Local Computation Algorithms:

Maximal Independent Set

Maximal Independent Set:

not maximum

NP-Complete

<u>def</u> $U \subseteq V$ is a "Maximal Independent Set" (MIS) if

(1) $\forall u_1, u_2 \in U$, $(u_1, u_2) \notin E$ "independent"

(2) $\nexists$ $w \in V \setminus U$ s.t. $U \cup \{w\}$ is independent "maximal"

Today's assumption:

G has max degree d

<u>Note</u>: MIS can be solved via greedy (<u>not</u> NP-Complete)

Distributed Algorithm for MIS:
"Luby's Algorithm" (actually a variant)

Maximal Independent Set:

def $U \subseteq V$ is a "Maximal Independent Set" (MIS) if
(1) $\forall u_1, u_2 \in U$, $(u_1, u_2) \notin E$
(2) $\nexists\, w \in V \setminus U$ s.t. $U \cup \{w\}$ is independent

- MIS $\leftarrow \varphi$
- all nodes set to "live"
- repeat K times in parallel: (K rounds)
  - $\forall$ nodes v, color self "red" with prob $\geq \frac{1}{2d}$, else "blue". Send color to all nbrs.
  - If v colors self "red" + no other nbr of v colors self red then
    - add v to MIS
    - remove v + all nbrs from graph (set to "dead")

(for purposes of analyses, continue to select selves after die, but don't send color to nbrs)

<u>Thm</u> $\Pr[\#\text{ phases til graph empty} \geq 8d\log n] \leq \frac{1}{n}$

<u>Corr</u> $E[\#\text{ phases}]$ is $O(d \log n)$ $\Leftarrow$ can improve!

Main Lemma

$\Pr[v \text{ live} + \text{added to MIS in round}] \geq \frac{1}{4d}$

Proof

$\Pr[v \text{ colors self red}] = \frac{1}{2d}$

$\Pr[\text{any } w \in N(v) \text{ colors self red}] \leq \sum_{w \in N(v)} \frac{1}{2d}$ (union bnd)

$\leq \frac{1}{2}$ (bound on degree)

- MIS $\leftarrow \varphi$
- all nodes set to "live"
- repeat K times in parallel:
  - $\forall$ nodes v, color self "red" with prob $\geq \frac{1}{2d}$, else "blue". Send color to all nbrs.
  - If v colors self "red" + no other nbr of v colors self red then
    - add v to MIS
    - remove v + all nbrs from graph (set to "dead")

$\therefore \Pr[v \text{ colors self red} + \text{no other nbr colors self red}] \geq \frac{1}{2d}\left(1-\frac{1}{2}\right) = \frac{1}{4d}$ ∎

$\Rightarrow$ Corr $\Pr[v \text{ live after } 4Kd \text{ rounds}] \leq \left(1-\frac{1}{4d}\right)^{4Kd} = e^{-K}$

Setting K:

if $K = O(\log n)$, $\Pr[v \text{ live at end}] \leq e^{-O(\log n)} = \frac{1}{n^c}$

(can do better)

See slides for

Local Computation Algorithm (LCA)

model

Problem when sequentially simulate k-round algorithm

get d^k complexity

$k = O(\log n) \Rightarrow$ not sublinear

What to do? run fewer rounds

many nodes will not be decided yet ← is it ok?

Local Computation Algorithm to

compute Luby's answer:

"Luby status"

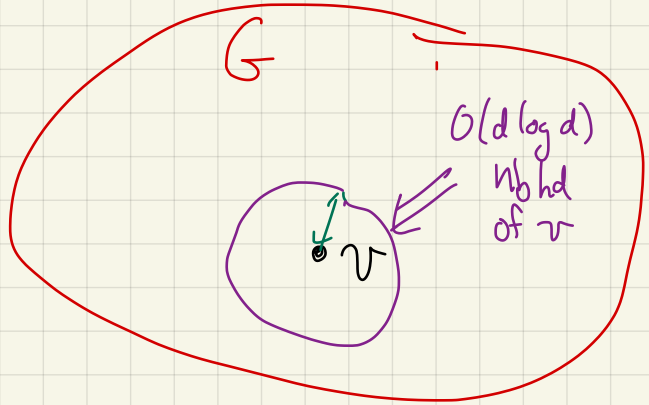
- Run Luby with $k = O(d \log d)$ rounds

at end, each node v is one of:

$\begin{cases} \text{live} \\ \text{in MIS} \quad \leftarrow \text{set self to red + no nbrs did} \\ \text{not in MIS} \quad \leftarrow \text{taken out by nbr} \end{cases}$

Luby:

- MIS $\leftarrow \varphi$
- all nodes set to "live"
- repeat k times in parallel:
  - $\forall$ nodes v, color self "red" with prob $\geq \frac{1}{2d}$, else "blue". Send color to all nbrs.
  - If v colors self "red" + no other nbr of v colors self red then
    - add v to MIS
    - remove v + all nbrs from graph (set to "dead")



- Use "Parnas - Ron" reduction:

simulate v's view of computation in sequential manner: $d^k = d^{O(d \log d)}$ queries

degree

rounds

+ determine whether v is live/in/not in

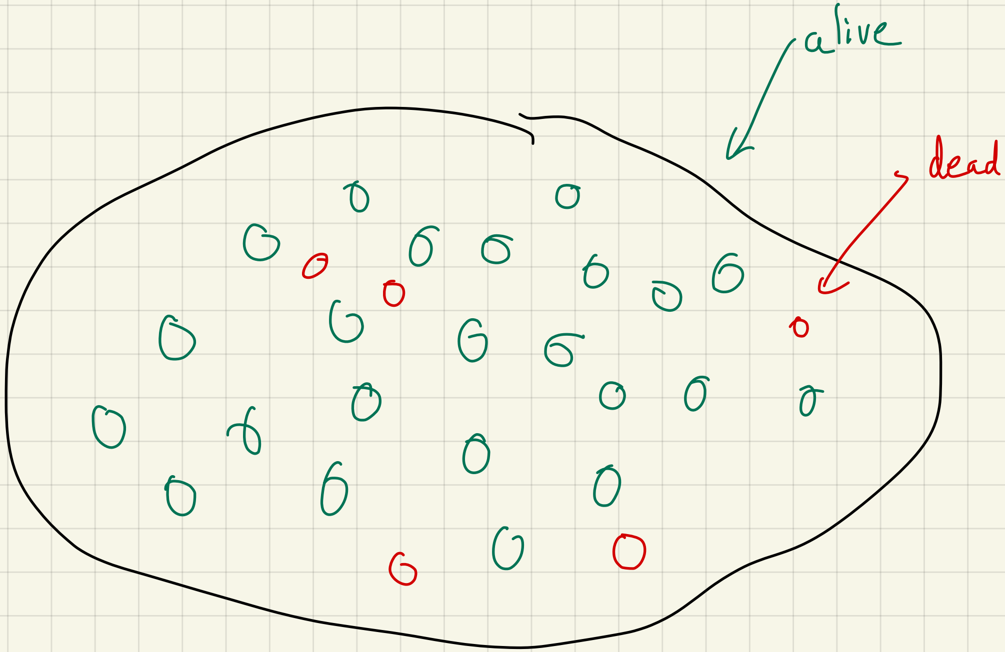
- if v is in/not in then done

else v is alive $\leftarrow$ What do we do?

Questions:

What is prob v is alive?

How are live nodes distributed after $O(d \log d)$ rounds?

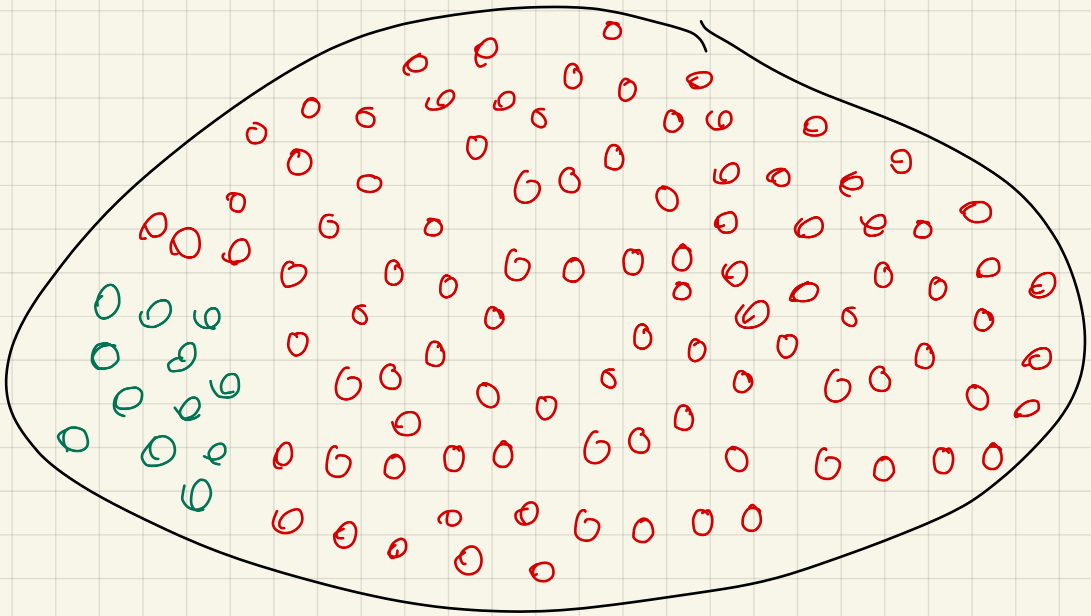


lots of live, few dead?

NO!

$$\Pr[v \text{ survives } O(d \log d) \text{ rounds}] \le e^{-O(\log d)} \le \frac{1}{d^c}$$

Most will "die" ← don't worry it won't be painful!

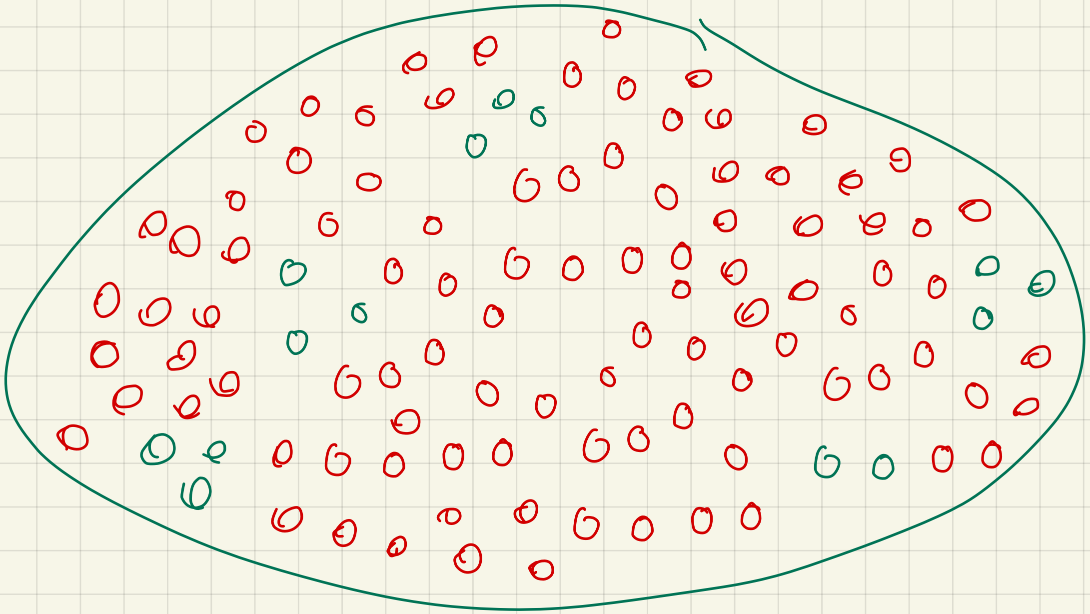


few live but clumped together?

NO!

surviving nodes will be in small connected components

Surviving nodes will be in small connected components



This relies heavily on degree bound of graph

- # conn subgraphs small
- survival of components ≈ independent

"Luby status" Luby with $k = O(d \log d)$:

given v, is it:

- live
- in MIS ← set self to red + no nbrs did
- not in MIS ← taken out by nbr

Luby:

- MIS ← φ
- all nodes set to "live"
- repeat k times in parallel:
  - $\forall$ nodes v, color self "red" with prob $\geq \frac{1}{2d}$, else "blue". Send color to all nbrs.
  - If v colors self "red" + no other nbr of v colors self red then
    - add v to MIS
    - remove v + all nbrs from graph (set to "dead")

LCA for MIS(v):

- run sequential version of Luby status(v)
- if it is in/out output answer + halt
- else, (1) do BFS to find v's connected component of live nodes
  - (2) Compute lexicographically 1st MIS M' for that connected component
  - (3) Output whether v in/out of M'

Run time

$d^{O(d \log d)}$

$d^{O(d \log d)} \times$ size of component

size of component

what is size of component?

Bounding size of connected components:

Claim After $O(d \log d)$ rounds, connected components of survivors of size $\leq O(\mathrm{poly} \log(d) \cdot \log n)$

$\Longrightarrow$ can find whole component via BFS

"brute force"

Main difficulty: survival of v & neighbors are not independent

Bounding survivors:

$$A_v = \begin{cases} 1 & \text{if } v \text{ survives all rounds} \\ 0 & \text{o.w.} \end{cases}$$

$$B_v = \begin{cases} 1 & \text{if } \nexists \text{ round s.t. } v \text{ colors self \& no } w \in N(v) \text{ colors self} \\ 0 & \text{o.w.} \end{cases}$$

might not mean that $v \in MIS$
e.g. if v died due to nbr being put in MIS

Luby:
- $MIS \leftarrow \varphi$
- all nodes set to "live"
- repeat K times in parallel:
  - $\forall$ nodes v, color self "red" with prob $\geq \frac{1}{2d}$, else "blue". Send color to all nbrs.
  - If v colors self "red" & no other nbr of v colors self red then
    - add v to MIS
    - remove v & all nbrs from graph (set to "dead")

Note: $A_v = 1 \Rightarrow B_v = 1$

e.g. v survives $\Rightarrow \nexists$ round s.t. v colors self & no $w \in N(v)$ colors self

these events are independent

$$\Pr[B_v = 1] \leq \underbrace{\left(1 - \frac{1}{4d}\right)}_{\text{prob survive one round}}^{c \cdot d \log d} \leq \frac{1}{8d^3} \quad \text{for} \quad c \geq 20$$

Bounding size of connected components:

$$A_v = \begin{cases} 1 & \text{if } v \text{ survives all rounds} \\ 0 & \text{o.w.} \end{cases}$$

$$B_v = \begin{cases} 1 & \text{if } \nexists \text{ round s.t. } v \text{ colors self \& no } w \in N(v) \text{ colors self} \\ 0 & \text{o.w.} \end{cases}$$

might not mean that $v \in MIS$
e.g. if v died due to nbr being put in MIS

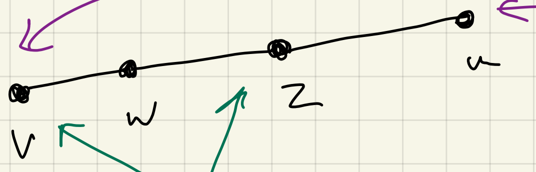
Luby:

- $MIS \leftarrow \varphi$
- all nodes set to "live"
- repeat K times in parallel:
  - $\forall$ nodes v, color self "red" with prob $\geq \frac{1}{2d}$, else "blue". Send color to all nbrs.
  - If v colors self "red" & no other nbr of v colors self red then
    - add v to MIS
    - remove v & all nbrs from graph (set to "dead")

Note: $A_v = 1 \Rightarrow B_v = 1$

eg. v survives $\Rightarrow \nexists$ round s.t. v colors self & no $w \in N(v)$ colors self

we care about A_v's, but B_v's have nice independence properties



distance 3:

B_v depends on B_w

B_u depends on B_z

but B_u & B_v indep!

distance 2:

B_v & B_z depend on w's colors so not independent

$\deg \leq d \Rightarrow$ each B_u depends on $\leq d^2$ other B_w's

Bounding size of connected components:

<u>Claim</u> After $O(d \log d)$ rounds, connected components of survivors of size $\leq O(\mathrm{poly}\log(d) \cdot \log n)$

$\Longrightarrow$ can find whole component via BFS
"brute force"

<u>Proof idea</u>:

- any large conn component has lots of nodes that are independent (distance ≥ 3)
- these indep nodes unlikely to simultaneously survive

do we need to union bnd over <u>all</u> sets of size w?
<u>NO</u>

Claim After $O(d \log d)$ rounds, connected components of survivors of size $\leq O(\text{poly}\log(d) \cdot \log n)$

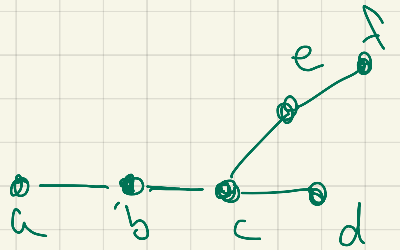
Proof

Let $G^{(3)} \leftarrow$ graph s.t. nodes $\sim B_v$

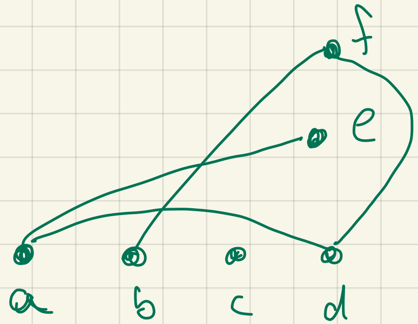
edges $\sim B_v$ & B_w $\underbrace{\text{distance}}_{}\geq 3$ in G

hope: represent independent events!!
(but not true yet)

$\deg(G^{(3)}) \leq d^3$



$G^{(3)}$



H

not even connected!

Great! we are good at counting trees!! $\Longrightarrow$

Observe: # components in $G^{(3)}$ of size w

$\leq$

size w subtrees in $G^{(3)}$

why? map each component C to arbitrary spanning tree of C

mapping is 1-1
but could have many spanning trees per component

Claim for (live) connected component S in G

$G^{(3)}$ contains tree with vertex set T

as subgraph s.t. (1) $|T| \geq \frac{|S|}{d^2+1}$

(2) $\text{dist}_G(u,v) \geq 3 \quad \forall\ u,v \in T$

Proof Pick T greedily

1. pick arbitrary $v \in S$
2. repeat until S empty
3. move v from S to T, remove all u with $\text{dist}_G(u,v) < 3$ from S
4. pick new node $v \in S$ s.t. $\text{dist}_G(u,v) = 3$ for some $u \in T$

Note: for each v put in T, remove $\leq d^2$ nodes from S

$+$ remove v

total $\leq d^2 + 1$

How big are remaining components?

Let $s \leftarrow \log \frac{n}{3}$

Let $\Upsilon_s = \{ T \subseteq V \mid |T| = s,$ all $u, v \in T$ have dist ≥ 3 in G & T is connected in $G^{(3)} \}$

$\Pr[\exists\, T \in \Upsilon_s$ s.t. all nodes in T survive$]$

see calculation on next slide

$$\leq \sum_{T \in \Upsilon_s} \Pr[\text{all nodes in } T \text{ survive}] \leq |\Upsilon_s| \cdot \left(\frac{1}{8d^3}\right)^s \leq N \cdot (4d^3)^s \cdot \frac{1}{(8d^3)^s} \leq \frac{N}{2^s} = \frac{1}{3}$$

($\Rightarrow$ unlikely any size s tree survives)

$\Rightarrow$ with prob $\geq 2/3$ all surviving conn. comp have size $\leq (d^2+1) \cdot \log n/3 = O(d^2 \log n)$

How many subtrees in a degree bounded graph?

<u>Known Thm</u> # non isomorphic trees on w nodes $\leq 4^w$ ← ignores names of nodes & root (just shape)

<u>Corr</u> # size w subtrees in N-node graph of degree $\leq D$ ← considers names of nodes & root

is $\leq N \cdot 4^w \cdot D^w = N(4D)^w$

why?

- Choose Root in H
- Choose size w tree (shape) from known thm
- choose placement in H via DFS sequence

<u>choices</u>

N

4^w

D choices for 1st child

" " " 2nd "

⋮

total # choices: $N \cdot 4^w \cdot D^w$