MPRI 2.18.1 (2019/20): distributed algorithms for networks, 2^{nd} part

Lecture 2: Rumor Spreading in Complete Graphs (End), Trees, and Grids

Benjamin Doerr, LIX

Outline:
Homework
Randomized rumor spreading in complete graphs: 2^{nd} and 3^{rd} phase
Path arguments: lower bound, degree-diameter bound
Rumor spreading in regular trees and grids
Methods: concentration, Chernoff bounds, Markov chain thinking

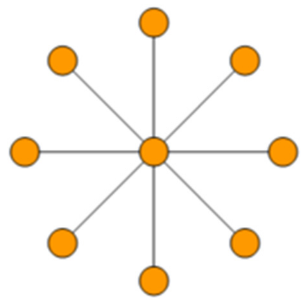
Note: The lecture materials for this lecture also includes a page (electronic) of homework problems.

Contents

- Randomized rumor spreading (RSS) in complete graphs:
  - Proof: From $n^{0.4}$ to $o(n)$ informed nodes in $\big(0.6 + o(1)\big)\log_2 n$ rounds
  - Proof: From $o(n)$ informed to $o(n)$ uninformed in very short time
- RSS in k-regular trees of height h: $\Theta(h \log n)$ rounds.
- RSS in k^d grids: $O\ (d^2 k)$ rounds
- Useful arguments:
  - Waiting time argument: Expected waiting time = 1 / success_probability
  - Counting random things: Indicator random variables & linearity of expectation
  - Concentration, tail bounds: Chernoff bounds, method of bounded differences
  - Tail bound + union bound !
  - Markov chain thinking: add expected waiting times

Homework 1.1: Rumors in Star Graphs

- Let G be the star graph on n vertices, that is, G has n vertices such that there is one *central node* that is connected to all others via a direct edge, and apart from this, there are no edges. What is the expected time $E[T]$ until randomized rumor spreading has informed this graph?

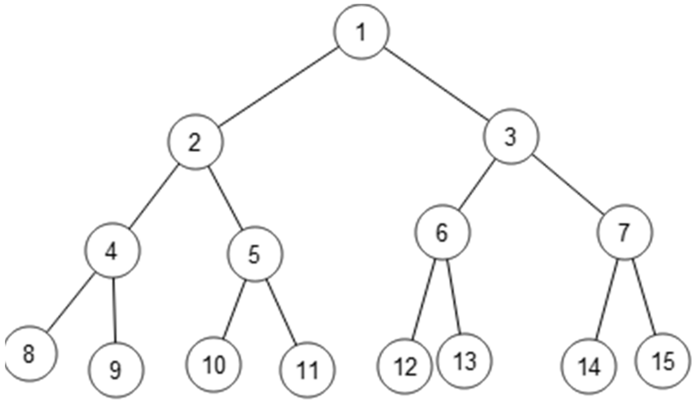


- Observation: More or less, this is the coupon collector problem!
- Details: If the central node and k of the leaves are informed, then one round with probability $p_k = \frac{n-1-k}{n-1}$ informs a new leaf, otherwise nothing changes.
  - Hence the expected *waiting time* for this event is $\frac{1}{p_k} = \frac{n-1}{n-1-k}$.
  - Rumor starts in center: $E[T] = \sum_{k=0}^{n-2} \frac{n-1}{n-1-k} = (n-1)\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n-1}\right)$
  - Rumor starts in a leaf: After one round, also the center is informed and thus $E[T] = 1 + \sum_{k=1}^{n-2} \frac{n-1}{n-1-k} = (n-1)\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n-1}\right)$
  - In both cases $\boldsymbol{E[T] = (n-1)H_{n-1}}$, where $H_k = 1 + \frac{1}{2} + \cdots + \frac{1}{k}$

Harmonic number H_k:
$\ln k \le H_k \le \ln(k) + 1$

Homework 1.4: Rumors in Trees

- Let G be a k-regular rooted tree of height h, that is, an undirected graph having $n = 1 + k + k^2 + \cdots + k^h$ vertices such that there is one "root" vertex which has k neighbors such that each of them is the root of a k-regular tree of height $h - 1$ (when we delete the original root and all edges incident with it).
- Assume that you run the randomized rumor spreading protocol in this graph, starting the rumor in the root.



- Warning: 3 different *waiting times* !
  - The expected waiting time for a node (except the root) to call a particular neighbor is exactly $k + 1$ (k for the root)
  - For each fixed node, the expected time it needs to call *all* its neighbors is $\Theta(k \log k)$ [coupon collector]
  - If we run n such coupon collector experiments independently in parallel, the time for *all* to finish is $\Theta(k \log(kn)) = \Theta(k \log(n^2)) = \Theta(k \log n)$ [proof of the upper bound: next slide]

Homework 1.4: Rumors in Trees (2)

- **Theorem:** With high probability after $O(hk \log n) = O(h^2 k \log k)$ rounds everyone is informed.

- Proof:
  - Let x be a node (of degree $k+1$ or smaller) and y be a neighbor of x.
  - The probability that x does not call y within $3(k+1)\ln n$ rounds *after becoming informed* is at most
  $$\left(1 - \frac{1}{k+1}\right)^{3(k+1)\ln n} \leq \exp\left(-\frac{1}{k+1}\, 3(k+1)\ln n\right) = n^{-3}.$$
  - Union bound over all pairs (x, y): The probability that there is a node that has not called all its neighbors within $3(k+1)\ln n$ rounds *after being informed* is at most $n\,(k+1)n^{-3} \leq n^{-1}$.
  - Hence with prob. $1 - n^{-1}$, each node calls all its neighbors within $3(k+1)\ln n$ rounds after being informed. In this case, all nodes are informed after $h \cdot 3(k+1)\ln n$ rounds (max. distance to root is h).

- Is this tight, that is, we really need $\Theta(hk \log n)$ round? – No, see later… ☹

Topics Today

- Finish the proof that randomized rumor spreading informs all nodes of the complete graph on n vertices in $(1+o(1))(\log_2 n + \ln n)$ rounds
  - Phase 2: from $n^{0.4}$ to $o(n)$ informed nodes. The number of informed nodes almost doubles each round → $\le \big(0.6 + o(1)\big) \log_2 n$ rounds.
  - Phase 3: from $i = o(n)$ informed nodes to $o(n)$ uninformed nodes in almost no time ☺

- Path arguments: How does a rumor “follow” a path?
  - rumor spreading in k-ary trees of height h: $\Theta(h \log n)$
  - rumor spreading in $k \times k$ grids: $\Theta(k)$

- Methods:
  - Simple Markov chain thinking
  - Concentration, large deviation bounds → get what you expect ☺

Reminder: RRS in Complete Graphs

<u>The rumor spreading process can be split into 4 phases:</u>

- 1st phase (up to $o(\sqrt{n})$ informed nodes): True doubling. With high probability, all calls in one round reach a “new” node, that is, the number of informed nodes doubles in these rounds. → “Birthday paradox.”

- 2nd phase (up to $o(n)$ informed nodes): Exponential growth. Each call still has a probability of $1 - o(1)$ of reaching a “new” node. Hence the number of informed nodes almost doubles, that is, increases by a factor of $(2 - o(1))$ in these phases

- 3rd phase: a short connecting phase between the exponential growth of the informed nodes and the exponential shrinking of the uninformed nodes

- 4th phase (from $n - o(n)$ informed nodes on): Exponential shrinking of the uninformed nodes by a factor of $(e - o(1))$. → “Coupon collector”

Phase 3: Short Connecting Phase (Also Homework 1.2)

- Lemma: Let $f, g = o(1)$. Assume that we have $i = nf$ nodes informed. Let $T = 1/(fg)$. Then with probability $1 - o(1)$, after T rounds at least $n(1-g)$ nodes are informed.

- Proof (same idea as for Phase 4):
  - The probability that an uninformed node remains uninformed during these T rounds is at most $\left(1 - \frac{1}{n}\right)^{nfT} \leq \exp(-1/g)$.
  - Thus the expected number of uninformed nodes is at most $(n - i)\exp(-1/g) \leq n\exp(-1/g)$,
  - and the probability that we have more than ng uninformed nodes is at most $n\exp(-1/g)/ng = o(1)$ by Markov's inequality.
    - $\Pr[X \geq \lambda] \leq E[X]/\lambda$ for all non-negative random variables X

- Comment: As in Phase 4, we ignore that newly informed nodes join the process. This now loses a lot, but the phase is very short, so we don't care.

Phase 2: Expected Number of Newly Informed Nodes (Hw 1.3)

- Reminder: We regard randomized rumor spreading (RSS)
  - in the complete graph on n vertices
  - with the assumption that nodes call random nodes including themselves

- Assume that a round starts with i nodes informed.
- Let X denote the number of nodes newly informed in this round
  - that is, the number of nodes that were not informed at the start of the round, but were informed at the end

- Claim: $E[X] \geq i\left(1 - \frac{3i}{2n}\right)$

- Two small technical difficulties:
  - How to count random objects?
  - Collisions: two informed nodes may call the same uninformed node

Removing the Collisions

- We pretend that the informed nodes do their calls one after the other.
- Node k is *successful* if it calls a node which at that moment is uninformed.
- → Number newly informed nodes = number of successful calls
- Comments:
  - All we do here is symmetry-breaking – if a node is called by several nodes, then we declare one of them as the successful caller
  - We used this trick already in the analysis of phase 1, but didn't say so ☺

Counting Successful Calls (=Newly Informed Nodes)

- *Indicator random variable* for a successful calls:
  - $X_k := \begin{cases} 1, \text{if the } k\text{th node calls a node uninformed at that time} \\ 0, \text{otherwise} \end{cases}$
- Counting: The number of successful calls is $X := \sum_{k=1}^{i} X_k$
- The expected number of successful calls is at least
  - $E[X] = E\left[\sum_{k=1}^{i} X_k\right] = \sum_{k=1}^{i} E[X_k] \geq \sum_{k=1}^{i} \frac{n-i-(k-1)}{n} \geq i\left(1 - \frac{3i}{2n}\right)$

Linearity of expectation

Expectation of binary random variables: $E[X_k] = \Pr[X_k = 1]$

- Generally a good way of counting random objects!

Alternative Proof: "Remaining Uninformed"

- As in Phase 3 and 4:
  - $\Pr[x \text{ remains uninformed}] = \left(1 - \frac{1}{n}\right)^i$
  - $\Pr[x \text{ becomes informed}] = 1 - \left(1 - \frac{1}{n}\right)^i \geq 1 - \frac{1}{1+i/n} \geq \frac{i}{n}\left(1 - \frac{i}{n}\right) = p$

Elementary estimate

  - $E[\text{newly informed nodes}] = (n-i)p = \frac{i}{n}\left(1 - \frac{i}{n}\right)^2 \geq \frac{i}{n}\left(1 - \frac{2i}{n}\right)$

- Gives the more or less the same result
  - exact result before the elementary estimate
  - weaker in the unimportant $1 - \Theta\left(\frac{i}{n}\right)$ term after the estimate
- Less intuitive as we see less why we get this bound

If the Expectation Was the Truth: Traditional Proof

- Lemma [previous slides, weaker constant]: A round starting with i informed nodes ends with an expected number of at least $2i\left(1-\frac{i}{n}\right)$ informed nodes
- *Assume that this was not an expectation, but the truth with probability 1.*
- Phase 2a: Going from i_0 to $i_1 = n/(\log n)^2$ informed nodes.
  - In each round, the number of informed nodes multiplies by at least $2(1-(\log n)^{-2})$ [uniform pessimistic estimate from Lemma]
  - If we would do this for $T_1 = \log(n/i_0) := \log_2(n/i_0)$ rounds, we would have $i_0 \cdot \left(2(1-(\log n)^{-2})\right)^{T_1} \geq n(1-T_1(\log n)^{-2}) \geq i_1$ informed nodes, hence we need at most T_1 rounds to go to i_1 informed nodes
    - Bernoulli inequality: $(1-x)^t \geq 1 - xt$
- Phase 2b: Going from i_1 to $i_2 := n/(\log\log n)^2$ informed nodes
  - same argument with (roughly) all $\log n$ replaced by $\log\log n$: $T_2 = 2\log\log n$ rounds do the job

If the Expectation Was the Truth: Modern Proof [D.&Künnemann'14]

- Lemma: *Assume that with prob. 1,* a round starting with i informed nodes ends with at least $2i\left(1-\frac{3i}{4n}\right)$ informed nodes [sharp constant]
- Start this process with i_0 informed nodes
- Let $i_t = 2i_{t-1}\left(1-\frac{3i_{t-1}}{4n}\right)$ for all $t \geq 1$
- Then
  - $i_t = 2^t i_0 \prod_{s=0}^{t-1}\left(1-\frac{3i_s}{4n}\right)$ [fake induction]
  - $\geq 2^t i_0\left(1-\frac{3}{4n}\sum_{s=0}^{t-1} i_s\right)$ [Bernoulli's inequality]
  - $\geq 2^t i_0\left(1-\frac{3}{4n}\sum_{s=0}^{t-1} 2^s i_0\right)$ [$i_s \leq 2i_{s-1}$ for all s]
  - $\geq 2^t i_0\left(1-\frac{3}{4n}\,2^t i_0\right)$
- Take t such that $2^t i_0 \in \left[\frac{1}{4}n, \frac{1}{2}n\right]$. Then $i_t \geq \frac{n}{4}\left(1-\frac{3}{4n}\cdot\frac{n}{2}\right) = \frac{5}{32}n$ ☺

Recall
$i_t \leq 2^t i_0 \leq \frac{1}{2}n$

Making the Expectation (Almost) the Truth: *Chebyshev's Inequality*

- $X_k := \begin{cases} 1, & \text{if the } k\text{th call informs a node uninformed at that time} \\ 0, & \text{otherwise} \end{cases}, \quad k = 1, \dots, i$
- $\Pr[X_k = 1] \geq \frac{n-i-(k-1)}{n}$
- Number of newly informed nodes: $X := \sum_{k=1}^{i} X_k$
- $E[X] = E\left[\sum_{k=1}^{i} X_k\right] = \sum_{k=1}^{i} E[X_k] \geq \sum_{k=1}^{i} \frac{n-i-(k-1)}{n} \geq i(1 - 1.5\, i/n)$
- Let's compute (estimate) the *variance* $(Var[X] := E[(X - E[X])^2])$:
  - $Var[X] = Var\left[\sum_{k=1}^{i} X_k\right] = \sum_{k=1}^{i} Var[X_k] + 2 \sum_{k=1}^{i} \sum_{l=k+1}^{i} Cov[X_k, X_l]$
    - $Var[X_k] \leq \Pr[X_k = 1] \leq i/n$
    - $Cov[X_k, X_l] = \Pr[X_k = X_l = 1] - \Pr[X_k = 1] \Pr[X_l = 1] < 0$
  - $Var[X] \leq i^2/n$
- *Chebyshev*: $\Pr[X \leq E[X] - i^{0.75}] \leq \frac{Var[X]}{(i^{0.75})^2} \leq i^{0.5} n^{-1} \leq n^{-0.5}$
- For $i \leq n/4$, we have $E[X] \geq 0.5i$ and $\mathbf{Pr\left[X \leq E[X]\left(1 - 2i^{-0.25}\right)\right] \leq n^{-0.5}}$

Chebyshev + Union Bound = Done

- Expectation&Chebyshev: Consider a round starting with $i \leq n/4$ informed nodes. Let X denote the number of informed nodes after the round.
  - $E[X] \geq 2i\left(1 - \frac{3i}{4n}\right)$
  - $\Pr[X \leq E[X](1 - 2i^{-0.25})] \leq n^{-0.5}$

- Union bound: With probability $1 - \frac{2\log n}{\sqrt{n}}$, we have
$$X \geq 2i\left(1 - \frac{3i}{4n}\right)(1 - 2i^{-0.25})$$
in all of the first $2\log n$ rounds that start with $i \leq n/4$ informed nodes.

- In this cases, using the “modern proof”, we have
  - $i_t = 2^t i_0 \prod_{s=0}^{t-1}\left(\left(1 - \frac{3i_s}{4n}\right)(1 - 2(i_s)^{-0.25})\right) \geq \frac{5}{32}n(1 - 2(i_0)^{-0.25})^t$
  - $\geq \frac{5}{32}n\,(1 - 2t(i_0)^{-0.25}) = \frac{5}{32}n\,(1 - o(1))$ when $i_0 \gg \log n$

Summary Phase 2

- Traditionally a challenge, but with the modern proof less frightening:
  The expected number of informed nodes almost doubles, but the gap to true doubling increases the closer we get to $n/2$ informed nodes.
  → makes it hard to treat all rounds the same!

- Result: The prices for the imperfect doubling is only a constant number of rounds

- “Only” a proof-problem: We can not argue with the expectation alone!
  → Solution: Argue that the true number of informed nodes with high probability is close to the expectation

Concentration!

- Strong concentration: A random variable X is said to be strongly concentrated (around its mean) when X is usually close to $E[X]$, that is, when the deviations $\Pr[|X - E[X]| \geq \lambda]$ are small.

- Application 1: If the concentrations are strong enough, we can (almost) pretend that our random variables are not random, but just equal the expectation.
  - example: the previous analysis of phase 2

- Application 2: strong concentration → good union bounds
  - example: analysis of rumor spreading in k-ary trees (earlier this lecture)
    - if T_{xy} is the time a node x of degree k needs to call a particular neighbor y, then $\Pr\left[T_{xy} \geq \lambda E\left[T_{xy}\right]\right] = (1-1/k)^{\lambda k} \leq \exp(-\lambda)$
    - union bound: all nodes x inform all neighbors y within $\lambda E\left[T_{xy}\right]$ rounds $\Pr\left[\exists x, y \colon T_{xy} \geq \lambda E\left[T_{xy}\right]\right] \leq \sum_{x,y} \Pr\left[T_{xy} \geq \lambda E\left[T_{xy}\right]\right] \leq nk \exp(-\lambda)$
    - → $\lambda = 2 \ln nk$ suffices!

Proving Strong Concentration

How can we show that some random variable X is, with high probability, not too far from $E[X]$ ("strong concentration")?

- Elementary tools:
  - direct methods: analyzing the distribution of X by hand
    - examples: coupon collector, degree-k node calling neighbors, …
  - Markov's inequality: $\Pr[X \geq \lambda E[X]] \leq 1/\lambda$ for non-negative X.
    - relatively weak, only bounds for the "upper tail"
  - Chebyshev's inequality: $\Pr\left[|X - E[X]| \geq \lambda\sqrt{Var[X]}\right] \leq 1/\lambda^2$.
    - equivalent: $\Pr[|X - E[X]| \geq \kappa] \leq Var[X]/\kappa^2$
- Less elementary: Chernoff bounds… (next slide)
  - exponentially small deviation probabilities (*when applicable*)

Chernoff Bounds

- Let $X_1, \ldots, X_n$ be independent random variables taking values in $[0,1]$.
  Let $X = X_1 + \cdots + X_n$.

- Multiplicative bounds: For $\delta \in [0,1]$,
  - $\Pr\left[X \leq (1-\delta)E[X]\right] \leq \exp\left(-\frac{\delta^2 E[X]}{2}\right)$,
  - $\Pr\left[X \geq (1+\delta)E[X]\right] \leq \exp\left(-\frac{\delta^2 E[X]}{3}\right)$.

- Additive bounds: For $\lambda \geq 0$,
  - $\Pr[X \geq E[X] + \lambda] \leq \exp\left(-\frac{2\lambda^2}{n}\right)$,
  - $\Pr[X \leq E[X] - \lambda] \leq \exp\left(-\frac{2\lambda^2}{n}\right)$.

Exponentially small deviation probabilities!!!

- There are lots of other/stronger Chernoff bounds, but these four suffice for most purposes. For much more on Chernoff bounds, see, e.g., https://arxiv.org/abs/1801.06733

Method of Bounded Differences

- Theorem (McDiarmid, **Azuma**, Hoeffding…): Let $X_1, \dots, X_m$ be independent random variables taking values in some sets $A_1, \dots, A_m$.
  - Let $f: A_1 \times \cdots \times A_m \to \mathbb{R}$.
  - Lipschitz condition: Assume that $|f(a) - f(b)| \le c_i$ whenever a, b differ only in the ith component.
  - Then we have the *additive* tail bounds
    - $\Pr[f(X_1, \dots, X_m) \le E[f(X_1, \dots, X_m)] - \lambda] \le \exp(-2\lambda^2 / \sum_{i=1}^m c_i^2)$,
    - $\Pr[f(X_1, \dots, X_m) \ge E[f(X_1, \dots, X_m)] + \lambda] \le \exp(-2\lambda^2 / \sum_{i=1}^m c_i^2)$.

- Application to one round in Phase 2 (starting with i informed nodes):
  - independent rvs: i random calls
  - f: number of newly informed nodes (as determined by the i calls)
  - influence of one call on the number of newly informed nodes: ≤ 1
  - Azuma with $m = i$ and $c_k = 1$: $\Pr[X \le E[X] - i^{0.75}] \le \exp(-2n^{0.5})$

Chebyshev vs. Chernoff

- Both prove what we need: If we have $n^{0.4}$ nodes informed, then after $0.6\ \log n + O(1)$ rounds, with high probability $\frac{5}{32}\, n\,(1 - o(1))$ nodes are informed.
  - failure probability with Chebyshev: $n^{-\Theta(1)}$
  - failure probability with Chernoff: $\exp\left(-n^{\Theta(1)}\right)$

- Chebyshev: Needs a bound for the variance (which can be nasty)

- Chernoff/Azuma: Needs independent random variables.
  - → the reason why many randomized algorithms build on independent random decisions!

Part 2: Path-Arguments

- So far: Analyze how many nodes become informed in a round starting with i arbitray informed nodes
  - Ignores the structure of the graph
  - Works well when the graph is highly symmetric (complete graphs, random graphs $G(n, p)$)

- Now: Use the argument "how long does it take for the rumor to traverse a given path"
  - Hope: Allows to take into account the graphs structure

A Simple Lower Bound

- Notation: Let $G = (V, E)$ be an undirected graph
  - $d(s, v)$: smallest length (=number of edges) of a path between s and v
  - $diam(G) = \max\{d(u, v) | u, v \in V\}$ *diameter*

- Trivial observation: If the rumor starts in s, then after k rounds all informed nodes v satisfy $d(s, v) \leq k$. [Formal proof: induction]

- **Lemma:** Regardless of where the rumor starts, the time to reach all nodes is at least $\frac{1}{2}\, diam(G)$.
- **Proof:** Let s be the node where the rumor starts.
  - Let v be a node with $d(s, v) \geq \frac{1}{2}\, diam(G)$ – note that such a v exists as otherwise $d(u, v) \leq d(u, s) + d(s, v) < diam(G)$ for all u, v (contradiction).
  - By the observation, v becomes informed not earlier than in round $d(s, v)$.

Upper Bounds: Expected Path Traversal Times

- **Path-Lemma for Expectations:** Let $P: x = x_0, x_1, \dots, x_k = y$ be any path from x to y in G. Assume that the rumor starts in x. Then the first time T_y when y is informed satisfies $\boldsymbol{E[T_y] \leq \sum_{i=0}^{k-1} \deg x_i}$.

- **Argument** (details: homework)
- When x_i is informed, then in each round with probability $\frac{1}{\deg x_i}$ is calls x_{i+1}.
- The expected waiting time for this event is $\deg x_i$
- Hence $E[T_{x_{i+1}}] \leq E[T_{x_i}] + \deg x_i$.
- By induction, we obtain $E[T_{x_i}] \leq \sum_{j=0}^{i-1} \deg x_j$.

- This is a Markov chain argument: What happens in the current round only depends on who is informed, not on how they became informed
→ allows to add up waiting times

Expectations vs. Tail Bounds

- Previous slide: Expected time for the rumor to traverse one path.
- Problem: This does not help when we want to say something on the time $T = \max_y T_y$ to inform all vertices.
  - $E[\max_y T_y]$ can be larger than $\max_y E[T_y]$
  - Example star graph, rumor starts in center c: $E[T_y] = n - 1$ for all $y \in V \setminus \{c\}$, but $E[T] = \Theta(n \log n)$.
- Plan: Tail bound and union bound!
  - Let t be such that $\Pr[T_y > t] \leq n^{-2}$ for all $y \in V$.
  - Then $\Pr[T > t] = \Pr[\exists y \in V : T_y > t] \leq \sum_{y \in V} \Pr[T_y > t] \leq n^{-1}$

Tail bound

Union bound

Path-Lemma (Tail Bound), Degree-Diameter Bound

- **Path-Lemma for Tail-Bounds:**
  - Let G be any graph. Assume that the rumor starts in a vertex x_0.
  - Let $P: x_0, \dots, x_k$ be any path of length k in G.
  - Let $\Delta := \max\{\deg x_i \mid i \in [0..k-1]\}$ be the maximum degree of the vertices on P.
  - Let $k' \geq k$. ["pretend longer path to get better probability"]
  - Then after $T = 2k'\Delta$ rounds, the whole path is informed with probability $1 - \exp(-k'/4)$.

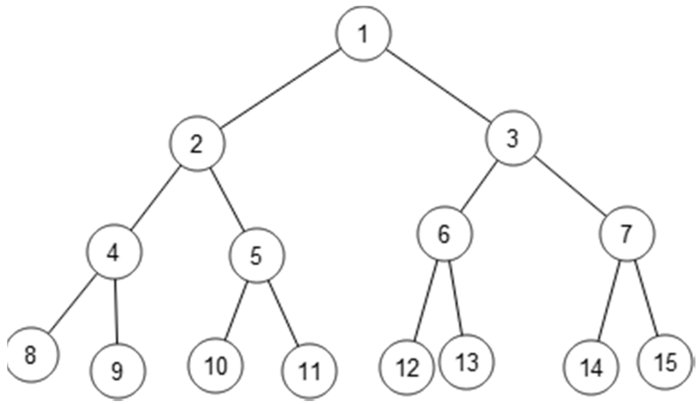
- **Corollary (degree-diameter bound):** All vertices are informed after $\boldsymbol{T = 2\Delta \max\{diam(G), 8 \ln n\}}$ rounds with probability $1 - n^{-1}$
  - Proof: Use path lemma with $k' = \max\{diam(G), 8 \ln n\}$.

- Proof of the path lemmas: Guided homework.

Rumor Spreading in Regular Trees

- Let G be a k-regular rooted tree of height h, that is, an undirected graph having $n = 1 + k + k^2 + \cdots + k^h$ vertices such that there is one "root" vertex which has k neighbors such that each of them is the root of a k-regular tree of height $h - 1$ (when we delete the original root and all edges incident with it).

- **Theorem:** Let T be the first round after which RSS, started in an arbitrary node, has informed all nodes of the k-regular tree of height h. Then $T = \Theta(hk \log k) = \Theta(k \log n)$.



- **Proof:**
- Upper bound:
  - diameter $\Theta(h) \leq O(\log n)$
  - max-degree $\Theta(k)$
  - degree-diameter bound: $T = O(\Delta \max\{diam(G), \log n\}) = O(k \log n)$
- Lower bound: Homework

Rumor Spreading in Grids

- Let $G = (V, E)$ be a d-dimensional grid of side length k, that is, $V = [1..k]^d$ and two vertices are neighbors if they differ in exactly one coordinate and this difference is exactly one.

- **Theorem:** A rumor starting in an arbitrary vertex reaches all vertices in expected time $O(d^2 k)$.

- **Proof:** Use the degree-diameter bound. $\Delta = \Theta(d)$ and $\mathrm{diam}(G) = d(k-1)$, the latter being $\Omega(\log n)$ for all values of d and k

- Comment:
  - For $d = \Theta(1)$, this is tight (diameter is a lower bound).
  - For $k = 2, d = \log_2 n$ (hypercube), the truth is $O(\log n)$ instead of the above $O(\log^2 n)$.

Summary

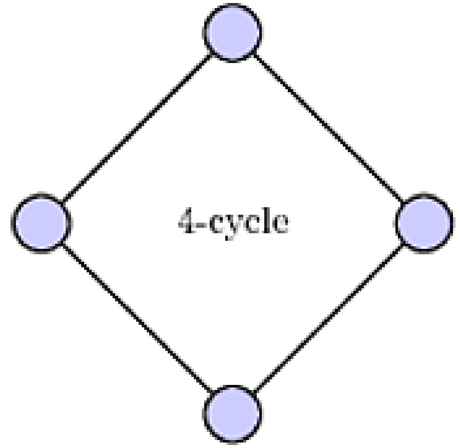
- Analysis of the 2nd phase of RRS in complete graphs.
  - expect almost doubling: factor $2\left(1 - \frac{3i}{4n}\right)$
  - “modern proof”: growing loss compared to perfect doubling can be taken care of with elementary arguments ☺
  - expectation → reality: strong concentration

- Path arguments:
  - give good bounds for regular trees and grids

- Methods:
  - Markov chain arguments: add waiting times
  - Strong concentration → tail bounds, work well with union bound.

Homework

- Homework 2.1: In expectation, how long does it take until randomized rumor spreading (with nodes calling random neighbors, not including themselves) informs a cycle on 4 vertices?



- Homework 2.2: Give an example showing the path lemmas can greatly over-estimate the time it takes to inform a vertex. Why?

- Homework 2.3: Prove the path lemmas (separate pdf document)

Homework (2)

- Homework 2.4: Prove the lower bound of our result on rumor spreading in k-regular trees of height h, that is, that a rumor started in an arbitrary node needs at least $T = \Omega(hk \log k) = \Omega(k \log n)$ rounds to reach all nodes.
Try the following *two proof ideas*:
  - *Path thinking:* Assume that the rumor starts in the root of the tree and construct a path to a leaf along which the rumor is slow.
  - *Coupon collector thinking:* Assume that all nodes except the leaves are already informed. Show that after $ck \ln n$ rounds, c a sufficiently small constant, with high probability a decent number of nodes is uninformed. Hint: compute the expected number of uninformed nodes and use the method of bounded differences to turn the expectation into a with-high-probability statement.

Homework (3)

- Homework 2.5: [continuation of homework 1.5]
  - [difficult, though purely graph theoretic] If you haven't solved 1.5, try again. Recall that the claim of 1.5 was that *in any graph on n vertices, for any initially informed vertex x and any target vertex y, the expected number of rounds after which y is informed is at most $3n$*. Since we want to use the path lemma, what is the purely graph theoretic result we need? Prove this innocent lemma.
  - Use the result of 1.5 to show that in any graph, regardless of the starting vertex, after $O(n \log n)$ rounds all vertices are informed with high probability [easy].
    - Hint: Find a very general elementary argument showing that any particular vertex with probability at least $1 - n^{-2}$ is informed after $O(n \log n)$ rounds. No Chernoff bound is needed for this concentration statement.