Today: Use pairwise indp't. random variables to derandomize an algorithm.

Idea: Present a randomized algorithm that works with constant probability & only uses pairwise indp't. random variables. Then, can iterate through all possible choices of to find a deterministic choice that is guaranteed to succeed.

Maximal independent set:

IS of a graph $G=(V,E)$ is a subset $S \subset V$ where for all $v,w \in S$, $(v,w) \notin E$ (so no edge is contained in S)

independent set

IS S is maximal if:

for all $v \in V$, either $v \in S$ or $N(v) \cap S \neq \emptyset$ (so can't add v to S)

Call it MIS = maximal independent set.

<u>Now:</u> Parallel algorithm for MIS due to [Luby '85].
$O(\log m)$ rounds & $\mathrm{poly}(n)$ processors
(CREW PRAM model
= concurrent read, exclusive write)

<u>Idea:</u> Have a current IS I.
Let $G' = G \setminus (I \cup N(I))$ be the remaining graph.

Any $v \in G'$ can be added to I.

Here's a simple MIS alg. (sequential alg.):

1. $I = \emptyset$, $V' = V$.
2. While $(V' \neq \emptyset)$ do:
   a) Choose any $v \in V'$.
   b) Set $I = I \cup \{v\}$.
   c) Set $V' = V' \setminus (\{v\} \cup N(v))$
3. Output I.

This may take $O(n)$ time/rounds.

Goal: find IS S of G' &
add S to I.

If $S \cup N(S)$ is constant fraction of G'
then $O(\log n)$ rounds needed.

How to find S?

Every vertex $v \in G'$ adds themselves to S
with prob. $p(v)$, independently
(or pairwise indp't.)

To make sure that S is an IS:

For all edges $(v,w) \in E$,
if v & w are in S
then remove ~~both~~.
lower deg. vertex
$d(v) \leq d(w)$.

Luby's MIS alg.: Given input $G=(V,E)$,

1. Set $I=\emptyset$, $V'=V$, $G'=G$.
2. While $(V' \neq \emptyset)$ do:
   a) Set $S=\emptyset$
   b) For each $v \in V'$,
      add v to S with prob. $\frac{1}{2d_{G'}(v)}$
      where $d_{G'}(v)=$ degree of v in G'.
   c) For every edge $(y,z) \in E(G')$,
      if $y \in S$ & $z \in S$,
      then remove lower degree in G',
      i.e., remove y where $d_{G'}(y) \le d_{G'}(z)$
      if $d_{G'}(y)=d_{G'}(z)$ choose arbitrarily.
      Call this new set S'.
   d) Set $I=I \cup S'$.
      Set $V'=V' \setminus (S' \cup N_{G'}(S'))$.
      Let G' be induced subgraph on V'.
3. Output I.

Analysis: Let $G_j = (V_j, E_j)$ be the graph G' after stage j.

Thus, $G_0 = G$.

Lemma: For some $c < 1$,

$$E\left[\,|E_j| \;\middle|\; E_{j-1}\right] < c\,|E_{j-1}|$$

Therefore, $O(\log m)$ rounds will be needed in expectation where $m = |E|$.

Proof of lemma:

For graph $G_j = (V_j, E_j)$ partition edges into GOOD & BAD.

First, vertex $v \in V_j$ is BAD if

$$\left|\{w \in N_{G_j}(v) : d_{G_j}(w) > d_{G_j}(v)\}\right| > \frac{2}{3} d_{G_j}(v)$$

more than $\frac{2}{3}$ of v's neighbors have higher degree

& v is GOOD if not BAD.

Then, edge $e=(v,w)\in E_j$ is BAD
if v & w are both BAD
& otherwise e is GOOD.

Claim 1: $\geq \frac{1}{2}$ the edges in G_j are GOOD.

And good edges have a good chance to get added to S' since few neighbors are higher degree.

Claim 2: If e is GOOD, then e is removed from G' with prob. $\geq \alpha := \frac{1}{2}\left(1-e^{-1/6}\right) \simeq .07676\ldots$

From these 2 claims we get the main lemma:

$$E\left[\,|E_{j+1}| \mid E_j\right] = \sum_{e\in E_j} \left(1-Pr(e \text{ gets deleted})\right)$$

$$\leq |E_j| - \alpha\,|\text{GOOD edges}|$$

$$\leq |E_j|\left(1-\alpha/2\right)$$

which proves the lemma with $c = 1-\frac{\alpha}{2}$.

From the lemma we have:

$$E[|E_j|] \le |E_0|\left(1-\frac{\alpha}{2}\right)^j$$
$$\le m e^{-j\alpha/2}$$
$$< 1 \quad \text{for } j > \frac{2}{\alpha}\log m$$

Moreover,
$$\Pr(E_j \ne \emptyset) \le \Pr(E_j \ge 1)$$
$$\le E[|E_j|]$$
$$\le \frac{1}{4} \quad \text{for } j > \frac{4}{\alpha}\log m.$$

Thus with prob. $\ge \frac{3}{4}$, we have $\le 60 \log m$ rounds.
(This is an RNC algorithm for MIS.)

So that will complete the analysis of the randomized algorithm once we prove the 2 claims.

Now let's prove the 2 claims.

Proof of claim 1:

— Let E_B = BAD edges of G_j.

— We'll define $f : E_B \to \binom{E_j}{2}$ so that:

for all $e_1 \neq e_2 \in E_B$, $f(e_1) \cap f(e_2) = \emptyset$

Thus, each $e \in E_B$ has a distinct pair of edges in E_j

& hence: $|E_B| \leq |E_j|/2$,

which proves the claim.

— Here's the function f:

for each $(v,w) \in E_j$, direct it from the lower degree endpoint to the higher degree one (choose arbitrarily if same degrees)

Suppose for $(v,w) \in E_B$ its directed $v \to w$

so $d_{G'}(v) \leq d_{G'}(w)$.

Since $(v,w) \in E_B$ so it's BAD

then v & w are BAD, by def'n.

Since v is BAD,

$\geq \frac{2}{3}$ of v's neighbors have $>$ degree.

So these edges point away from v.

& $\leq \frac{1}{3}$ of the edges incident v point to v.

So ≥ 2 times as many out-edges from v as in-edges to v.

Hence, for each BAD edge e directed into v, there are a pair of out edges out of v that we can uniquely assign to e.

That's the mapping: for each BAD edge, look at its orientation, take the incoming endpoint & there are a unique pair of out edges. (those out edges are incoming to the other endpoint so are not assigned elsewhere).

∎

Now for claim 2:

We'll prove 2 things:

a) if v is GOOD then it's likely to have a neighbor in S.

b) if $w \in S$ then with prob. $\geq \frac{1}{2}$ $w \in S'$.

then the claim follows.

Claim a: If v is GOOD, then $\Pr(N_{G'}(v) \cap S \neq \emptyset) \geq 2\alpha$

where $\alpha := \frac{1}{2}(1 - e^{-1/6})$.

Proof: Let $L(v) := \{w \in N_{G'}(v) : d_{G'}(w) \leq d_{G'}(v)\}$

= neighbors of v with lower degree.

Note, for v GOOD, then $|L(v)| \geq d_{G'}(v)/3$.

$$\Pr(N_{G'}(v) \cap S \neq \emptyset) = 1 - \Pr(N_{G'}(v) \cap S = \emptyset)$$

$$= 1 - \prod_{w \in N_{G'}(v)} \Pr(w \notin S)$$

*Uses full independence

$$\geq 1 - \prod_{w \in L(v)} \Pr(w \notin S)$$

$$= 1 - \prod_{w \in L(v)} \left(1 - \frac{1}{2d_{G'}(w)}\right) \quad \left(\text{by def'n. of } P(w)\right)$$

$$\geq 1 - \prod_{w \in L(v)} \left(1 - \frac{1}{2d_{G'}(v)}\right) \quad \text{Since } d_{G'}(w) \leq d_{G'}(v).$$

$$\geq 1 - e^{-|L(v)|/2d_{G'}(v)}$$

$$\geq 1 - e^{-1/6} \quad \text{Since } |L(v)| \geq \frac{d_{G'}(v)}{3}.$$

∎

Now let's prove:

<u>Claim b:</u> $\Pr(w \notin S' \mid w \in S) \leq 1/2$.

<u>Proof:</u> Let $H(w) = N_{G'}(w) \setminus L(w) = \{z \in N_{G'}(w) : d_{G'}(z) > d_{G'}(w)\}$

$$\Pr(w \notin S' \mid w \in S) = \Pr(H(w) \cap S \neq \emptyset \mid w \in S)$$

Since throw out lower degree endpoint

$$\leq \sum_{z \in H(w)} \Pr(z \in S \mid w \in S)$$ by union bound

$$= \sum_{z \in H(w)} \frac{\Pr(z \in S, w \in S)}{\Pr(w \in S)}$$

$$= \sum_{z \in H(w)} \frac{\Pr(z \in S)\Pr(w \in S)}{\Pr(w \in S)}$$ Using Pairwise independence

$$= \sum_{z \in H(w)} \Pr(z \in S)$$

$$= \sum_{z \in H(w)} \frac{1}{2d_{G'}(z)}$$

$$\leq \sum_{z \in H(w)} \frac{1}{2d_{G'}(w)}$$

$$\leq \frac{1}{2}$$ ∎

Now from these claims a & b:

$$Pr\left(v \in N_{G'}(S') \mid v \text{ is GOOD}\right)$$

$$= Pr\left(N_{G'}(v) \cap S' \neq \emptyset \mid v \text{ is GOOD}\right)$$

$$= Pr\left(N_{G'}(v) \cap S' \neq \emptyset \mid N(v) \cap S \neq \emptyset, v \text{ GOOD}\right) \cdot Pr\left(N(v) \cap S \neq \emptyset \mid v \text{ GOOD}\right)$$

$$\geq \left(\frac{1}{2}\right)(2\alpha) \text{ by these 2 claims.}$$

$$= \alpha.$$

Therefore, if v is GOOD then
v is deleted with prob. $\geq \alpha$
(since if $N(S')$ are deleted)

That proves Claim 2. since ≥ 1 endpoint is GOOD for a GOOD edge.

& That finishes the analysis of the randomized algorithm.

We used independence for the following:

$$\Pr(N_{G'}(v) \cap S \neq \emptyset) = 1 - \Pr(N_{G'}(v) \cap S = \emptyset)$$

$$= 1 - \prod_{w \in N_{G'}(v)} \Pr(w \notin S)$$

We need a lower bound on this

We'll prove:

<u>Lemma</u>: For $X_i \in \{0,1\}$ where $p_i = \Pr(X_i = 1)$ & X_i's are pairwise indpt.,

$$\Pr\left(\sum_{i=1}^{\ell} X_i > 0\right) \geq \frac{1}{2} \min\left\{\frac{1}{2}, \sum_{i=1}^{\ell} p_i\right\}$$

Letting: $X_i = \begin{cases} 1 & \text{if } w_i \in S \\ 0 & \text{o/w} \end{cases}$ we have:

$$\Pr(N_{G'}(v) \cap S \neq \emptyset) \geq \frac{1}{2} \min\left\{\frac{1}{2}, \sum_{i=1}^{d_{G'}(v)} \frac{1}{2d_{G'}(w_i)}\right\}$$

$$\geq \frac{1}{2} \min\left\{\frac{1}{2}, \sum_{i=1}^{L(v)} \frac{1}{2d_{G'}(v)}\right\} \geq \frac{1}{12}.$$

Proof of Lemma:

Case 1: $\sum_i p_i \leq 1$.

$$\Pr\Big(\sum_i X_i > 0\Big) \geq \Pr\Big(\sum_i X_i = 1\Big)$$

$$\geq \sum_i \Pr(X_i = 1) - \frac{1}{2}\sum_{i \neq j} \Pr(X_i = 1, X_j = 1)$$

$$= \sum_i p_i - \frac{1}{2}\sum_i p_i \sum_{j \neq i} p_j$$

because they are pairwise indep't.

$$\geq \sum_i p_i - \frac{1}{2}\Big(\sum_i p_i\Big)^2$$

$$= \sum_i p_i \Big(1 - \frac{1}{2}\sum_i p_i\Big)$$

$$\geq \frac{1}{2}\sum_i p_i \quad \text{when } \sum_i p_i \leq 1$$

Case 2: $\sum_i p_i > 1$:

Find $S \subset \{1, \dots, \ell\}$ where $\frac{1}{2} \leq \sum_{i \in S} p_i \leq 1$.

Always exists such a S b/c: either all i, $p_i < \frac{1}{2}$ or $\exists j\ \frac{1}{2} \leq p_j \leq 1$ So we can set $S = \{j\}$.

~~If no such~~

Then Do above proof for S:

$$\Pr\Big(\sum X_i > 0\Big) \geq \Pr\Big(\sum_{i \in S} X_i = 1\Big) \geq \frac{1}{2}\sum_{i \in S} p_i \geq \frac{1}{4}.$$