Order Statistics

Order Statistics

- Given a set of n numbers $\{a_1,\ldots,a_n\}$ and an integer $i \in \{1,\ldots,n\}$. Find out the i-th smallest number in $\{a_1,\ldots,a_n\}$.
- Input: $\{a_1,\ldots,a_n\}$ and i
- Output: the i-th smallest number
- Sample Input: {0,3,5,7,1,2,1,2,1}, i=5
- Sample Output: 2

Order Statistics

- Static version: the set will not change
  - Special case: $\Theta(n)$ if $i=1$ or $i=n$
  - Randomized partition: $\Theta(n)$ (Expected)
  - Deterministic divide and conquer: $\Theta(n)$
- Dynamic: the set will change
  - Using Augmented BST (see Chap. 14)
    - Insert / Delete / Select in $O(\log n)$

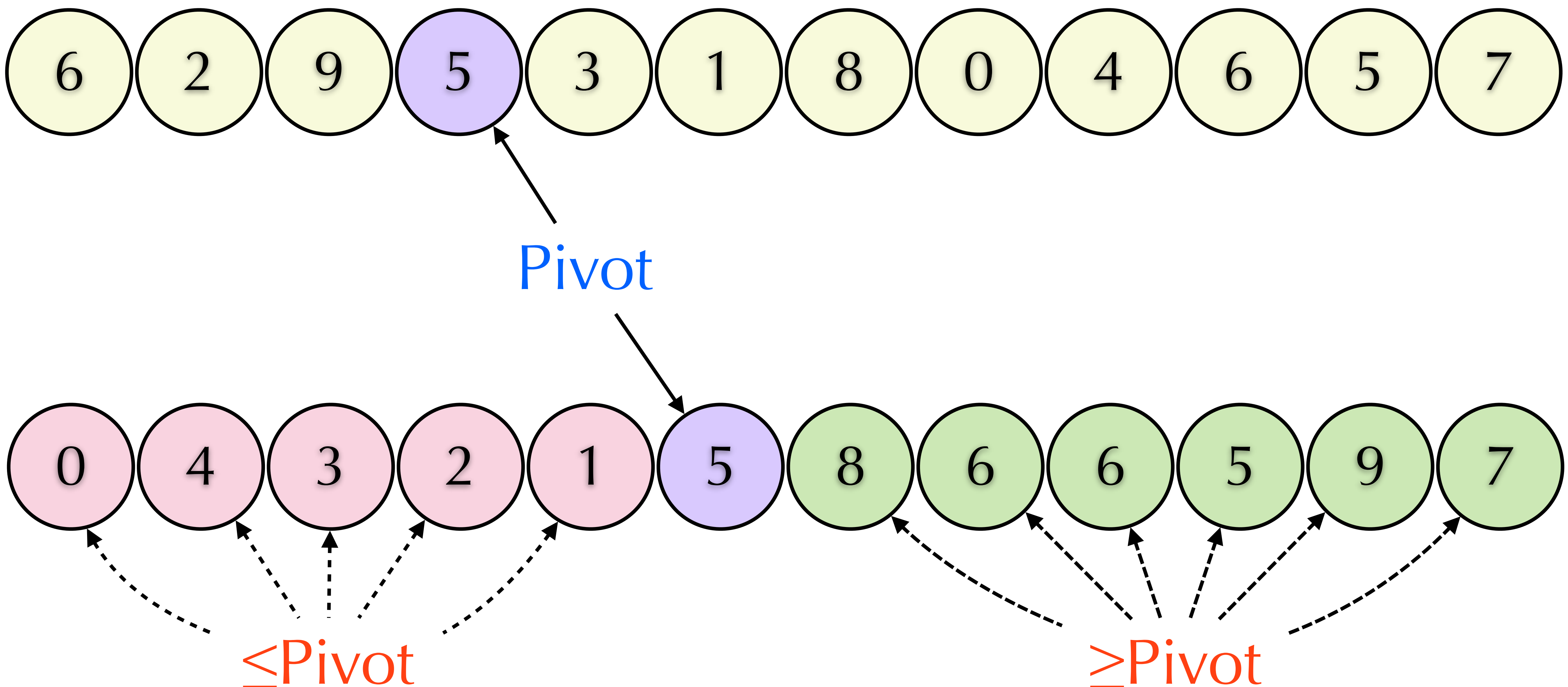
Special Case

- Finding the minimum (i=1)
  - ```
    minIndex=1
    for j = 2 to n do
        if a_minIndex>a_j then minIndex=j
    ```
    $a_{minIndex} > a_j$
- Finding the maximum (i=n)
  - ```
    maxIndex=1
    for j = 2 to n do
        if a_maxIndex<a_j then maxIndex=j
    ```
    $a_{maxIndex} < a_j$

Order Statistic: Randomized Partition

- Recall the partition method in quick sort
  - Input: A[1..n]
  - Reorder A and find m such that
    - For $1 \leq j < m < k \leq m$, $A[j] \leq A[m] \leq A[k]$.
- If m=i, then we are done.
- If m<i, then compute the (i−m)-th smallest element in A[m+1..n].
- If m>i, then compute the i-th smallest element in A[1..m−1].

Randomized Partition



Time Complexity: Worst Case

- $T(n)=\Theta(n)$ ... $i=m$
- $T(n)=\max_{1\leq m\leq n}T(m-1)+\Theta(n)$ ... $i<m$
- $T(n)=\max_{1\leq m\leq n}T(n-m)+\Theta(n)$ ... $i>m$
- Goal: $T(n)=\Theta(n^2)$
- Lower bound: $\Omega(n^2)$
  - $T(n)\geq T(n-1)+\Theta(n)$ .... $m=1<i=n$
    $\geq T(n-2)+\Theta(n-1)+\Theta(n)$
    $\geq\ldots$
    $\geq\Theta(1+2+\ldots+n)=\Theta(n^2)$

Time Complexity: Worst Case

- Upper bound: $O(n^2)$
  - Guess: $T(n) \leq cn^2$
  - $T(n) \leq \max(T(n-m), T(m-1)) + \Theta(n)$
    $\leq c(n-m)^2 + c(m-1)^2 + \Theta(n)$
    $\leq c(n-1)^2 + \Theta(n)$
    $\leq cn^2 - 2cn + c + c'n$
    $= cn^2 - (2cn - c - c'n)$
    $\leq cn^2$ ... picking $c \geq c'$ and $n \geq 1$

Time Complexity: Average Case

- $T(n)=\Theta(n)$ ... $i=m$
- $T(n)=\max_{1\leq m\leq n} T(m-1)+\Theta(n)$ ... $i<m$
- $T(n)=\max_{1\leq m\leq n} T(n-m)+\Theta(n)$ ... $i>m$
- Goal: $E[T(n)]=\Theta(n)$
- Lower bound: $\Omega(n)$
  - $E[T(n)]\geq E[\Theta(n)]=\Theta(n)$

Time Complexity: Average Case

- Upper bound: $O(n^2)$
- Guess: $E[T(n)] \le cn$
- $E[T(n)] \le E[\max(T(n-m), T(m-1))] + \Theta(n)$
  $\le c'n + E[T(\max(n-m, m-1))]$
  $= c'n + \Sigma_{1 \le k \le n} \Pr[m=k] T(\max(n-k, k-1))$
  $\le c'n + (c/n) \Sigma_{1 \le k \le n} \max(n-k, k-1)$
  $\le c'n + (2c/n) \Sigma_{\lfloor n/2 \rfloor \le k \le n} k$

Note: $T(n)$ is monotonic!

Time Complexity: Average Case

- E[T(n)]
  $\leq c'n+(2c/n)\Sigma_{\lfloor n/2\rfloor\leq k\leq n}k$
  $\leq c'n+(3/4)cn+c$
  $\leq cn-(cn/4-c-c'n)$ pick c,n, s.t. $c\geq 4c/n+4c'$
  $\leq cn$

$$\sum_{k=\lfloor\frac{n}{2}\rfloor}^{n} k = \left\lceil\frac{n}{2}\right\rceil \frac{\lfloor\frac{n}{2}\rfloor + n}{2} \leq \frac{3n(n+1)}{8} \leq \frac{3n^2}{8} + \frac{n}{2}$$

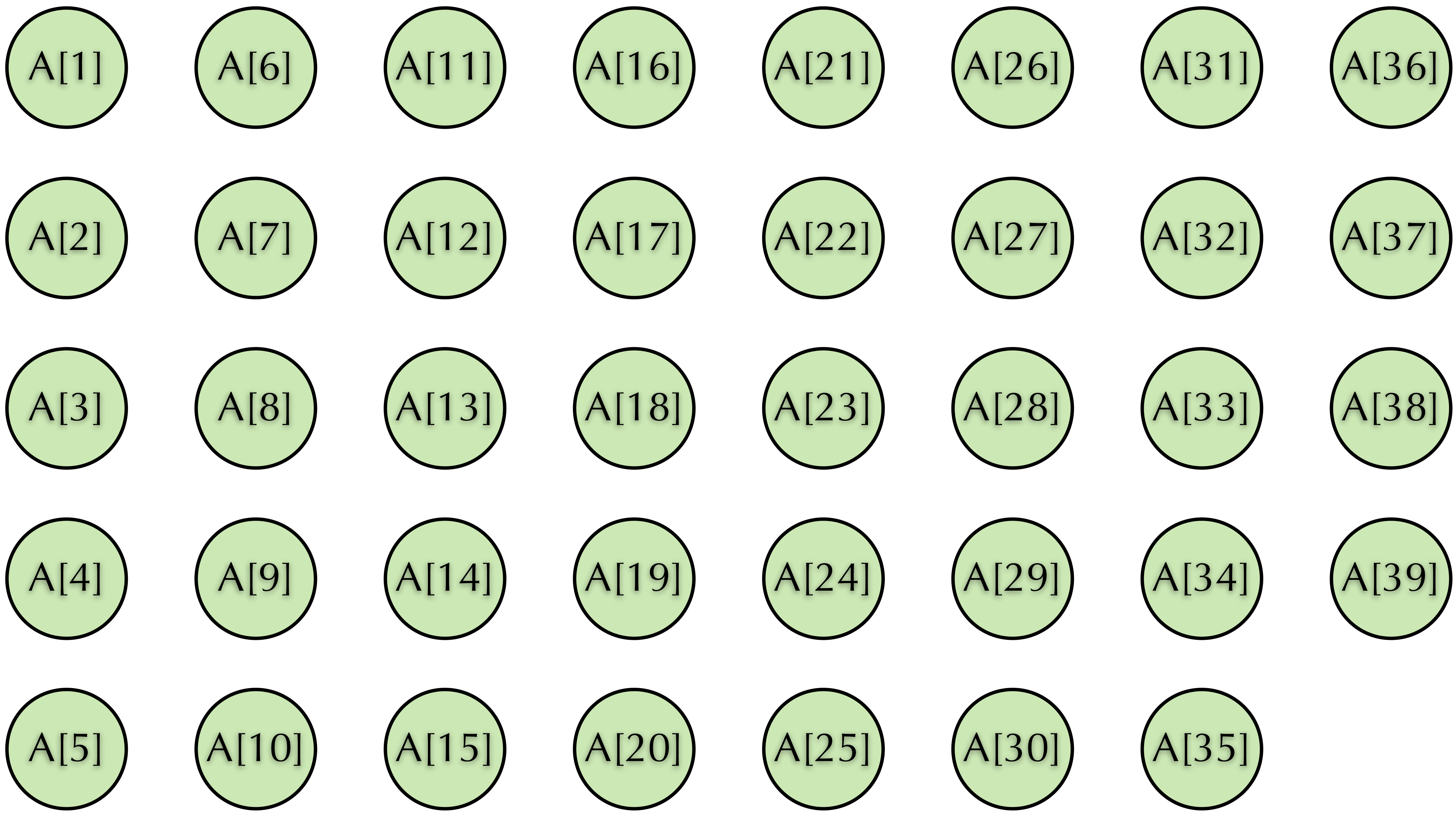
Divide and Conquer

- Termination: sort the input and output the i-th smallest element if $n \le 60$
- Divide Phase 1:
  - Divide A into $B_1,\ldots,B_{\lceil n/5 \rceil}$
  - $B_j=\{A[k]: 5j-4 \le k \le \min(5j,n)\}$ $\Theta(n)$
  - Compute C[j]: median of B_j $\Theta(n)$???
- Conquer Phase 1:
  - Compute p: median of $C[1..\lceil n/5 \rceil]$ $T(\lceil n/5 \rceil)$

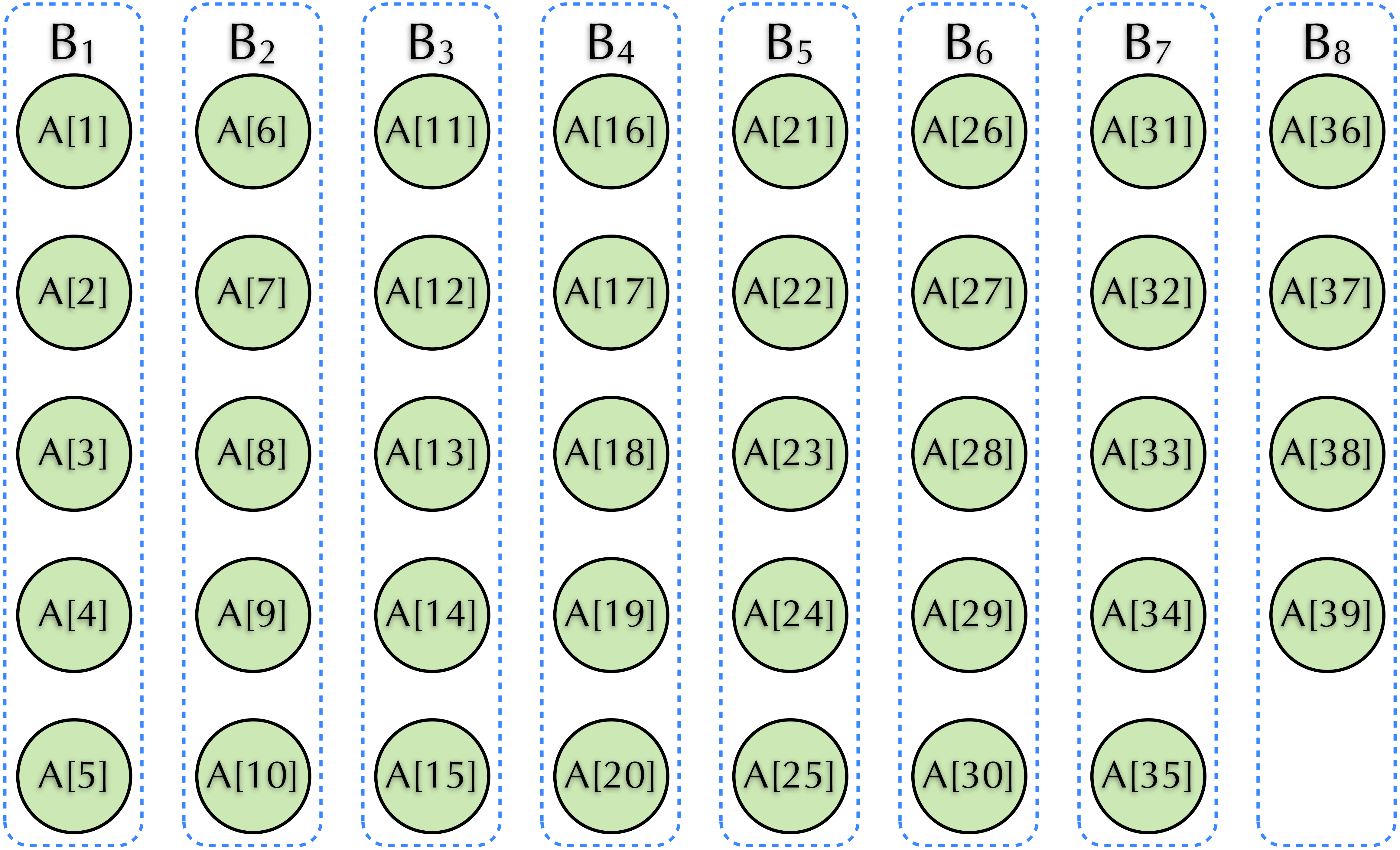
Divide and Conquer

- Divide Phase 2:
  - Compute m by partition with pivot p $\Theta(n)$
- Conquer Phase 2:
  - If $m=i$, then we are done. $\Theta(1)$
  - If $m<i$, then compute the $(i-m)$-th smallest element in $A[m+1..n]$. $T(7n/10+2)$?
  - If $m>i$, then compute the i-th smallest element in $A[1..m-1]$. $T(7n/10)$?
- Combine: The result of previous step

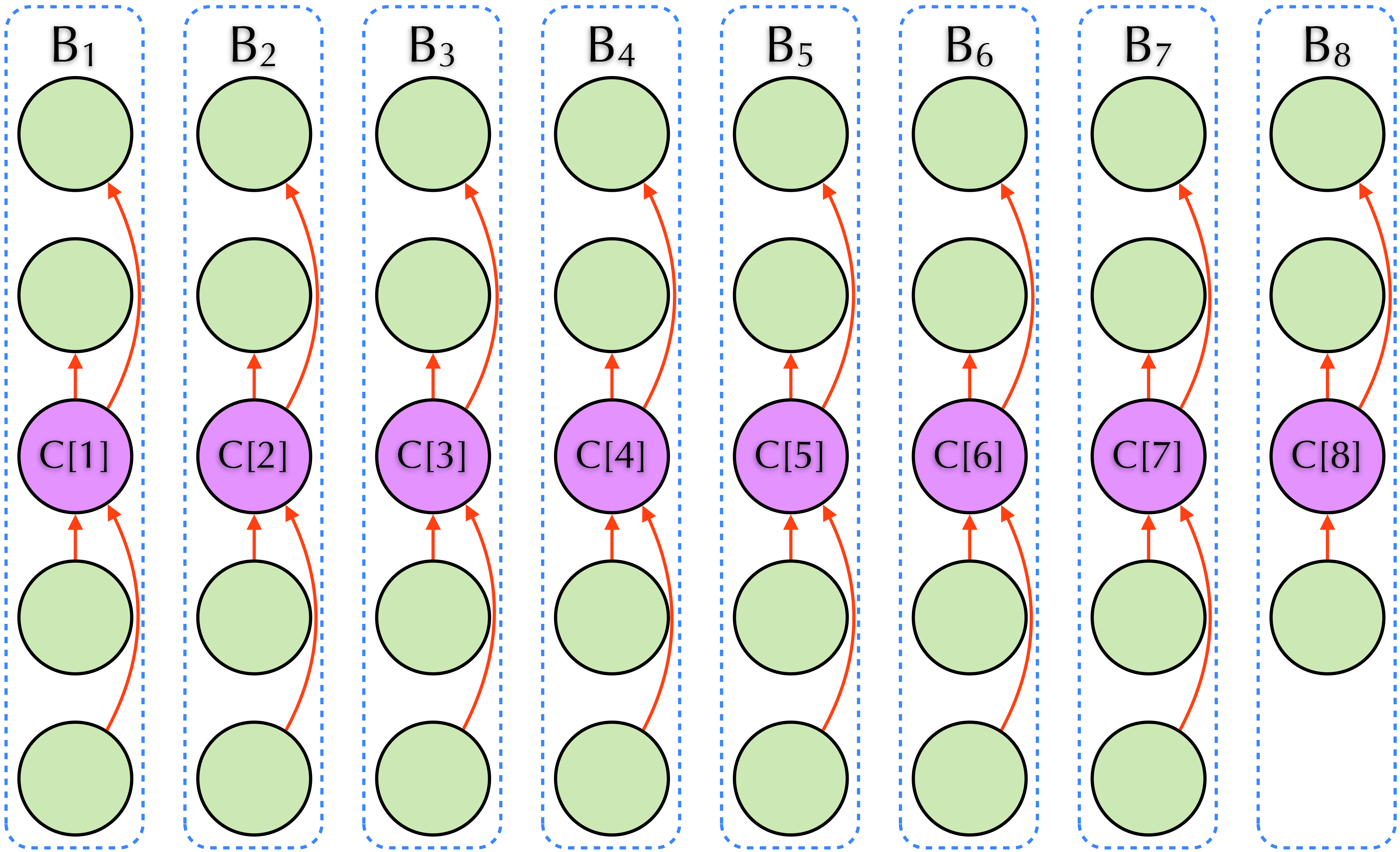
Divide: Phase 1



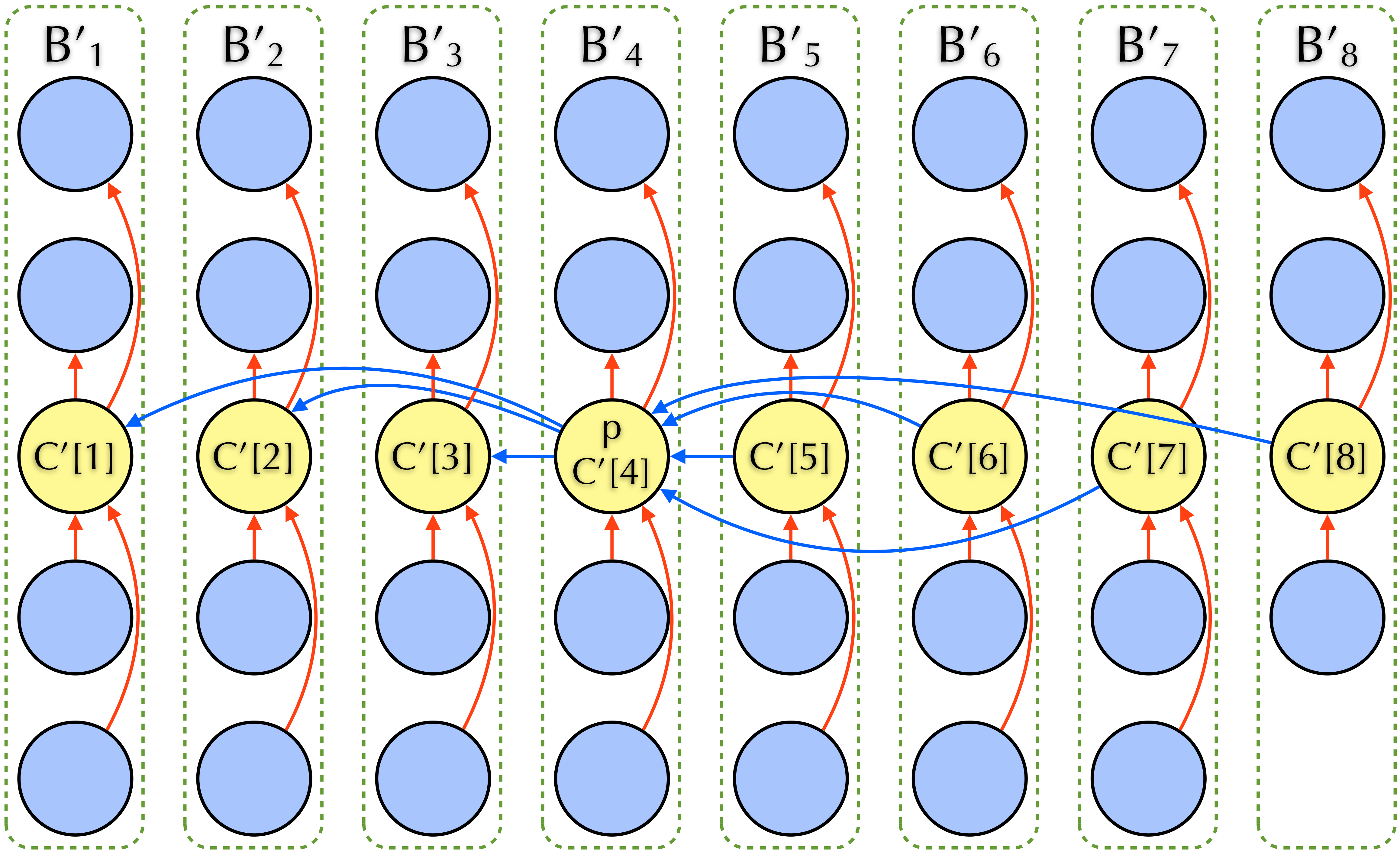
Divide: Phase 1



Divide: Phase 1



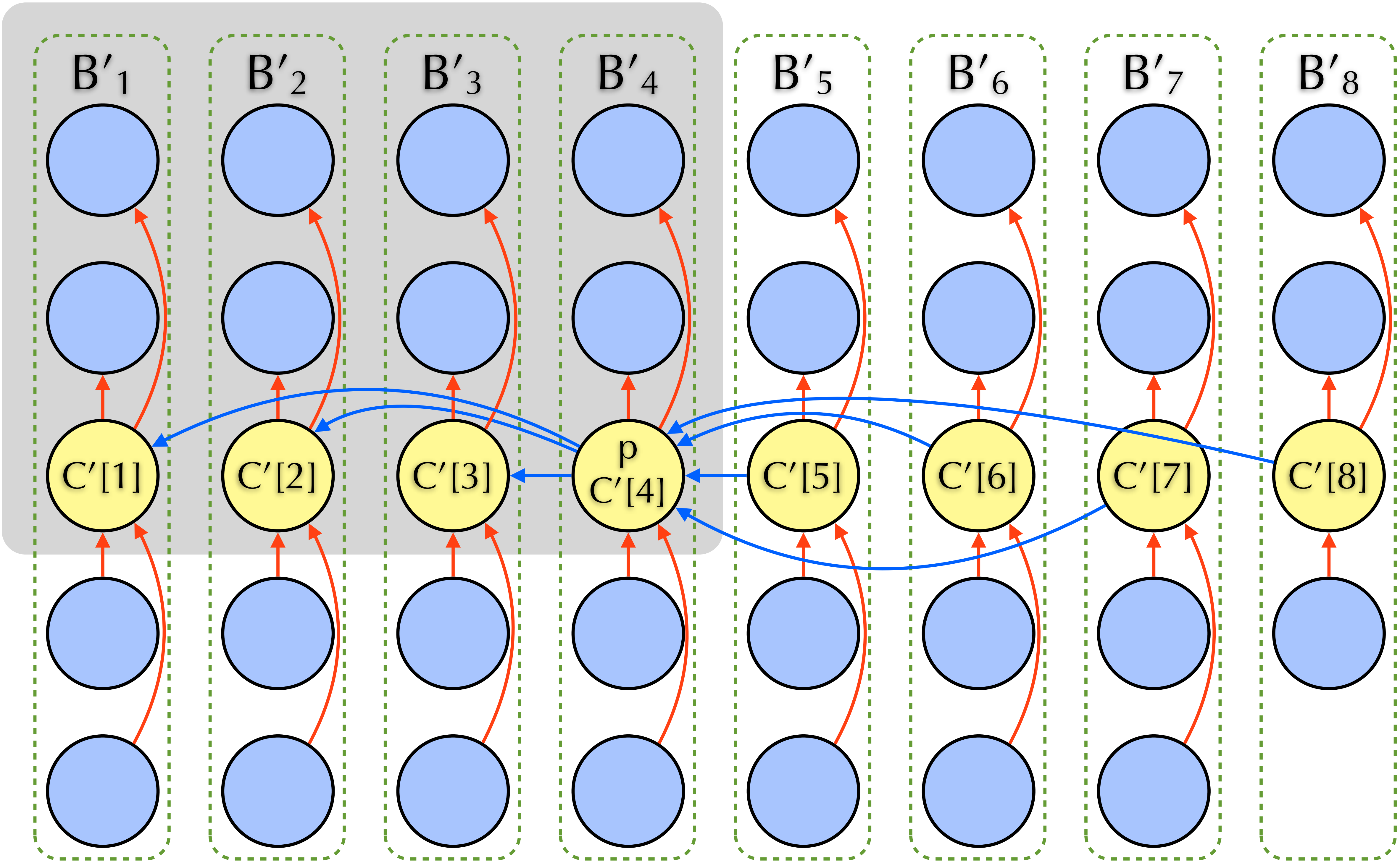
Conquer: Phase 1



Numbers ≤ p

$$m \geq 3\left\lceil \frac{\lceil \frac{n}{5} \rceil}{2} \right\rceil \geq \frac{3n}{10}$$

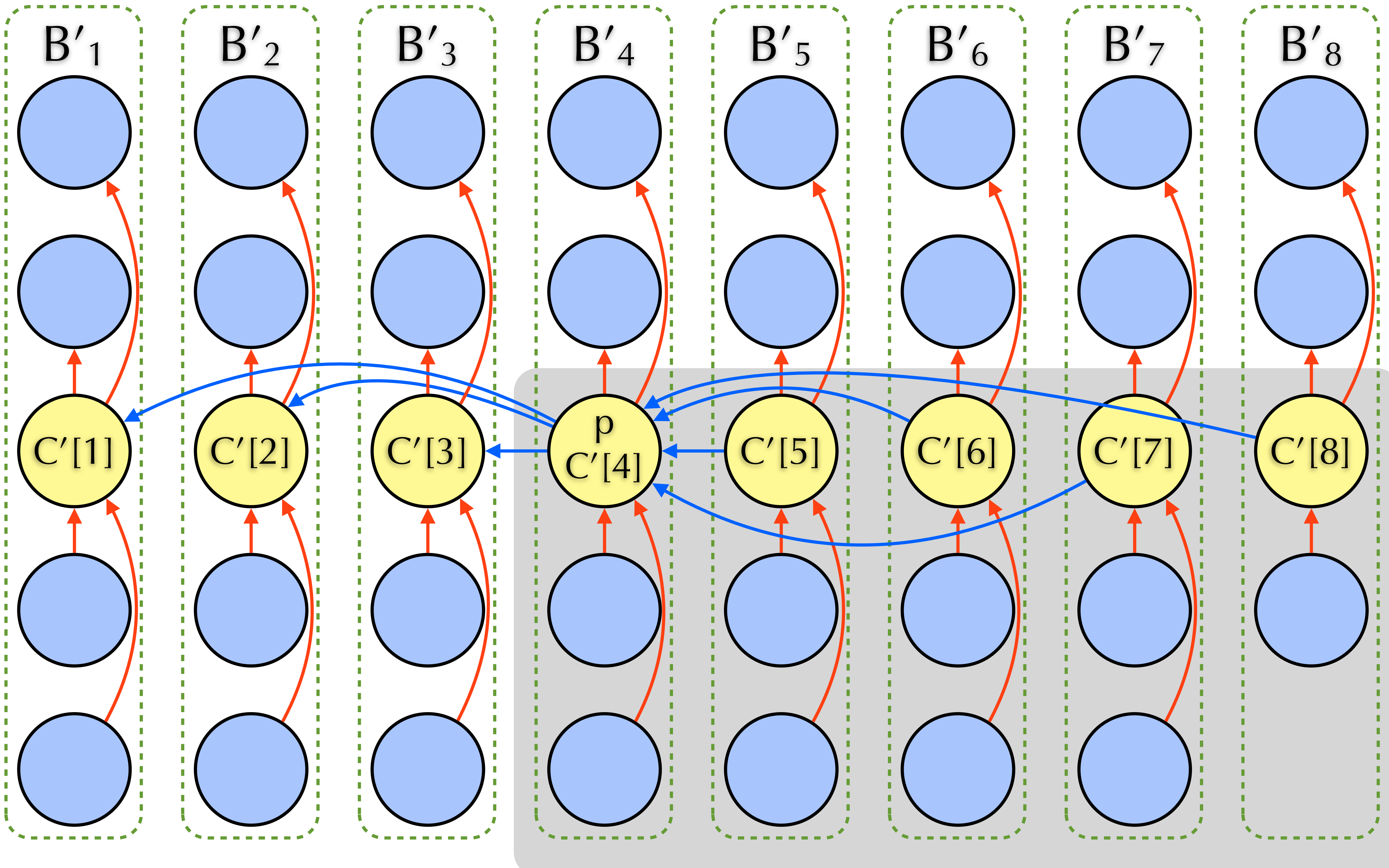
$$n - m \leq n - \frac{3n}{10} \leq \frac{7n}{10}$$



Numbers ≥ p

$$n - m + 1 \geq 3\left\lceil \frac{\left\lceil \frac{n}{5} \right\rceil}{2} \right\rceil - 2 \geq \frac{3n}{10} - 2$$

$$m - 1 \leq n - \frac{3n}{10} + 2 \leq \frac{7n}{10} + 2$$



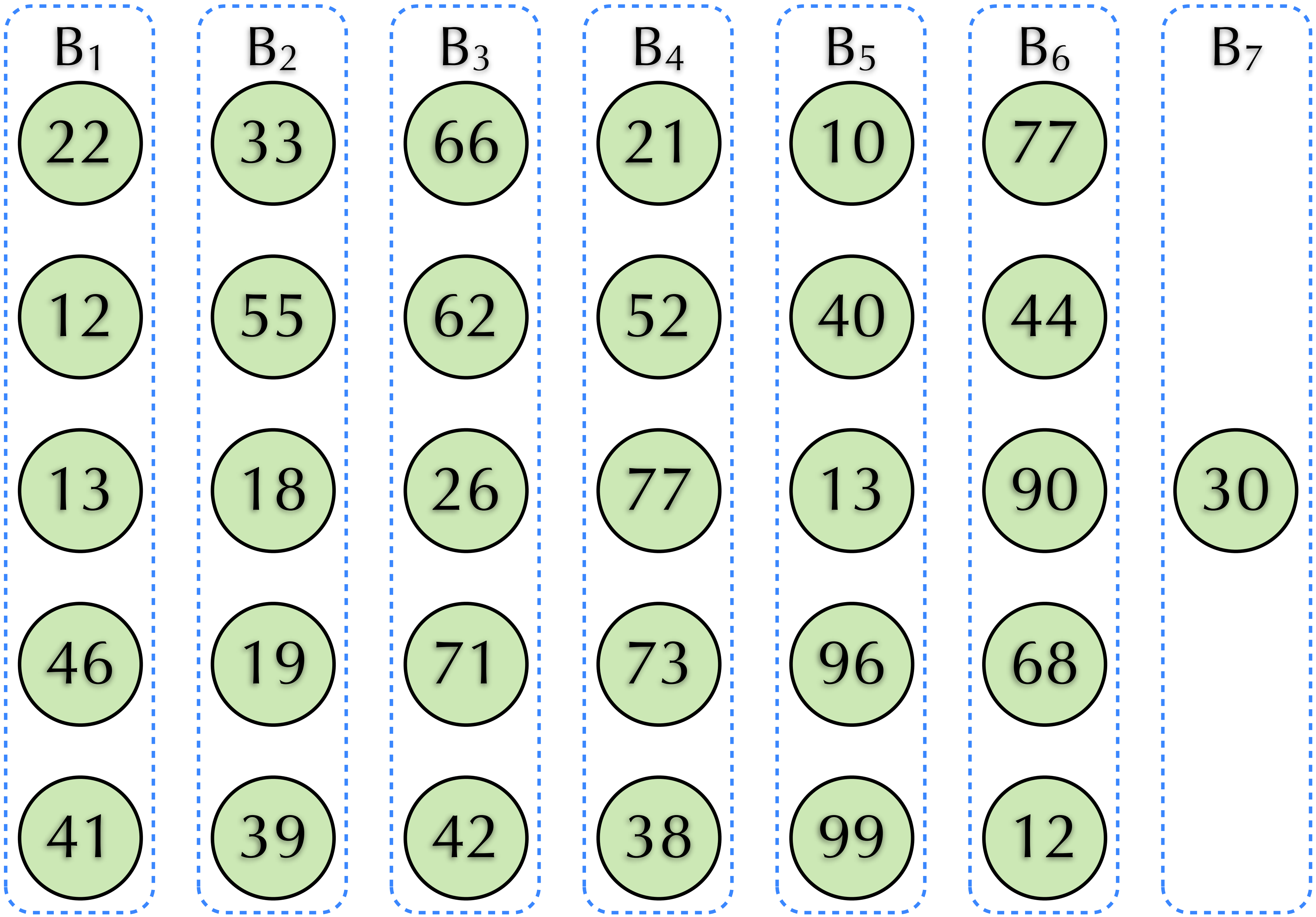
The analysis is different from the one in the textbook!

Time Complexity

- Goal: $T(n)=T(\lceil n/5 \rceil)+T(k)+\Theta(n)=\Theta(n)$ where k is the size of subproblem 2.
- Lower bound: trivial
- Upper bound: Assume $T(n) \le cn$
  - $T(n) \le c\lceil n/5 \rceil + c(7n/10+2)+c'n$
    $\le c(9n/10)+3c+c'n$
    $=cn-(cn/10-3c-c'n)$
    $\le cn$ ... $c \ge 30c/n+10c' \ge c/2+10c'$ $(n \ge 60)$

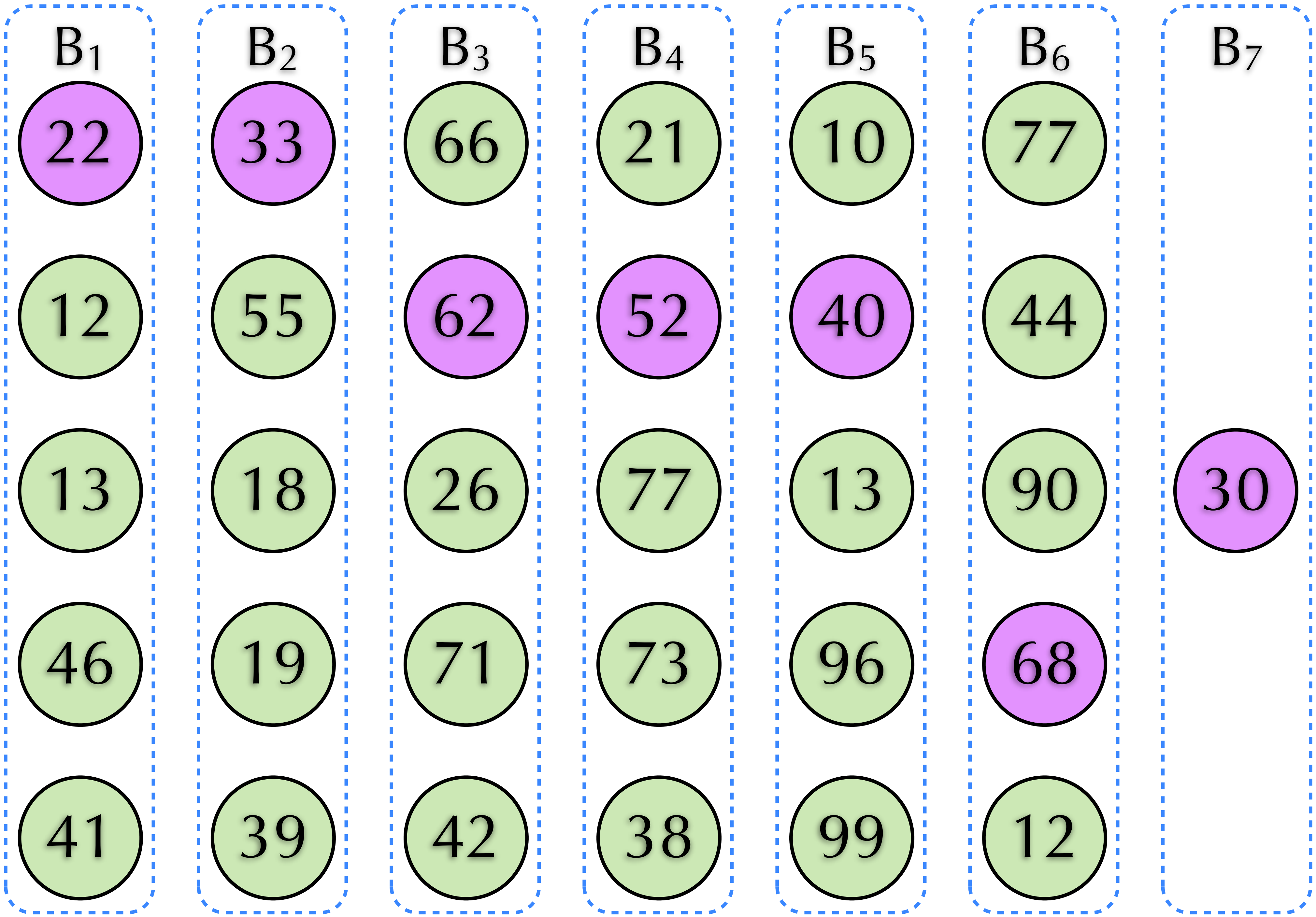
Note: T(n) is monotonic!

Example


B1
22
12
13
46
41
B2
33
55
18
19
39
B3
66
62
26
71
42
B4
21
52
77
73
38
B5
10
40
13
96
99
B6
77
44
90
68
12
B7
30

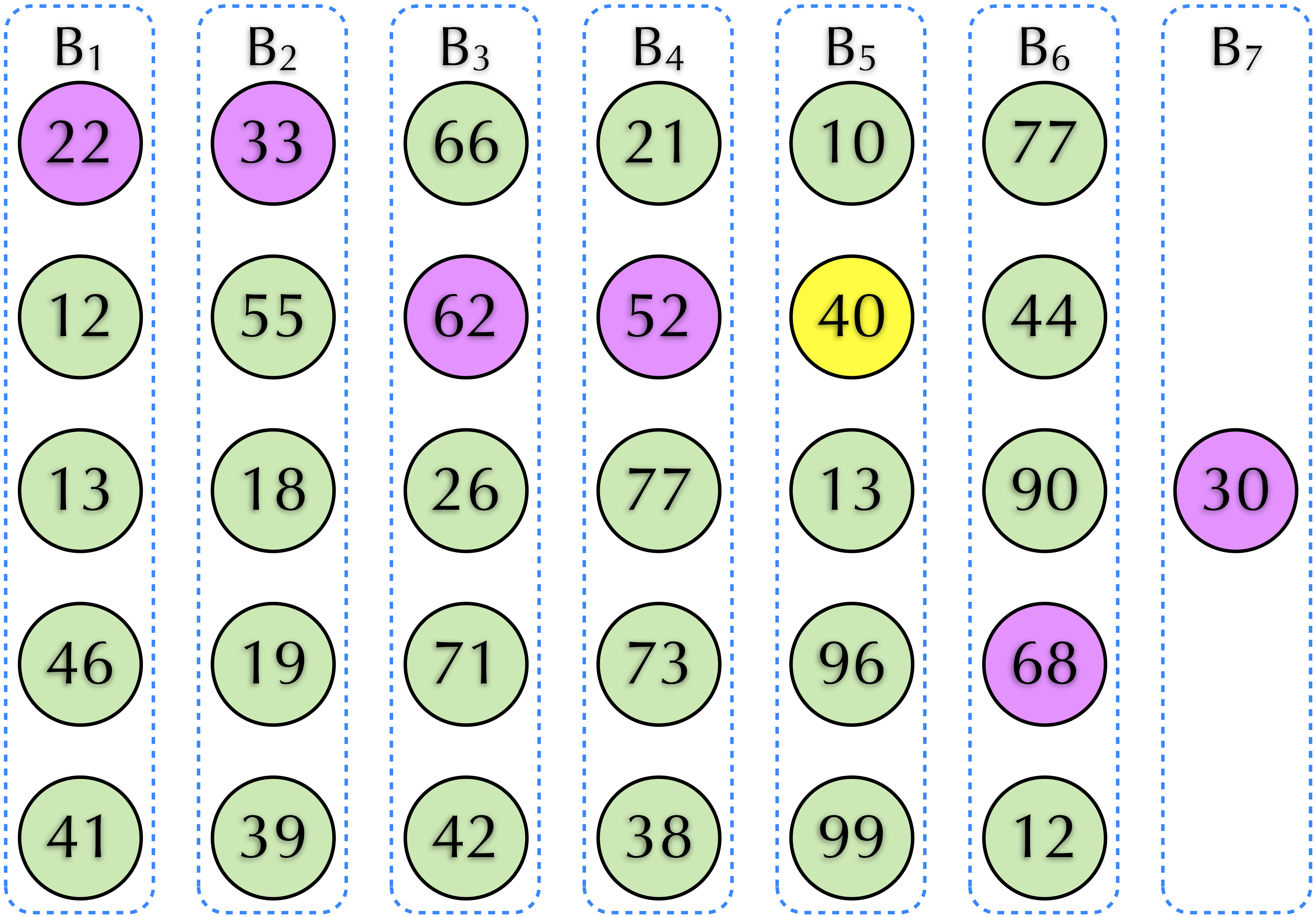
Note: Iteratively in $\Theta(n)$

Compute C[j]

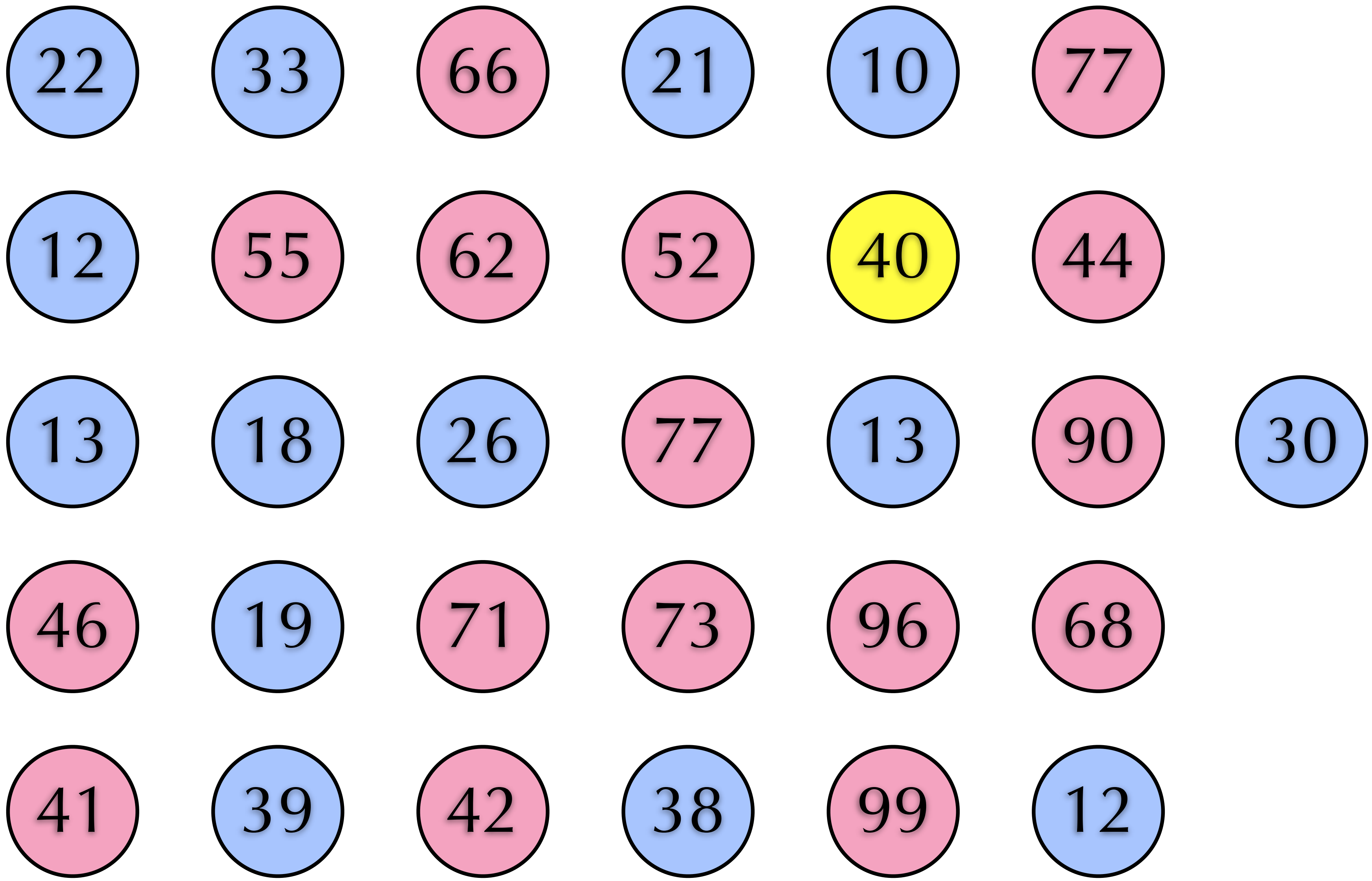


Note: Recursively in T(⌈n/5⌉)

Find Median of C[j]'s



Partition with p



Conquer Subproblem 2

Recursively in $T(7n/10)$ if $i<m$

22 12 13 33 18 19 39 26 21 38 10 13 12 30 40

Done if $i=m$

46 41 55 66 62 71 42 52 77 73 96 99 77 44 90 68

Recursively in $T(7n/10+2)$ if $i>m$