6.852: Distributed Algorithms
Fall, 2015

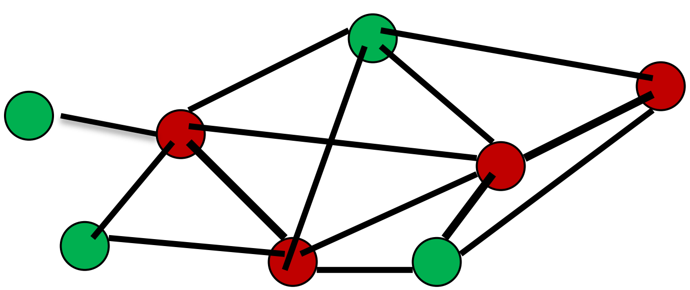
Lecture 4

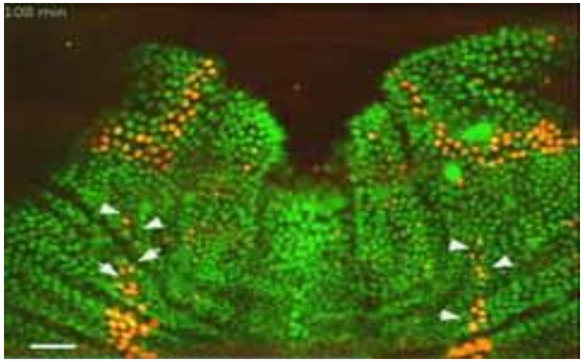
Last time

- Shortest Paths: Distributed Bellman-Ford
- Minimum Spanning Tree (MST): Gallager, Humblet, Spira.
- Maximal Independent Set (MIS): Luby

- Today:
  - Finish Luby MIS analysis
  - Then on to Graph Coloring (Stephan Holzer)

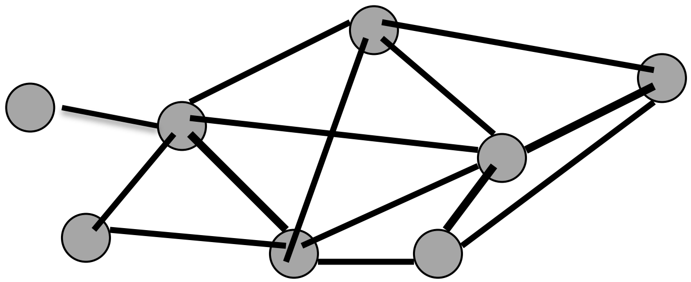
Maximal Independent Set



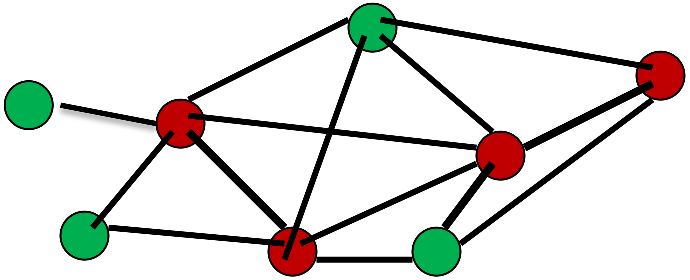


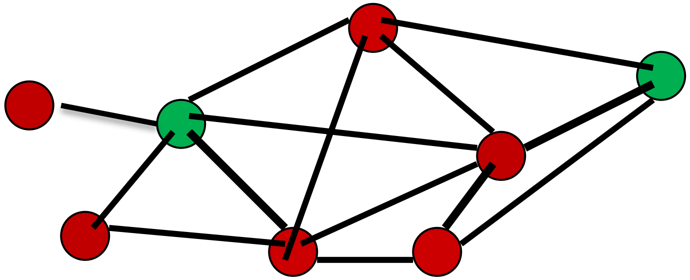
Maximal Independent Set (MIS)

- Assume a general undirected graph network:



- Problem: Select a subset S of the processes, so that they form a Maximal Independent Set.
- Independent: No two neighbors are both in the set.
- Maximal: We can't add any more nodes without violating independence.





Distributed MIS Problem

- Assume:
  - No UIDs
  - Processes know a good upper bound on n.
- Require:
  - Compute an MIS S of the network graph.
  - Each process in S should output $winner$, others output $loser$.

Luby's Algorithm

- Executes in 2-round phases.
- Initially all nodes are $active$.
- At each phase, some active nodes decide to be $winners$, others decide to be $losers$, algorithm continues to the next phase with a smaller graph (removing decided nodes and all their incident edges)
- Repeat until all nodes have decided.

- Behavior of active node i at phase ph:
- Round 1:
  - Choose a random value r in $\{1,2, \ldots, n^5\}$, send it to all neighbors.
  - Receive values from all active neighbors.
  - If r is strictly greater than all received values, then join the MIS, output $winner$.
- Round 2:
  - If you joined the MIS, announce it in messages to all (active) neighbors.
  - If you receive such an announcement, decide not to join the MIS, output $loser$.
  - If you decided one way or the other at this phase, become $inactive$.

Properties

- If it ever finishes, it produces a Maximal Independent Set (showed last time).

- It eventually finishes (with probability 1).
- The expected number of rounds until it finishes is $O(\log n)$.

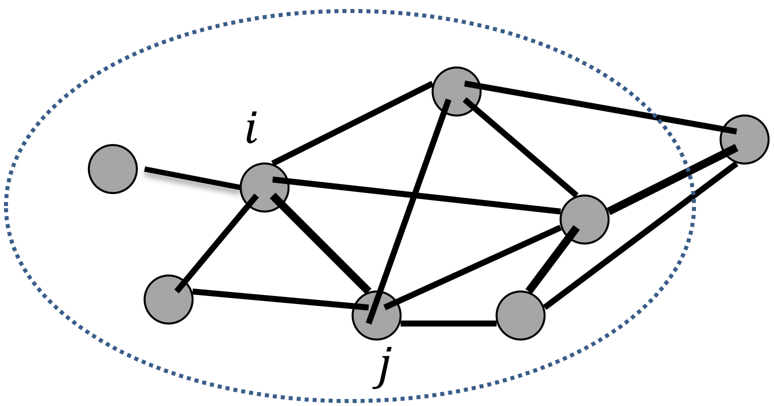
- The following implies both termination facts:
- Theorem 3: With probability at least $1 - \frac{1}{n}$, all nodes decide within $4 \log n$ phases.

Termination

- Theorem 3: With probability at least $1 - \frac{1}{n}$, all nodes decide within $4 \log n$ phases.
- Technical lemma:
- Lemma 4: With probability at least $1 - \frac{1}{n^2}$, in each phase $1, \dots, 4 \log n$, all nodes choose different random values.
- So we can essentially pretend that, in each phase, all the random numbers chosen are different.
- Key idea: Show the graph gets sufficiently "smaller" in each phase.
- Lemma 5: For each phase ph, the expected number of edges that are live (connect two active nodes) at the end of the phase is at most half the number that were live at the beginning of the phase.

Termination

- Lemma 5: For each phase ph, the expected number of edges that are live (connect two active nodes) after the phase is at most half the number that were live before the phase.
- Proof:
  - If node i has some neighbor j whose chosen value is greater than those of all of j's neighbors and all of i's other neighbors, then i must become a $loser$ in phase ph.
  - The probability that j chooses such a value is at least $\frac{1}{\deg(i)+\deg(j)}$.
  - Then the probability node i is "killed" by some neighbor in this way is at least $\Sigma_{j\in\Gamma(i)} \frac{1}{\deg(i)+\deg(j)}$.
  - Now consider an undirected edge $\{i, j\}$.



Termination, cont'd

- Proof:
  - Probability i killed $\geq \Sigma_{j \in \Gamma(i)} \frac{1}{\deg(i)+\deg(j)}$.
  - Probability that edge $\{i,j\}$ "dies"

    $\geq \frac{1}{2}$ (probability i killed + probability j killed).
  - So the expected number of edges that die

    $\geq \frac{1}{2} \Sigma_{\{i,j\}}$ (probability i killed + probability j killed).
  - The sum includes the "kill probability" for each node i exactly $\deg(i)$ times.
  - So rewrite the sum as:

    $\frac{1}{2} \Sigma_i \deg(i)$ (probability i killed).
  - Plug in the kill probability lower bound:

    $\geq \frac{1}{2} \Sigma_i \deg(i) \; \Sigma_{j \in \Gamma(i)} \frac{1}{\deg(i)+\deg(j)}$.

    $= \frac{1}{2} \Sigma_i \Sigma_{j \in \Gamma(i)} \frac{\deg(i)}{\deg(i)+\deg(j)}$.

Termination, cont'd

- Proof:
  - Expected number of edges that die $\geq \frac{1}{2} \Sigma_i \Sigma_{j \in \Gamma(i)} \frac{\deg(i)}{\deg(i)+\deg(j)}$.
  - Write this expression equivalently as a sum over directed edges (i, j):
    $\frac{1}{2} \Sigma_{(i,j)} \frac{\deg(i)}{\deg(i)+\deg(j)}$.
  - Here each undirected edge is counted twice, once for each direction, so this is the same as the following sum over undirected edges $\{i, j\}$.

    $\frac{1}{2} \Sigma_{\{i,j\}} \frac{\deg(i)+\deg(j)}{\deg(i)+\deg(j)} = \frac{1}{2} \Sigma_{\{i,j\}} 1.$
  - This is half the total number of undirected edges, ½ $|E|$, as needed!
- Thus we have:
- Lemma 5: For each phase ph, the expected number of edges that are live (connect two active nodes) at the end of the phase is at most half the number that were live at the beginning of the phase.

Termination, cont'd

- Lemma 5: For each phase ph, the expected number of edges that are live (connect two active nodes) at the end of the phase is at most half the number that were live before the phase.
- Theorem 3: With probability at least $1 - \frac{1}{n}$, all nodes decide within $4 \log n$ phases.
- Proof sketch:
  - Lemma 5 implies that the expected number of edges still live after $4 \log n$ phases is at most $\frac{n^2}{2} \div 2^{4 \log n} = \frac{1}{2n^2}$.
  - Then the probability that *any* edges remain live is $\leq \frac{1}{2n^2}$ (by Markov).
  - The probability that the algorithm doesn't terminate within $4 \log n$ phases $\leq \frac{1}{2n^2} + \frac{1}{n^2} < \frac{1}{\mathrm{n}}$.