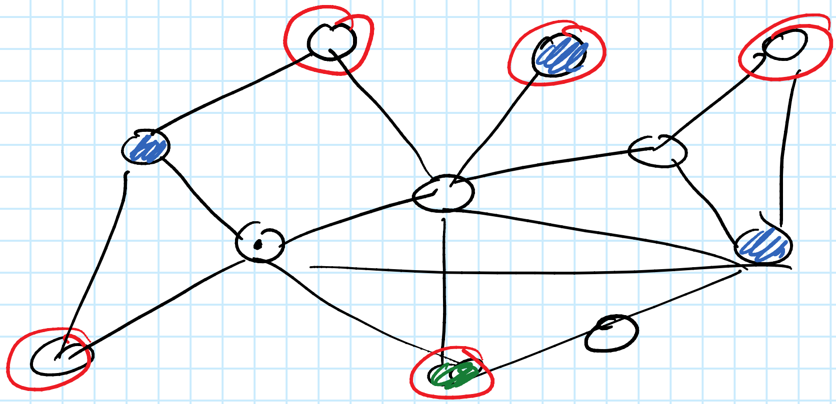
Maximal Independent Sets

Thursday, July 3, 2014
9:01 AM

- classical distr. graph alg.
- first randomized alg.



independent set : set of nodes where no two are connected by an edge

maximal ind. set (MIS) : cannot be extended

maximum ind. set : ind. set of max. cardinality

NP-hard to find $n^{1/2-\varepsilon}$-approximation

maximal I.S. vs. maximum I.S.

M.I.S.

maximum IS

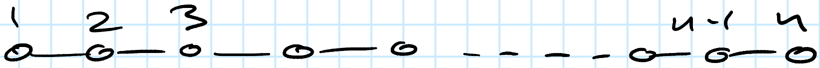
Goal: distributed algorithm to compute MIS

Distr. MIS Alg.

Nodes have IDs

Rule: v has smallest ID among undecided neighbors,
v joins MIS
↳ neighbors of v don't join MIS

Time complexity: $O(n)$



If done carefully, $O(m)$ messages.

If we are given a $\underbrace{C\text{-coloring}}$ of G
each node has a color $\in \{1, \dots, C\}$

⟹ do the same thing using colors (instead of IDs)

⟹ time complexity : $O(C)$

In rooted trees, ring, bounded degr. graphs
↳ $O(1)$-coloring in $O(\log^* n)$ rounds

⟹ MIS in $O(\log^* n)$ rounds

General graphs:

Idea: use randomization!

randomly pick some nodes
→ check whether they're adjacent

→ what probability to use?

→ tie-breaking

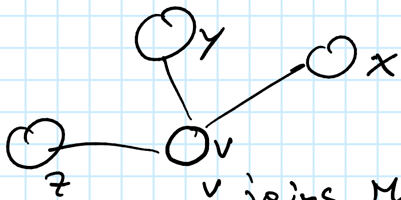
→ use random IDs
↳ computes an MIS in $O(\log^2 n)$ rounds

We use random IDs, but
pick new random IDs in every iteration/phase

Phase:

1) Every undec. node v picks random
$r(v) \in [0, 1]$
v sends $r(v)$ to all neighbors
2) If $r(v) < r(w)$ for all undec. neighbors w,
v joins MIS, inform neighbors
3) If some neighb. joined the MIS, v becomes dec.

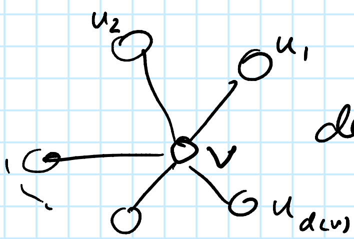
Correct!



v joins MIS $\Rightarrow r(v) < r(x), r(y), r(z)$
$\Rightarrow x, y, z$ do not join MIS

Not clear: running time

Probability that node v joins MIS in one phase
$\Rightarrow$ at beginning of phase: $\underbrace{G = (V, E)}_{\text{graph of undec. nodes}}$

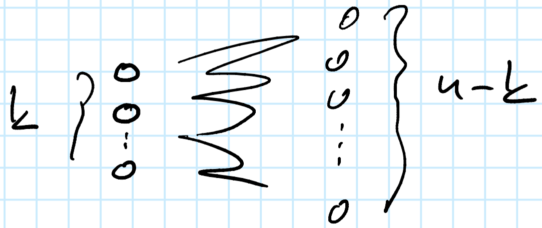


degree of v: $d(v)$

$\{v, u_1, \ldots, u_{d(v)}\}$ $\quad \Pr(v \text{ joins MIS}) = \dfrac{1}{d(v)+1}$

want to show: $O(\log n)$ phases

A const. fraction of the nodes get decided?

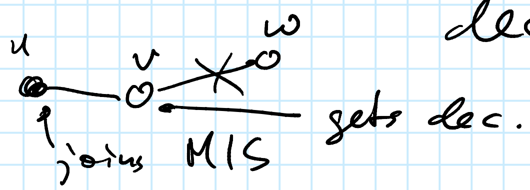


roughly $\frac{n}{k}$ nodes that join MIS

$\Longrightarrow$ # nodes that get decided $\Theta\left(\frac{n}{k}\right)$

A const. fraction of the edges gets elim. in each phase.

edge eliminated: if one of its nodes is decided



Random variables $X_1, X_2, \ldots, X_n$

$$X = X_1 + X_2 + \ldots + X_n$$

$$\Longrightarrow \underbrace{\mathbb{E}[X] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \ldots + \mathbb{E}[X_n]}_{\text{linearity of expectation}}$$

$$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

$$\mathbb{E}[X+Y] = \sum_{(X,Y)=(x,y)} \underbrace{\Pr((X,Y)=(x,y))} \cdot (x+y)$$

$$= \Pr(X=x) \cdot \Pr(Y=y \mid X=x)$$
$$= \Pr(Y=y) \cdot \Pr(X=x \mid Y=y)$$

$$\underline{E[X+Y]} = \sum_{(X,Y)=(x,y)} Pr((X,Y)=(x,y))\cdot(x+y)$$

$$= \sum_x \sum_y Pr(X=x)\cdot Pr(Y=y \mid X=x)\cdot x$$

$$+ \sum_y \sum_x Pr(Y=y)\cdot Pr(X=x \mid Y=y)\cdot y$$

$$= \underbrace{\sum_x Pr(X=x)\cdot x}_{E[X]} \cdot \underbrace{\sum_y Pr(Y=y \mid X=x)}_{=1}$$

$$+ \underbrace{\sum_y P(Y=y)\cdot y}_{E[Y]} \cdot \underbrace{\sum_x Pr(X=x \mid Y=y)}_{=1}$$

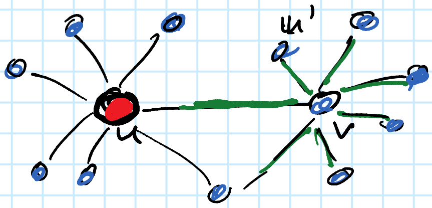
$$= \underline{E[X] + E[Y]}$$

R : # edges removed (elim.)

For every edge $\{u,v\}$ introduce two events

$\mathcal{E}_{u\to v}$ and $\mathcal{E}_{v\to u}$

$\underline{\mathcal{E}_{u\to v}}$

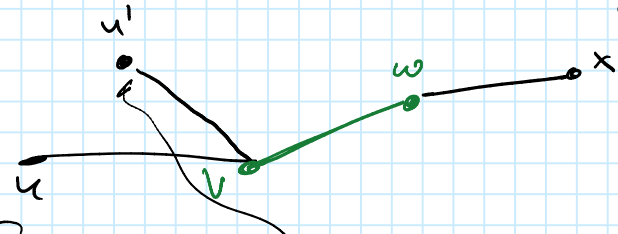


neighbors of u : $N(u)$

u joins MIS : $\forall w \in N(u) : r(u) < r(w)$

$\mathcal{E}_{u\to v}$: $\underline{\forall w \in N(u) \cup N(v) \setminus \{u\}} : r(u) < r(w)$

$$Pr(\mathcal{E}_{u\to v}) \geq \frac{1}{d(u)+d(v)}$$



$\mathcal{E}_{x\to w}$ possible only for one x

$\underline{\mathcal{E}_{u\to v}} \longrightarrow \underline{\mathcal{E}_{u'\to v}}$ cannot happen

Random var.

$$X_{u\to v} = \begin{cases} d(v) & \text{if } \mathcal{E}_{u\to v} \text{ happens} \\ 0 & \text{otherwise} \end{cases}$$

$$X = \sum_{\{u,v\}\in E} X_{u\to v} + X_{v\to u}$$

Claim: $R \geq X/2$

Proof: count every edge at most twice

$$\mathbb{E}[R] \geq \frac{\mathbb{E}[X]}{2}$$

$$\mathbb{E}[X] = \sum_{\{u,v\}\in E} \left(\mathbb{E}[X_{u\to v}] + \mathbb{E}[X_{v\to u}] \right)$$

lin. of exp.

$$= \sum_{\{u,v\}\in E} \left(\Pr(\mathcal{E}_{u\to v}) \cdot d(v) + \Pr(\mathcal{E}_{v\to u}) \cdot d(u) \right)$$

$$\geq \sum_{\{u,v\}\in E} \underbrace{\frac{d(v)}{d(u)+d(v)} + \frac{d(u)}{d(u)+d(v)}}_{=1} = |E|$$

$$\Longrightarrow \mathbb{E}[R] \geq \frac{|E|}{2}$$

Lemma: With prob. $\geq \frac{1}{3}$,
at least $1/4$ of the edge sets eliminated

Proof!

To show: $\underline{\Pr\left(R \geq \frac{|E|}{4}\right) \geq \frac{1}{3}}$ assume this not done.

$\downarrow p < \frac{1}{3}$

$$\underline{E[R]} = \sum_{r=0}^{|E|} \Pr(R=r) \cdot r$$

$$= \sum_{r \geq \frac{|E|}{4}} \Pr(R=r) \cdot r + \sum_{r < \frac{|E|}{4}} \Pr(R=r) \cdot r$$

$$< |E| \cdot \underbrace{\Pr\left(R \geq \frac{|E|}{4}\right)}_{p} + \frac{|E|}{4} \cdot \underbrace{\Pr\left(R < \frac{|E|}{4}\right)}_{1-p}$$

$$= |E| \left(\frac{1}{4} + \frac{3}{4} p\right)$$

$$< |E| \left(\frac{1}{2}\right) = \frac{|E|}{2}$$

---

phase is good if $\geq 1/4$ of edges are eliminated

$\Pr(\text{phase good}) \geq \frac{1}{3}$

$\Rightarrow$ for every phase, indep. of what happened before

need $\underbrace{\log_{4/3}(n^2)}_{\Theta(\log n)}$ good phases

T_i : # of phases until the first good phase after the $(i-1)^{st}$ good phase

$$| \underbrace{B\,B\,G}_{T_1} \; \underbrace{G}_{T_2} \; \underbrace{B\,B\,B\,G}_{T_3} \; \underbrace{B\,G}_{T_4} \; \ldots$$

#phases:

$$T = T_1 + \dots + T_{O(\log n)}$$

running time

$$\mathbb{E}[T] = \sum_i \mathbb{E}[T_i] = \underline{O(\log n)}$$

$$\mathbb{E}[T_i] \leq 3$$

Alg. terminates in $O(\log n)$ phases (rounds) in expectation.

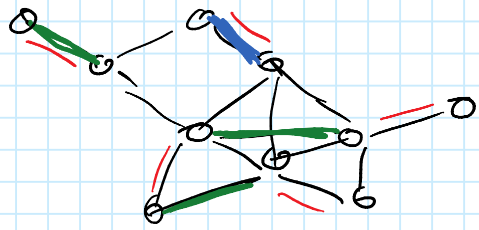
One can also show that for some const. $\alpha > 0$

$$Pr(T > \underline{c \cdot \alpha \cdot \log n}) < \frac{1}{n^c}$$

Alg. is succ. within $c\alpha\log n$ phases with prob. $\geq 1 - \frac{1}{n^c}$ (with high probability)

c can be picked arbitrarily (affects const. in $O(\cdot)$)

Maximal Matching



matching

maximal matching

matching that cannot be extended

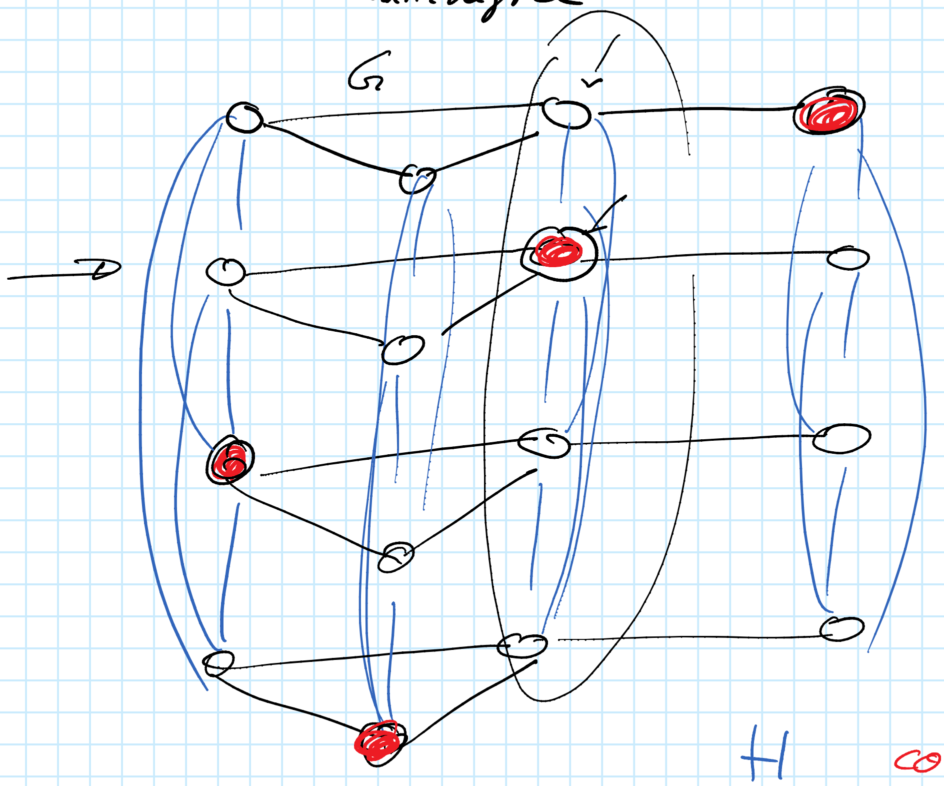
Can use MIS alg. to get a maximal matching in $O(\log n)$ time

Coloring

$(\Delta+1)$-coloring

max. degree

$\Delta = 3$



$\Delta+1$ copies

H

compute MIS

at most one copy of each node of G

at least one copy of each node of G

$\Longrightarrow$ exactly one copy of each node of G

gives the color

H can be simulated by
a distr. alg. on G

Nodes need to know Δ ?

sufficient to use $\underline{\underline{d(v)+1}}$ copies for node v

$\Longrightarrow$ $(\Delta+1)$-coloring in $O(\log n)$ rounds

---

MIS / $(\Delta+1)$-coloring

can be solved in $O(\log n)$ time
by using randomization

What about deterministic algorithms?

<u>best known</u> : $2^{\Theta(\sqrt{\log n})} = n^{\frac{1}{\Theta(\sqrt{\log n})}}$

$\text{polylog} = 2^{\Theta(\log\log n)}$

<u>MIS lower bound</u>

$\Omega(\sqrt{\log n})$