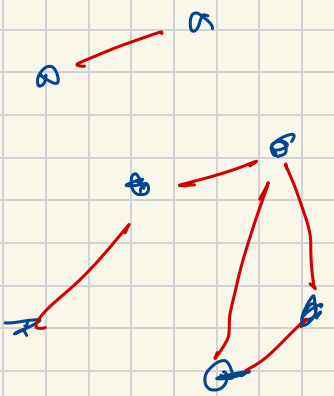
Graph $G = (V, E)$

people / entities

connections / "binary"

Modeling choice

Social Networks



Facebook

friends ≤ 7000

follows Ronaldo ~200M.

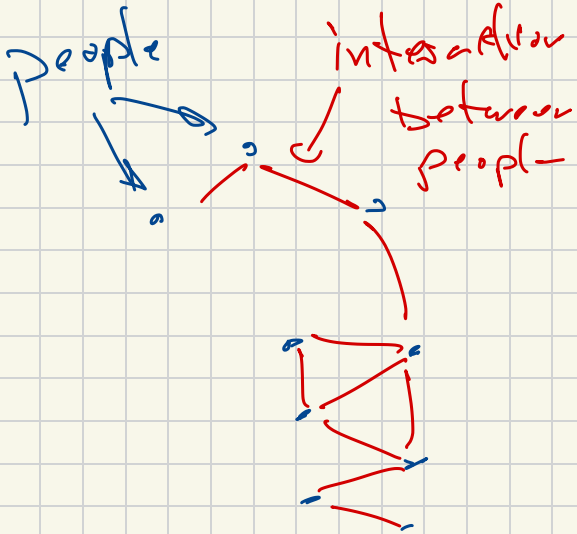
Early Social Networks

Sociology

questionnaires

manual process

$|V| \simeq [100, 200]$



planar

close to

planar

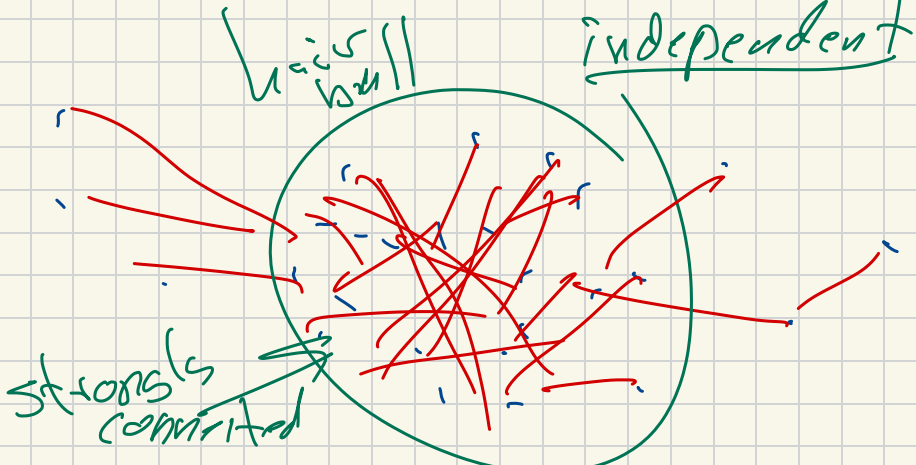
Mathematical Model of Graphs

Erdös-Renyi

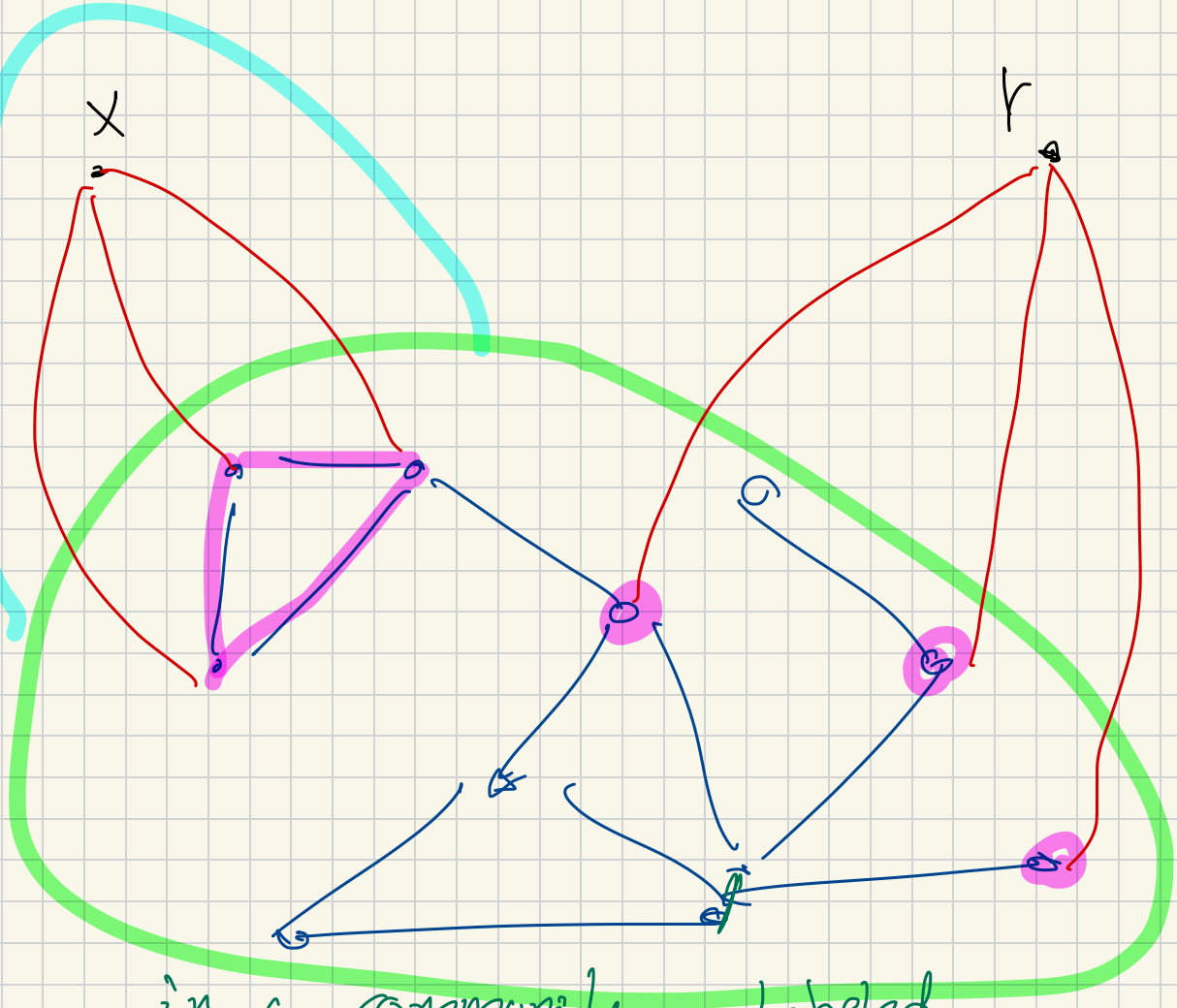
$|V| \geq 1000$

$n = |V|$

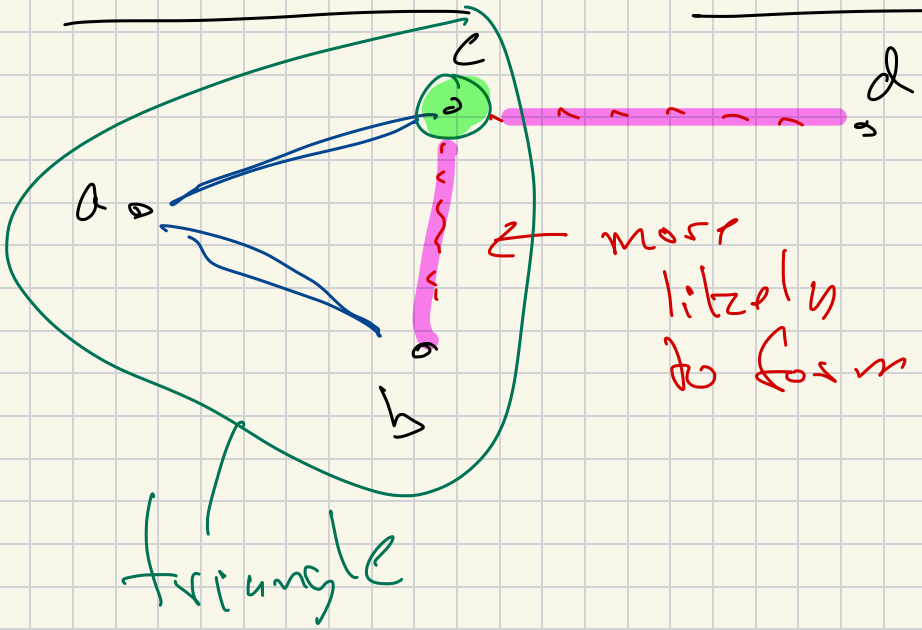
$p \in (0,1)$ = probability any pair of vertices had an edge independent



Who is more likely to join the group. X ? Y ?



Preferential Attachment



more likely to form

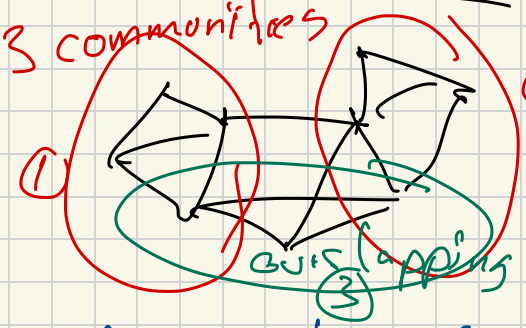
triangle

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

if it forms triangle

(a, b, c)

How do we find/define communities in graphs



- Spectral clustering

decomposes vertices into disjoint sets

Communities ① can overlap

② not all vertices must be in 1 or more

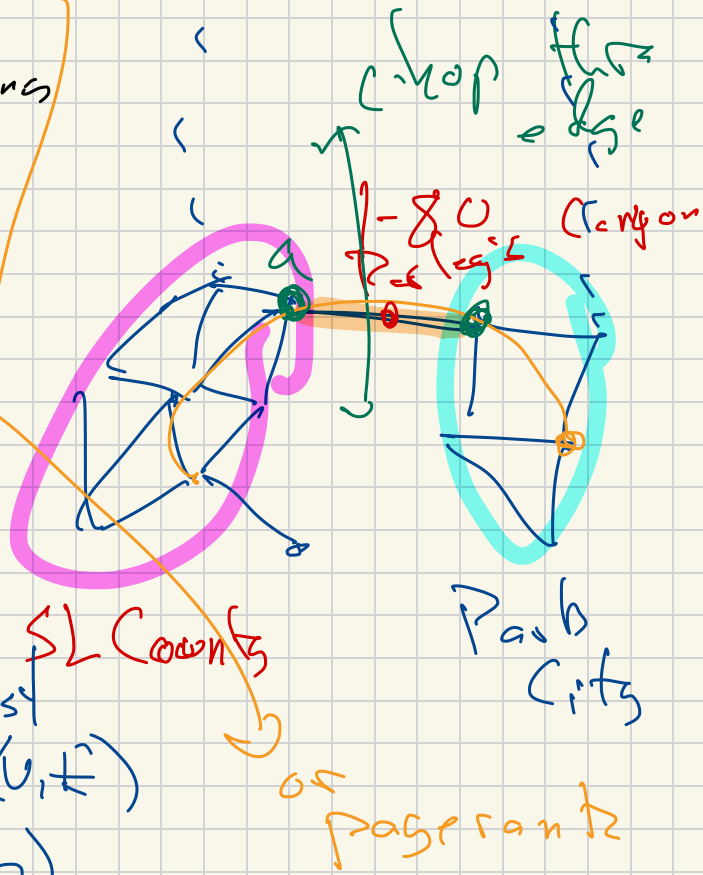
↳ tightly connected subset of vertices: two $v, v' \in C$ $P((v,v') \in E)$ high

Betweenness

(top-down view)

→ remove highly connecting edges

↳ remaining connected components are communities



betw(a,b) =

fraction of all shortest paths on graph G=(V,E) that use edge (a,b)

Clustering Coefficient

$e_i \in E$ $\quad e_i = (v, v')$

$$CC : c_i = \frac{2 \cdot (\#\text{ triangles})}{\deg(v)(\deg(v')-1)} = \frac{2\,|N(v) \cap N(v')|}{\deg(v)(\deg(v')-1)}$$

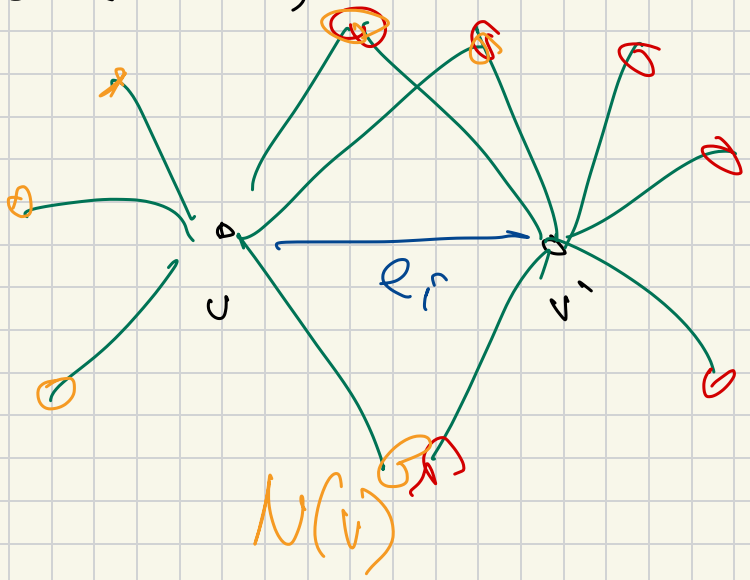
$N(v_i)$ edges incid. v_i

$N(v')$

proportional to expect # triangles

low clustering coefficient

1. good cuts
2. edges w/ low degree vertices



Modularity

score of how well-formed is a community

$Q(C)$ $C \subset V$ ← community

larger, the better.

$$Q(C) = \left(\text{fraction of edges in } (C, E_C)\right) - \left(\text{expected \# edges in } (C, E_C)\right)$$

adjacency matrix A, $A_{ij} = 1$ iff $e_{ij} \in E$

$$E_{ij} = \frac{\deg(j) \cdot \deg(i)}{2|E|}$$

frac edges C:

$$= \frac{1}{|E|} \sum_{ij \in C} A_{ij}$$

$$Q(C) = \frac{1}{E} \sum_{ij \in C} A_{ij} - \frac{\deg(i) \cdot \deg(j)}{2|E|}$$

$\in (-.5, 1)$ usually > 0.3