Today: Use pairwise independent random variables to \underline{derandomize} an algorithm.

Idea: Present a randomized algorithm that works with constant probability and only uses pairwise independent random variables. Then, can iterate through all possible choices of \underline{to find a deterministic choice that is guaranteed to succeed}.

Maximal independent set:

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

IS \( S \) is \underline{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 \( \text{MIS} = \text{Maximal independent set} \).
Now: Parallel algorithm for MIS due to [Luby 85].

$O(\log m)$ rounds & Poly(n) processors

(CREW PRAM model
  = concurrent read, exclusive write)

Idea:

Have a current IS $I$.

Let $G' = G \setminus (I \cup \{I(V|I)\})$ 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 $V' = V' \setminus (\{v\} \cup \{I(V|\{v\})\})$
3. Output $I$. 
This may take $O(n)$ time/rounds.

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

If $\text{SUN}(S)$ 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 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. \( \frac{1}{2d_6(v)} \)
      where \( d_6(v) = \) degree of \( v \) in \( G \).
   c) For every edge \( (y,z) \in E(G) \)
      if yes & yes,
      then remove lower degree in \( G \),
      i.e., remove \( y \) where \( d_6(y) \leq d_6(z) \)
      if \( d_6(y) = d_6(z) \) choose
      Call this new set \( S' \).
   \( G' = G - (S \cup S') \)

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 $c < 1$,

$$E[|E_j| \mid E_{j-1}] < c|E_{j-1}|$$

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

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

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:** \( \frac{1}{2} \) the edges in \( E_j \) 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. \( \geq \alpha := \frac{1}{2} \left( 1 - e^{-\frac{1}{c}} \right) \approx 0.07679 \).

From these 2 claims we get the main lemma:

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

\[
\leq |E_{j+1}| - \alpha | \text{GOOD edges} | 
\]

\[
\leq |E_{j+1}| \left( 1 - \alpha/2 \right) 
\]

which proves the lemma with \( c = 1 - \frac{\alpha}{2} \).
From the lemma we have:

\[ E[1_{E_j}] \leq |E_0|(1 - \frac{x}{2}) \]

\[ \leq m e^{-jx/2} \]

\[ < 1 \quad \text{for} \quad j > \frac{2}{x} \log m \]

Moreover,

\[ \Pr(E_j \neq \emptyset) \leq \Pr(E_j \geq 1) \]

\[ \leq E[1_{E_j}] \]

\[ \leq \frac{1}{4} \quad \text{for} \quad j > \frac{4}{x} \log m. \]

Thus with prob. \( \geq \frac{3}{4} \), we have \( \leq 60 \log m \) 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 $E_B = \text{BAD edges of } G_j$.

We'll define $f: E_B \to (\mathbb{E})$ 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 $\delta_G^-(v) \leq \delta_G^+(w)$.

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 $\geq$ degree.

So these edges point away from $v$.

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

Point to $v$.

So $\geq 2$ times as many out-edges from $v$

as in-edges to $v$.

Hence, for each BAD edge directed into $v$,

there are a pair of out edges out of $v$

that we can uniquely assign to each BAD edge

Thats 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. \( \geq \frac{1}{3} \) \( w \in S \).

Then the claim follows.

Claim a: If \( v \) is good, then \( \Pr(N_6(v) \cap S \neq \emptyset) \geq 2\alpha \)
where \( \alpha := \frac{1}{2} (1 - e^{-\frac{1}{6}}) \)

Proof: Let \( L(v) := \{ w \in N_6(v) : \sigma_6(w) \leq \sigma_6(v) \} \)
= neighbors of \( v \) with lower degree.

Note, for \( v \) good, then \( |L(v)| \geq \sigma_6(v) / 3 \).

\[
\Pr(N_6(v) \cap S \neq \emptyset) = 1 - \Pr(N_6(v) \cap S = \emptyset)
\]
\[
= 1 - \prod_{w \in N_6(v)} \Pr(w \in S) \quad \text{**uses full independence**}
\]
\[
\geq 1 - \prod_{w \in L(v)} \Pr(w \in S)
\]
\[ = 1 - \prod_{\text{wel}(v)} \left(1 - \frac{1}{2d_0(w)}\right) \quad (\text{by defn. of } \Pr(w)) \]
\[ \geq 1 - \prod_{\text{wel}(v)} \left(1 - \frac{1}{2d_0(v)}\right) \quad \text{since } d_0(w) \leq d_0(v) \]
\[ \geq 1 - e^{-\frac{1}{2}} \quad \text{since } 1 \leq v \leq \frac{d_0(v)}{3} \]

Now let's prove:

**Claim b:** \( \Pr(w \notin S \mid \text{wel}) \leq \frac{1}{2} \)

**Proof:** Let \( H(w) = N_6(w) \setminus L(w) = \{ z \in N_6(w) : d_6(z) > d_6(w) \} \)
\[ \Pr(w \& S^* \mid \text{wes}) = \Pr(H(w) \& S^* = \emptyset \mid \text{wes}) \]

Since throw out lower degree endpoint

\[ \leq \sum_{z \in H(w)} \Pr(z \in S \mid \text{wes}) \quad \text{by union bound} \]

\[ = \sum_{z \in H(w)} \frac{\Pr(z \in S \& \text{wes})}{\Pr(\text{wes})} \]

\[ = \sum_{z \in H(w)} \frac{\Pr(z \in S) \Pr(\text{wes})}{\Pr(\text{wes})} \]

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

\[ = \sum_{z \in H(w)} \frac{1}{2g_c(z)} \]

\[ = \sum_{z \in H(w)} \frac{1}{2g_c'(w)} \]

\[ \leq \frac{1}{2} \]
Now from these claims a & b:

\[
\Pr(\nu \in N(S') \mid \nu \text{ is Good})
\]

\[
= \Pr(N_G(\nu) \cap S' \neq \emptyset \mid \nu \text{ is Good})
\]

\[
= \Pr(N_G(\nu) \cap S' \neq \emptyset \mid N(\nu) \cap S' \neq \emptyset, \nu \text{ Good}) \cdot \Pr(N(\nu) \cap S' \neq \emptyset \mid \nu \text{ Good})
\]

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

\[
= \alpha.
\]

Therefore, if \( \nu \) is Good then

\( \nu \) is deleted with prob. \( \geq \alpha \)

(Since if \( N(S') \) are deleted)

That proves Claim 2. Since \( \geq 1 \) endpoint is Good for a Good edge.

& That finishes the analysis of the randomized algorithm.
We used independence for the following:

\[ \Pr(N_6(v) \neq \emptyset) = 1 - \prod_{w \in \mathbb{N}_6(v)} \Pr(w \neq \emptyset) \]

We need a lower bound on this.

We'll prove:

Lemma: For \( X_i \in \Xi, i = 1, \ldots, l \) where \( p_i = \Pr(X_i = 1) \) & \( X_i \)'s are pairwise independent,

\[ \Pr(\sum_{i=1}^{l} X_i > 0) \geq \frac{1}{2} \min \left\{ \frac{1}{a}, \sum_{i=1}^{l} p_i \right\} \]

Letting: \( X_i = 1 \) if \( w_i \in S \)

\[ \Pr(N_6(v) \neq \emptyset) \geq \frac{1}{2} \min \left\{ \frac{1}{a}, \sum \frac{h(v)}{26(v)} \right\} \]

\[ \geq \frac{1}{2} \min \left\{ \frac{1}{a}, \sum \frac{h(v)}{26(v)} \right\} \geq \frac{1}{12} \]
Proof of Lemma:

Case 1: $\sum P_i \leq 1$.

$$Pr(\sum X_i > 0) \geq Pr(\sum X_i = 1)$$

$$\geq \# \sum Pr(X_i = 1) - \frac{1}{2} \sum Pr(X_i = 1 | X_j = 1)$$

$$= \sum P_i - \frac{1}{2} \sum P_i \sum P_j$$

$$\geq \sum P_i - \frac{1}{2} (\sum P_i)^2$$

$$= \sum P_i (1 - \frac{1}{2} \sum P_i)$$

$$\geq \frac{1}{2} \sum P_i \text{ when } \sum P_i \leq 1$$

Case 2: $\sum P_i > 1$:

Find $S \subseteq \{i_1, \ldots, i_j\}$ where $\sum P_{i \in S} \leq 1$.

Find such $S$ so that $\sum P_{i \in S} \leq 1$.

Then do above proof for $S$:

$$Pr(\sum X_i > 0) \geq Pr(\sum X_i = 1) \geq \frac{1}{2} \sum P_i$$

Always exists such a $S$ b/c:

either all $i, P_i < \frac{1}{2}$

or $\exists j \delta \leq P_j \leq 1$

So we can set $S$ for $\frac{1}{2} \sum P_i \geq \frac{1}{2}$