Recitation 2

Median Finding & Randomized Algorithms

Rebecca Lin | Friday, September 13th, 2024

How’s everyone doing?

Median Finding

Given a set S of n distinct elements and a number $i \in \{1,2,\ldots,n\}$, find the element $x \in S$ such that $\text{RANK}(x) = i$, that is, the ith smallest element.

8	-1	19	11	12	0	-7	3	-4

-7	-4	-1	0	3	8	11	12	19

sorting requires $O(n \log n)$ time

can we do better?

Analysis

Bad: If $|L| = 0$ at each level, then

$$T(n) = T(n-1) + O(n) = O(n^2)$$

Good: If $|L|, |G| \leq cn$ for some constant $c < 1$ at each level, then:

$$T(n) = T(cn) + O(n) = O(n)$$

Goal: In $O(n)$ time, pick an x^* that is "c-balanced":

$$\max\{\text{RANK}(x^*), n - \text{RANK}(x^*)\} \leq cn$$

Algorithm

SELECT(i, S)

"clever" selection

1. Divide S into $\frac{n}{5}$ groups of 5 elements each, padded by large numbers, if necessary
2. Find the median of each 5-element group by sorting
3. Recursively SELECT the median x^* of the $\frac{n}{5}$ group medium as the pivot
4. Compute $L = \{y \in S : y < x^*\}$ and $G = \{y \in S : y > x^*\}$
5. Since so RANK$(x^*) = |L| + 1$:
   - If RANK$(x^*) = i$, then $x = x^*$
   - If RANK$(x^*) > i$, then SELECT(i, L)
   - If RANK$(x^*) < i$, then SELECT$(i - |L| - 1, G)$

as before (pg.2)

Algorithm

SELECT(i, S)

1. Divide S into $\frac{n}{5}$ groups of 5 elements each, padded by large numbers, if necessary
2. Find the median of each 5-element group by sorting

$O(n)$

3. Recursively SELECT the median x^* of the $\frac{n}{5}$ group medium as the pivot

$T\left(\frac{n}{5}\right)$

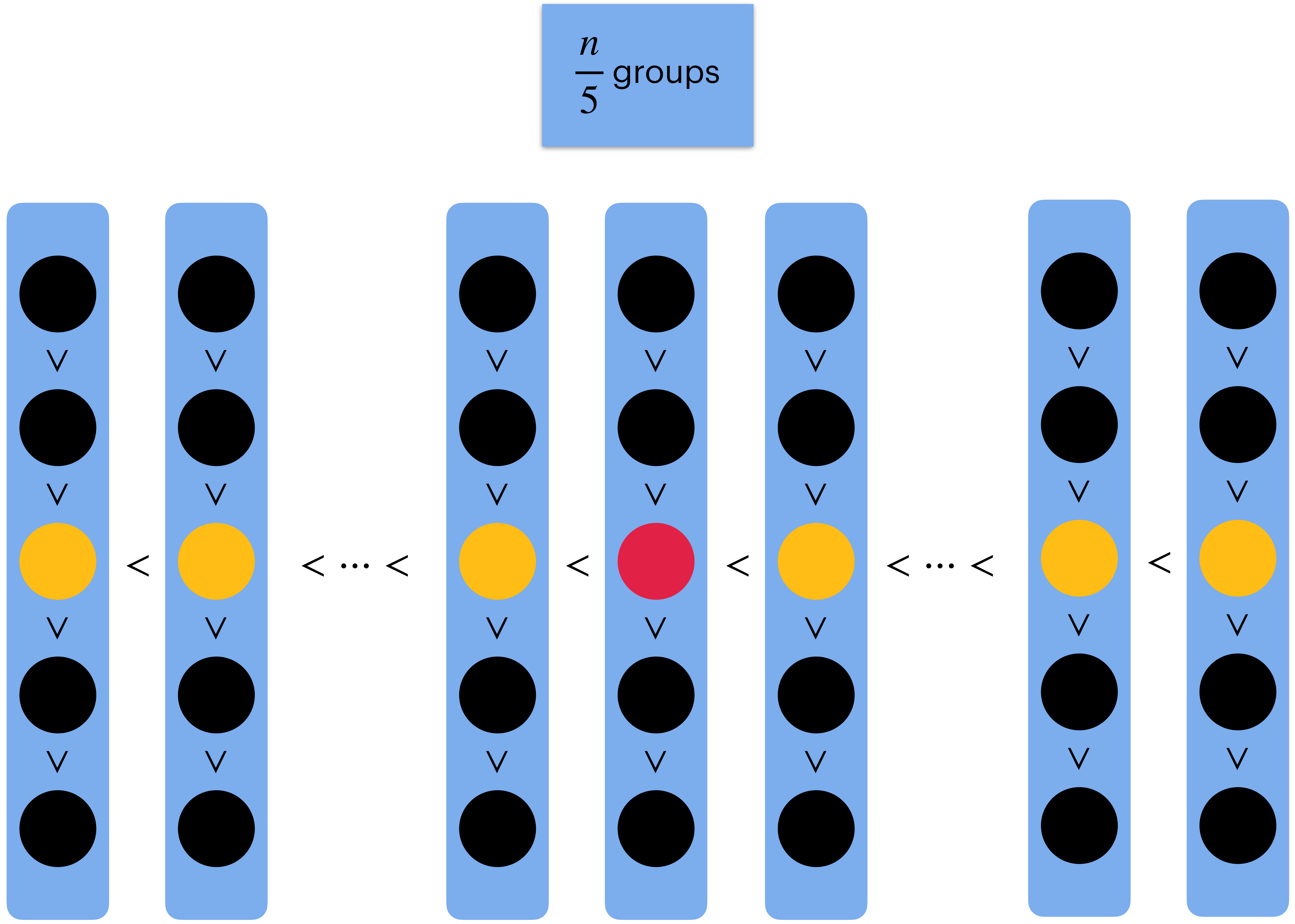
4. Compute $L = \{y \in S : y < x^*\}$ and $G = \{y \in S : y > x^*\}$

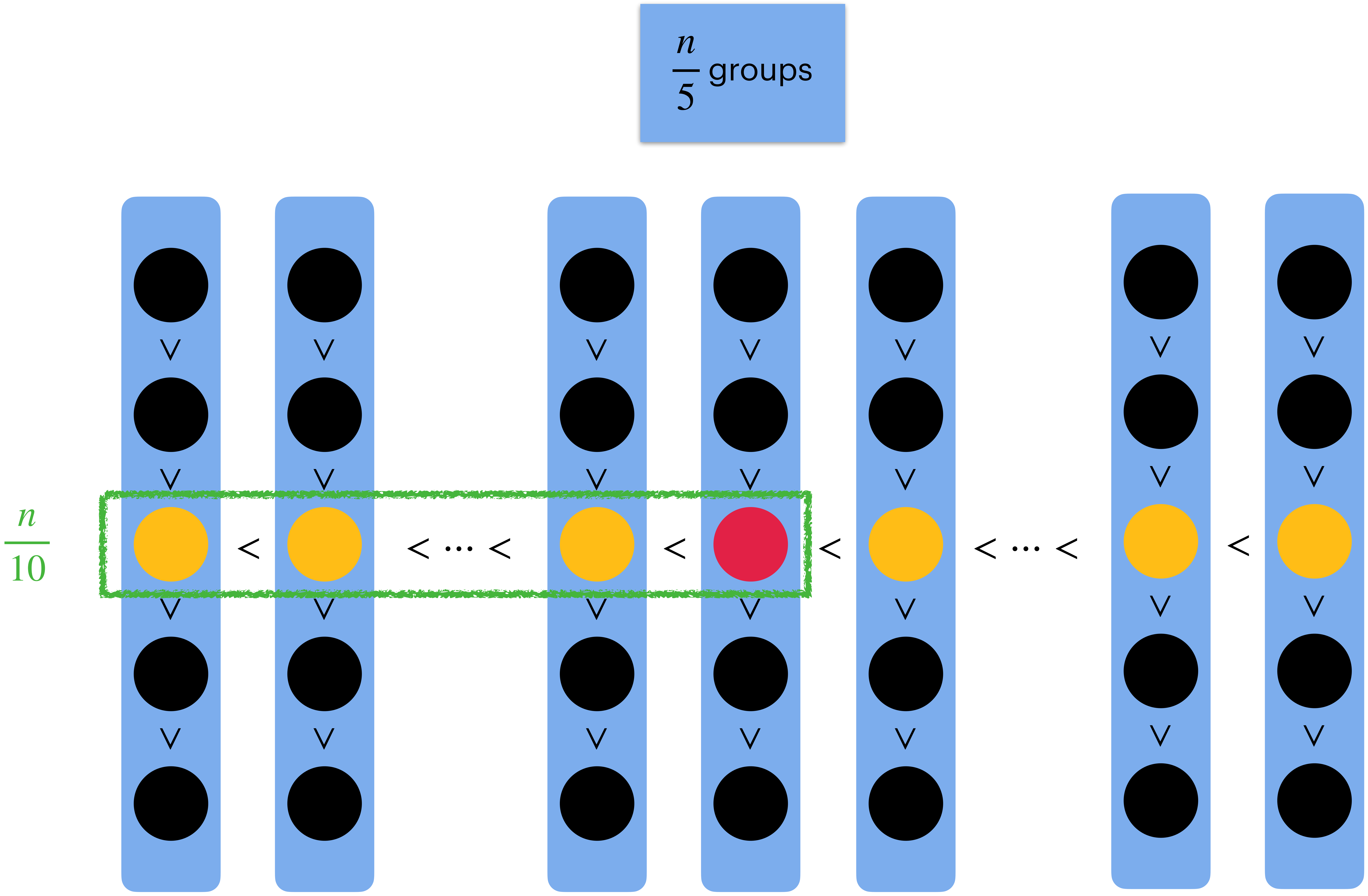
$O(n)$

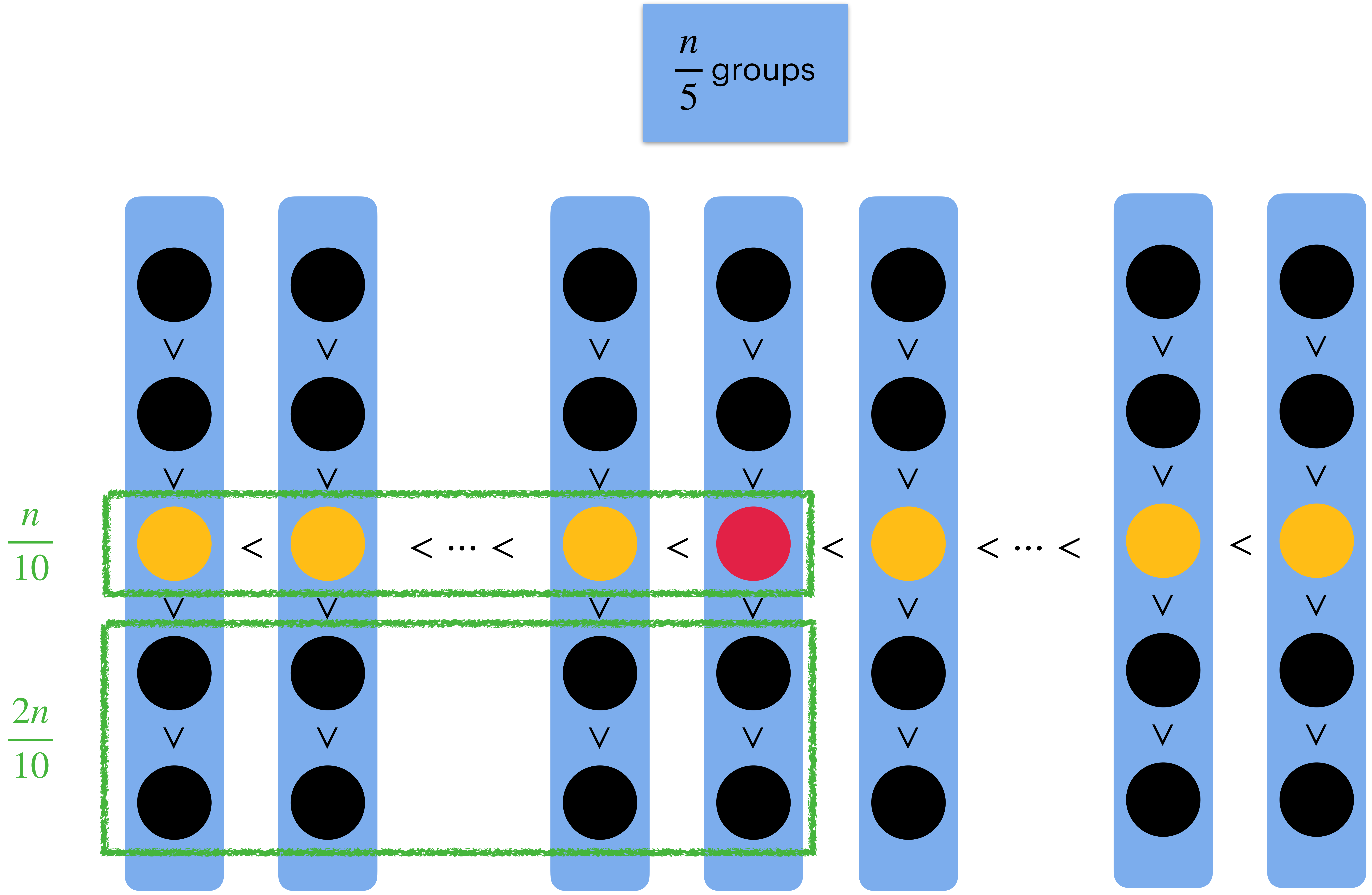
5. Since so RANK$(x^*) = |L| + 1$:
   - If RANK$(x^*) = i$, then $x = x^*$
   - If RANK$(x^*) > i$, then SELECT(i, L) — $T(|L|)$
   - If RANK$(x^*) < i$, then SELECT$(i - |L| - 1, G)$ — $T(|G|)$

Analysis

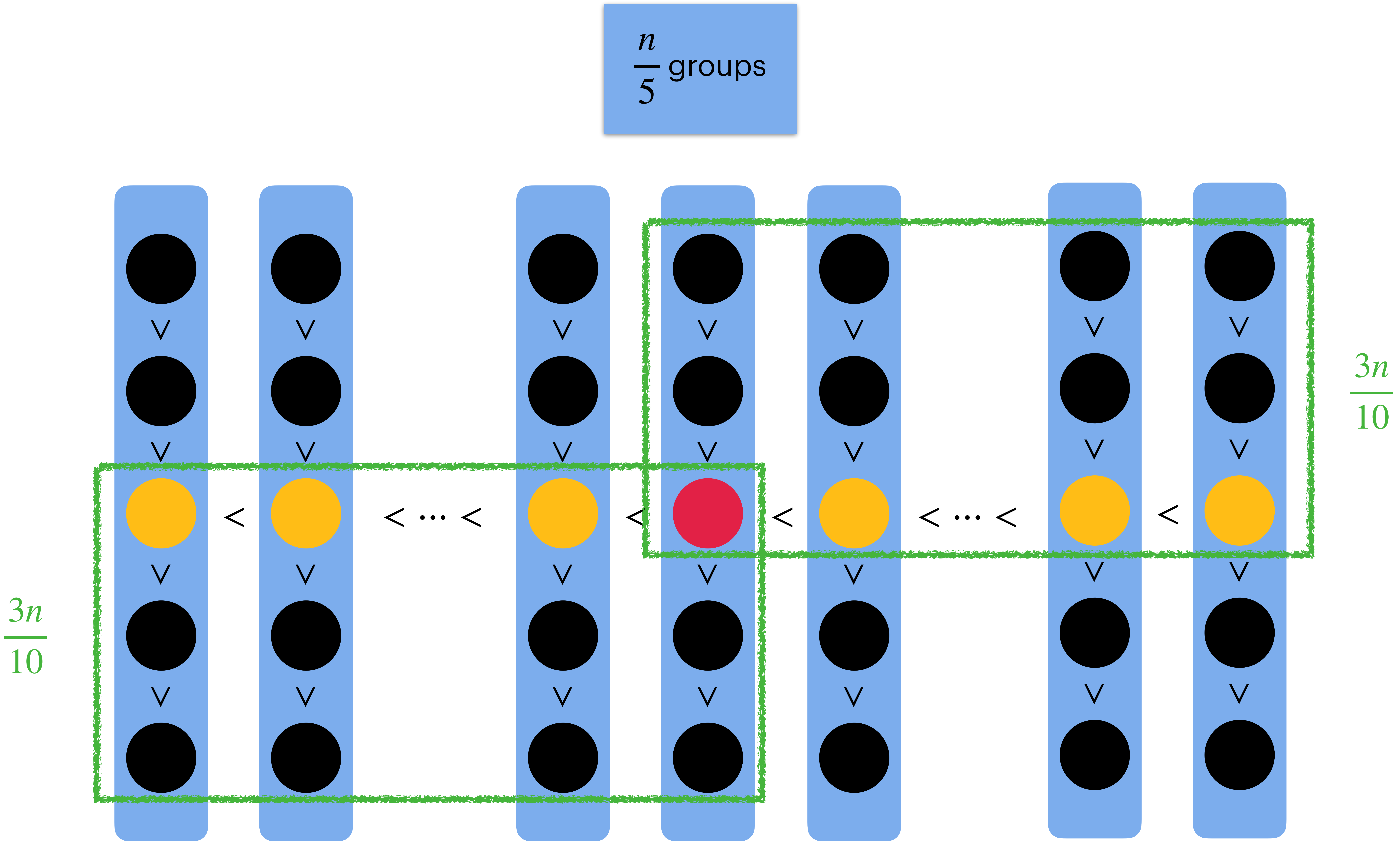
Claim: x^* is $\frac{7}{10}$-balanced, that is $\max\{|L|, |G|\} \leq \frac{7n}{10}$.


$\frac{n}{5}$ groups


$\frac{n}{5}$ groups
$\frac{n}{10}$



n/5 groups
n/10
2n/10



$\frac{n}{5}$ groups
$\frac{3n}{10}$
$\frac{3n}{10}$

Analysis

Claim: x^* is $\frac{7}{10}$-balanced, that is $\max\{|L|, |G|\} \le \frac{7n}{10}$.

$$|L| + 1 \ge \frac{3n}{10} \implies |G| \le \frac{7n}{10}$$

$$|G| + 1 \ge \frac{3n}{10} \implies |L| \le \frac{7n}{10}$$

Analysis

Recurrence: $T(n) = T\left(\frac{n}{5}\right) + T\left(\frac{7n}{10}\right) + O(n)$

- What if we had chosen groups of $\frac{n}{3}$ elements?
- How about groups of $\frac{n}{7}$ elements?

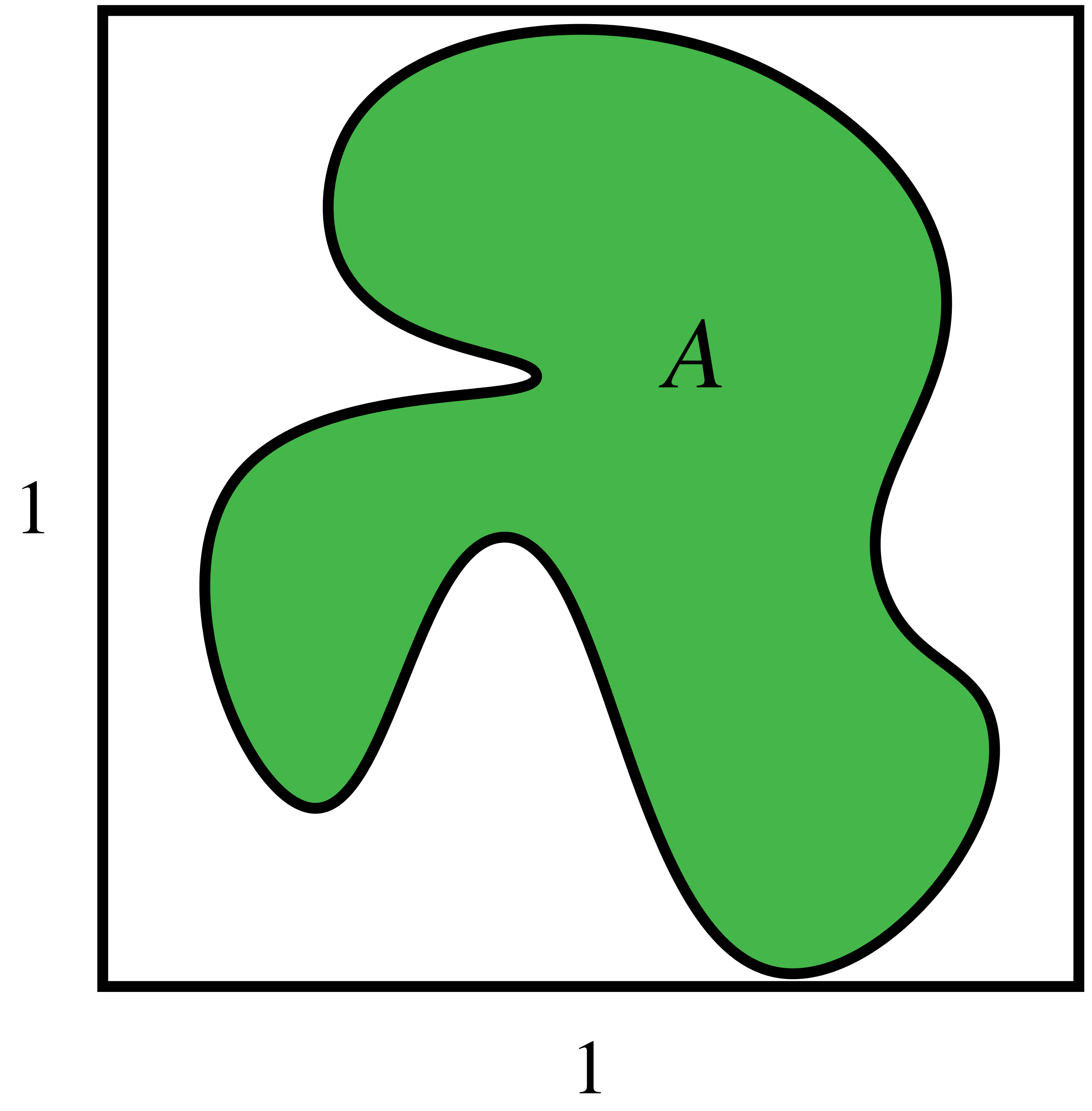
Randomized Algorithms

- Execution not deterministic, but depends on random choices.
- Two types:
  - Las Vegas: Always correct output, typically good runtime
  - Monte Carlo: Always good runtime, typically correct output
- Today: Monte Carlo

Text on randomized algorithms: https://www.cs.ubc.ca/~nickhar/Book1.pdf

Estimating Area

Give a randomized algorithm to estimate A.



Estimating Area

Via Sampling

Idea: The probability of a random sample landing in the shape is A, so sample n random points and compute the proportion $\hat{A}$ of points in the shape.

Let $X_i = 1$ if point i is in the shape and 0, otherwise.

Our estimate: $\hat{A} = \dfrac{X}{n}$ where $X = \sum_{i=1}^{n} X_i$

Claim: $\hat{A}$ is an unbiased estimator, i.e., $\mathbb{E}[\hat{A}] = A$.

Proof. On the board.

What are possible concerns?

How often is $\hat{A}$ actually close to A?

Estimating Area

Concentration

Suppose $A > \frac{1}{10}$ is given. Provide a bound on the number of samples required to ensure the estimate $\hat{A}$ is within $(1 \pm \epsilon)A$ with probability $1 - \delta$ for $0 < \epsilon, \delta < 1$.

On the board.