Οι διαφάνειες βασίζονται σε αυτές του ακόλουθου μαθήματος:

Introduction to Algorithms (6-046J), MIT

https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-046j-introduction-to-algorithms-sma-5503-fall-2005/

Διατακτικές Στατιστικές

Επίλεξε το i μικρότερο από n στοιχεία (το στοιχείο με ***βαθμό*** i).

- $i = 1$: ***ελάχιστο***
- $i = n$: ***μέγιστο***
- $i = \lceil (n+1)/2 \rceil$ or $\lfloor (n+1)/2 \rfloor$: ***μεσαίος***.

Απλός αλγόριθμος: Ταξινόμησε και επίστρεψε το στοιχείο στη θέση i

Χρόνος χειρότερης περίπτωσης $= \Theta(n \lg n) + \Theta(1) = \Theta(n \lg n)$, χρησιμοποιώντας συγχωνευτική ταξινόμηση ή συγχώνευση σωρού (όχι ταχυταξινόμηση).

Τυχαιοκρατικός Διαίρει και Βασίλευε Αλγόριθμος

RAND-SELECT(A, p, q, i) $\quad \triangleright$ ith smallest of $A[p \,..\, q]$

  if $p = q$ **then return** $A[p]$

  $r \leftarrow$ RAND-PARTITION(A, p, q)

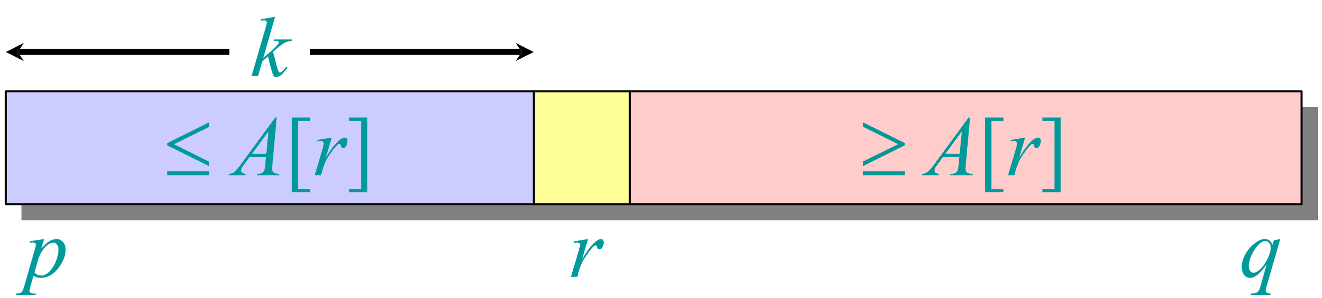
  $k \leftarrow r - p + 1$ $\quad \triangleright$ $k = \text{rank}(A[r])$

  if $i = k$ **then return** $A[r]$

  if $i < k$

    then return RAND-SELECT$(A, p, r - 1, i)$

    else return RAND-SELECT$(A, r + 1, q, i - k)$



Παράδειγμα

Επίλεξε το $i = 7$ μικρότερο στοιχείο:

6	10	13	5	8	3	2	11

$i = 7$

pivot

Διαμέριση:

2	5	3	6	8	13	10	11

$k = 4$

Επίλεξε το $7 - 4 = 3$ μικρότερο στοιχείο αναδρομικά.

Διαίσθηση για την ανάλυση

(Η ανάλυση υποθέτει ότι όλα τα στοιχεία είναι διαφορετικα.)

Τυχεροί:

$T(n) = T(9n/10) + \Theta(n)$ $\quad n^{\log_{10/9} 1} = n^0 = 1$

$= \Theta(n)$ $\quad$ Περίπτωση 3

Άτυχοι:

$T(n) = T(n-1) + \Theta(n)$ $\quad$ Αριθμητική

$= \Theta(n^2)$ $\quad$ πρόοδος

Χειρότερα από ταξινόμηση!

Ανάλυση αναμενόμενου χρόνου

Η ανάλυση ακολουθεί αυτή της τυχαιοκρατικής ταξινόμησης.

Έστω $T(n) =$ η τυχαία μεταβλητή για το χρόνο εκτέλεσης της Rand-Select σε μία είσοδο μεγέθους n, υποθέτοντας ότι οι τυχαίοι αριθμοί είναι ανεξάρτητοι.

Για $k = 0, 1, \ldots, n–1$, ορίζουμε τη ***δείκτρια τυχαία μεταβλητή***

$$X_k = \begin{cases} 1 & \text{αν η Partition δημιουργεί μία } k : n-k-1 \text{ διαμέριση,} \\ 0 & \text{αλλιώς.} \end{cases}$$

Ανάλυση (συν.)

Για να λάβουμε ένα πάνω όριο, υποθέτουμε ότι το i-στο στοιχείο πάντα βρίσκεται στο μεγαλύτερο υποπίνακα της διαμέρισης:

$$T(n) = \begin{cases} T(\max\{0, n-1\}) + \Theta(n) & \text{αν } 0 : n-1 \text{ δμρση,} \\ T(\max\{1, n-2\}) + \Theta(n) & \text{αν } 1 : n-2 \text{ δμρση,} \\ \quad\vdots & \\ T(\max\{n-1, 0\}) + \Theta(n) & \text{αν } n-1 : 0 \text{ δμρση,} \end{cases}$$

$$= \sum_{k=0}^{n-1} X_k \big(T(\max\{k, n-k-1\}) + \Theta(n) \big).$$

Υπολογισμός αναμενόμενης τιμής

$$E[T(n)] = E\left[\sum_{k=0}^{n-1} X_k \left(T(\max\{k, n-k-1\}) + \Theta(n)\right)\right]$$

$$= \sum_{k=0}^{n-1} E\left[X_k \left(T(\max\{k, n-k-1\}) + \Theta(n)\right)\right]$$

$$= \sum_{k=0}^{n-1} E[X_k] \cdot E[T(\max\{k, n-k-1\}) + \Theta(n)]$$

$$= \frac{1}{n} \sum_{k=0}^{n-1} E[T(\max\{k, n-k-1\})] + \frac{1}{n} \sum_{k=0}^{n-1} \Theta(n)$$

$$\leq \frac{2}{n} \sum_{k=\lfloor n/2 \rfloor}^{n-1} E[T(k)] + \Theta(n)$$

Οι πάνω όροι εμφανίζονται δύο φορές.

Δύσκολη αναδρομή

$$E[T(n)] = \frac{2}{n} \sum_{k=\lfloor n/2 \rfloor}^{n-1} E[T(k)] + \Theta(n)$$

■

Απόδειξη: $E[T(n)] \le cn$ για μία σταθερά $c > 0$.

- Η σταθερά c μπορεί να επιλεχθεί μεγάλη αρκετά προκειμένου $E[T(n)] \le cn$ και για τις εναρκτήριες περιπτώσεις.

Ισχύει:

$$\sum_{k=\lfloor n/2 \rfloor}^{n-1} k \le \frac{3}{8} n^2$$

Μέθοδος αντικατάστασης

$$
\begin{aligned}
E[T(n)] &\le \frac{2}{n} \sum_{k=\lfloor n/2 \rfloor}^{n-1} ck + \Theta(n) \\
&\le \frac{2c}{n}\left(\frac{3}{8}n^2\right) + \Theta(n) \\
&= cn - \left(\frac{cn}{4} - \Theta(n)\right) \\
&\le cn,
\end{aligned}
$$

αν το c επιλεχθεί αρκετά μεγάλο ώστε το $cn/4$ να κυριαρχεί του $\Theta(n)$.

Περίληψη της τυχαιοκρατικής επιλογής

- Εκτελείται γρήγορα: γραμμικός αναμενόμενος χρόνος.
- Εξαιρετικός αλγόριθμος στην πράξη.
- Αλλά, στη χειρότερη περίπτωση είναι ***πολύ*** κακός: $\Theta(n^2)$.

$\blacksquare$

Ε. Υπάρχει αλγόριθμος που τρέχει σε γραμμικό χρόνο στη χειρότερη περίπτωση;

Α. Ναι, λόγω των Blum, Floyd, Pratt, Rivest, and Tarjan [1973].

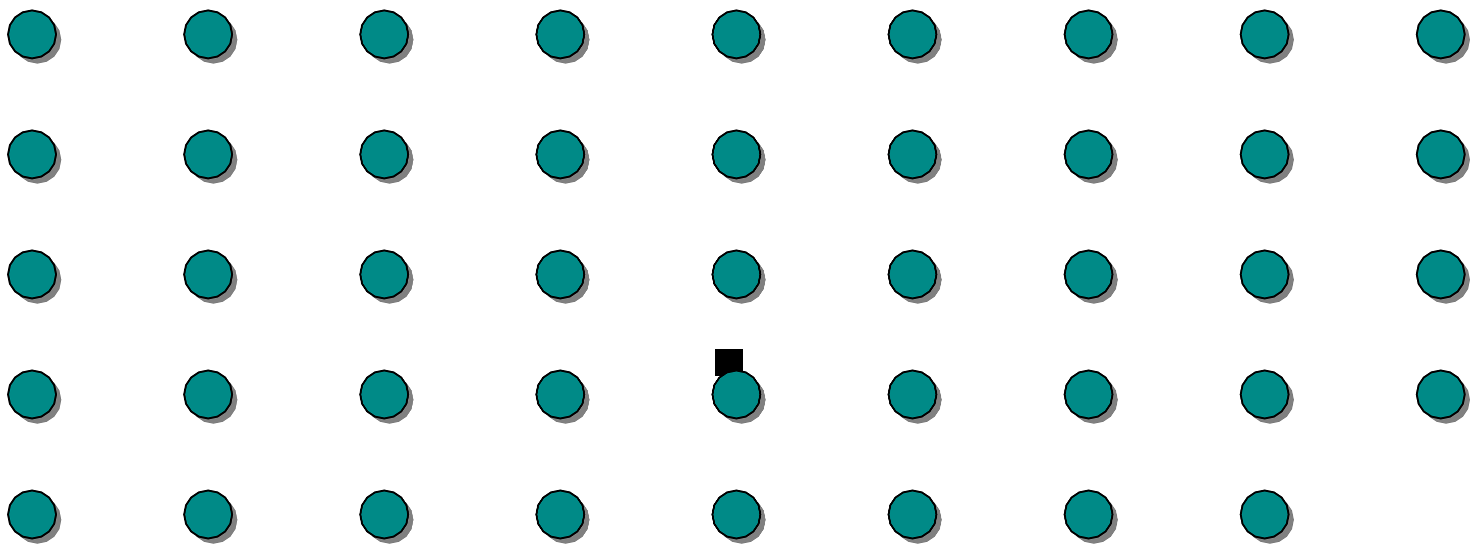
Ιδέα: Δημιούργησε ένα καλό οδηγό αναδρομικά.

Επιλογή με γραμμικό χρόνο χειρότερης περίπτωσης

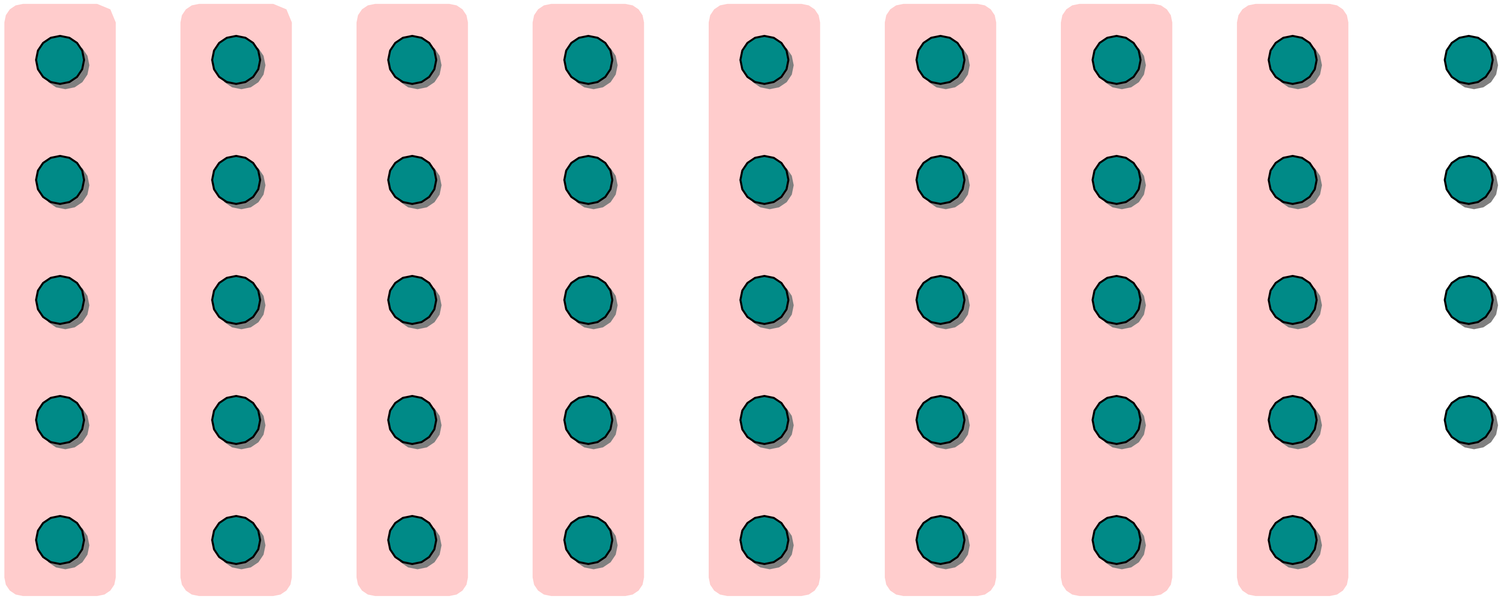
SELECT(i, n)

1. Διαίρεσε τα n στοιχεία σε ομάδες των 5. Βρες το μέσο καθεμίας ομάδας με ενθετική ταξινόμηση.
2. Βρες το μέσο x των $\lfloor n/5 \rfloor$ μέσων των ομάδων με αναδρομική εκτέλεση της SELECT
3. Διαμέρισε γύρω από τον οδηγό x. Έστω $k = rank(x)$
4. **if** $i = k$ **then return** x
   elseif $i < k$
   then αναδρομικά βρες το i-μικρότερο στοιχείο στο χαμηλότερο υποπίνακα
   else αναδρομικά βρες το $(i–k)$- μικρότερο στοιχείο στον υψηλότερο υποπίνακα

Επιλογή του οδηγού

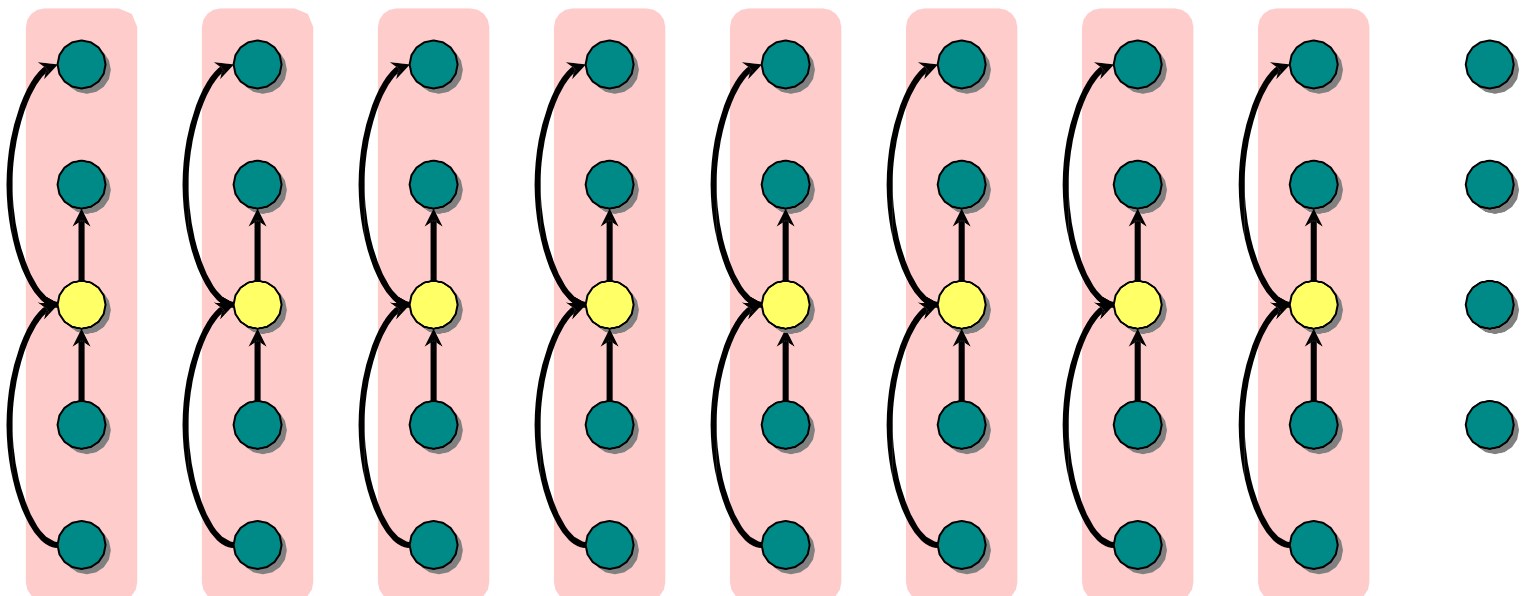


Επιλογή του οδηγού

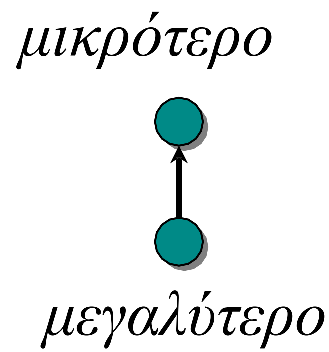


1. Διαίρεσε τα *n* στοιχεία σε ομάδες των 5.

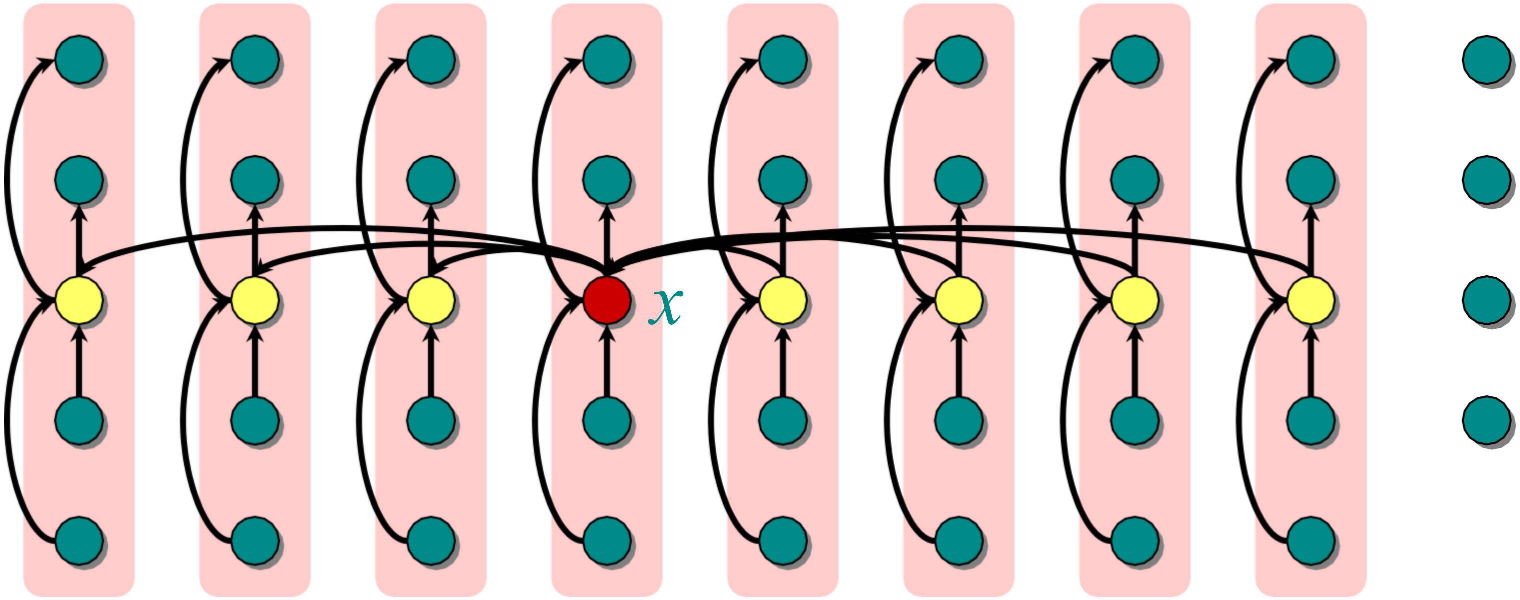
Επιλογή του οδηγού



1. Διαίρεσε τα *n* στοιχεία σε ομάδες των 5. Βρες το μέσο κάθε ομάδας με ενθετική ταξινόμηση.

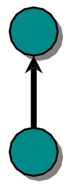


Επιλογή του οδηγού



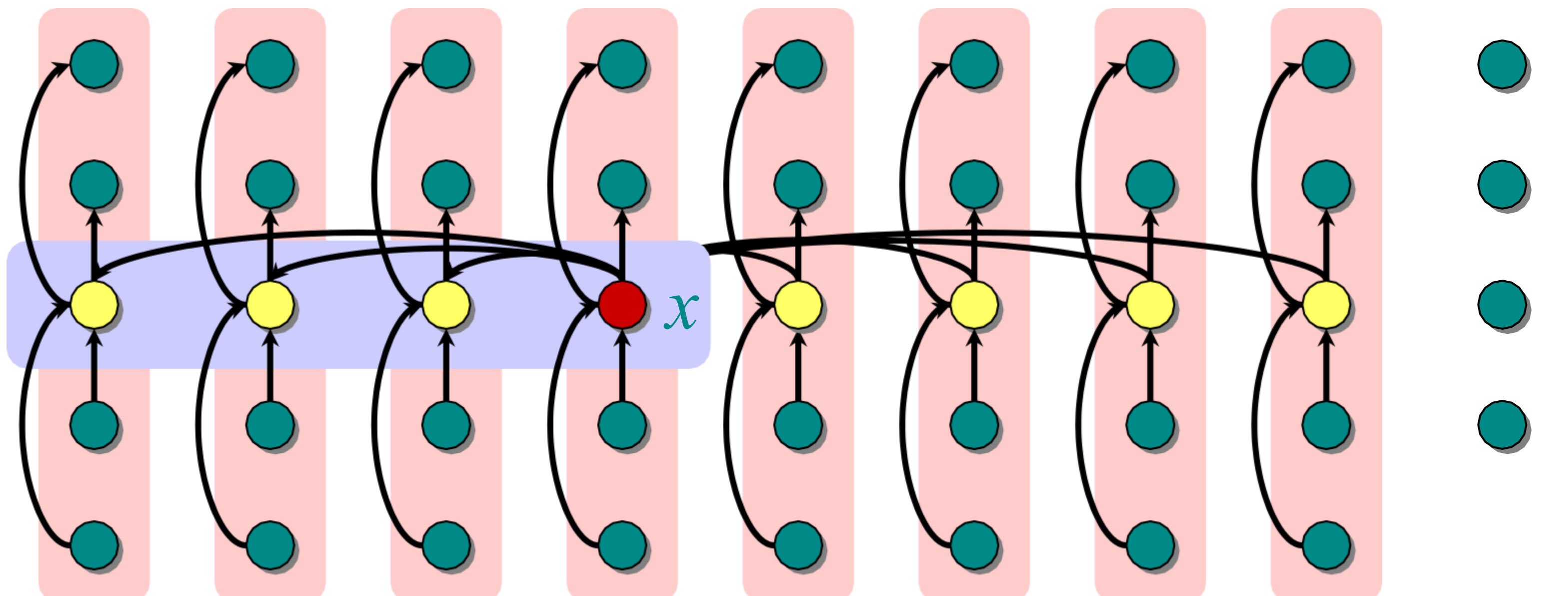
1. Διαίρεσε τα n στοιχεία σε ομάδες των 5. Βρες το μέσο κάθε ομάδας με ενθετική ταξινόμηση.
2. Βρες το μέσο x των $\lfloor n/5 \rfloor$ μέσων των ομάδων με αναδρομική εκτέλεση της SELECT

μικρότερο



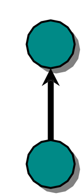
μεγαλύτερο

Ανάλυση



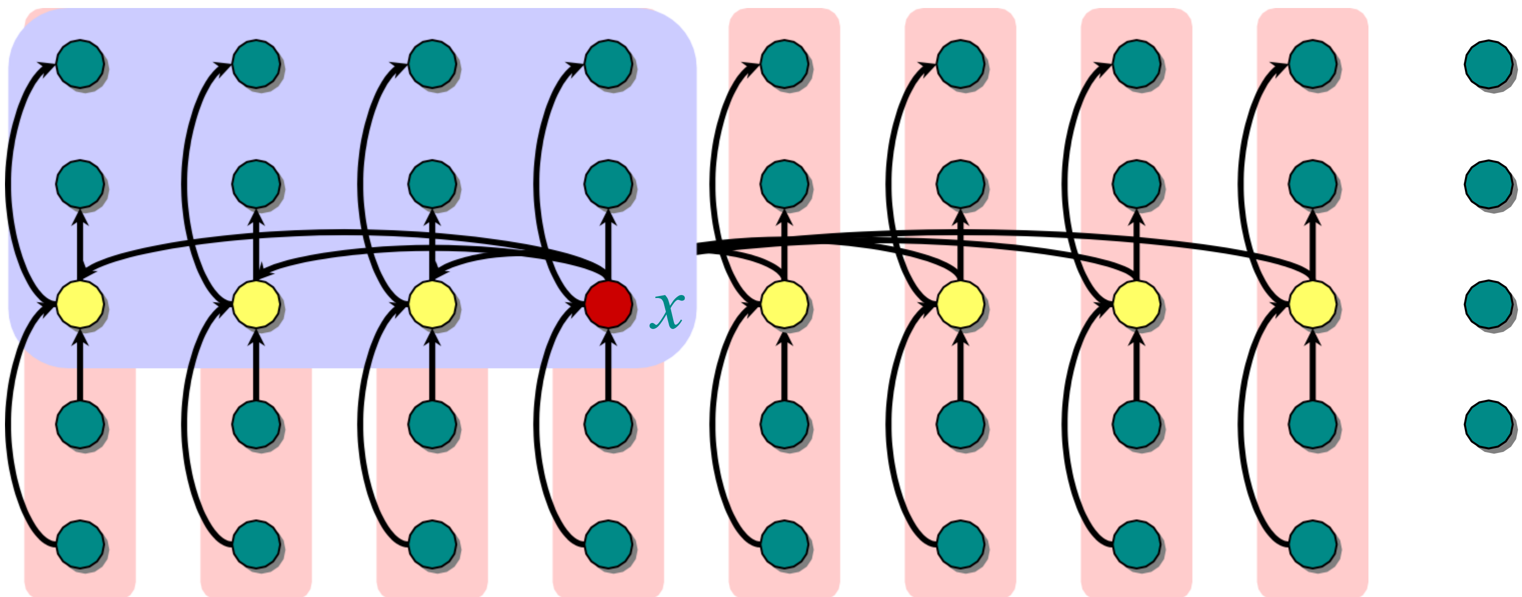
Τουλάχιστον μισοί από τους μέσους των ομάδων είναι $\leq x$, οι οποίοι είναι τουλάχιστον $\lfloor \lfloor n/5 \rfloor /2 \rfloor = \lfloor n/10 \rfloor$ μέσοι ομάδων

Μικρότερο



μεγαλύτερο

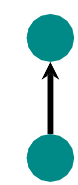
Ανάλυση (υποθέτοντας ότι τα στοιχεία είναι διαφορετικά)



Τουλάχιστον μισοί από τους μέσους των ομάδων είναι $\le x$, οι οποίοι είναι τουλάχιστον $\lfloor \lfloor n/5 \rfloor /2 \rfloor = \lfloor n/10 \rfloor$ μέσοι ομάδων

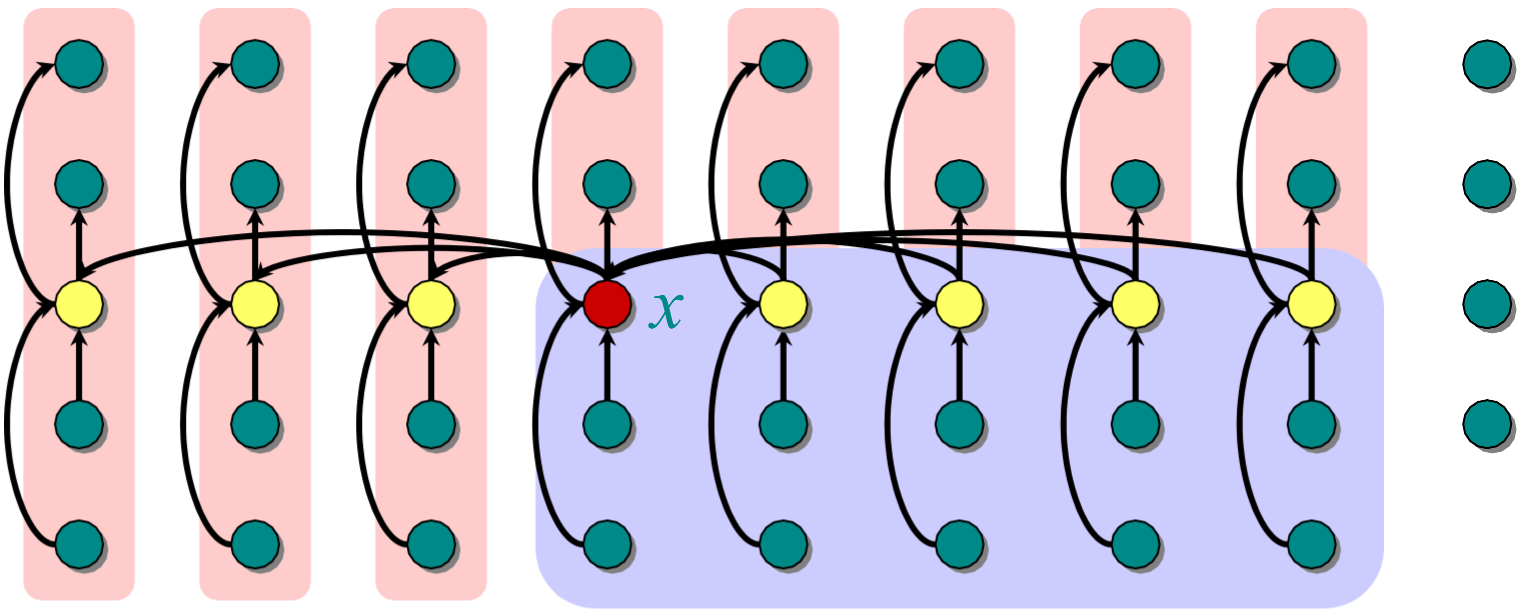
- Επομένως, τουλάχιστον $3\lfloor n/10 \rfloor$ στοιχεία είναι $\le x$.

Μικρότερο

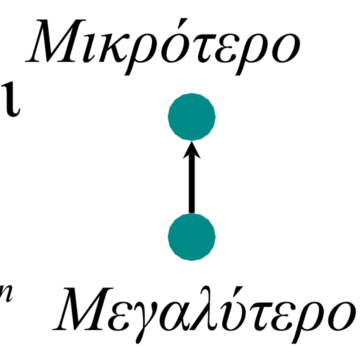


Μεγαλύτερο

Ανάλυση (υποθέτοντας ότι τα στοιχεία είναι διαφορετικά)



- Ομοίως , τουλάχιστον $3\lfloor n/10 \rfloor$ στοιχεία είναι $\geq x$.



Απλοποίηση

- Για $n \geq 50$, έχουμε $3\lfloor n/10 \rfloor \geq n/4$.
- Επομένως, για $n \geq 50$ η αναδρομική κλήση της SELECT στο βήμα 4 εκτελείται αναδρομικά σε $\leq 3n/4$ στοιχεία.
- Έτσι, το βήμα 4 απαιτεί χρόνο $T(3n/4)$ στη χειρότερη περίπτωση.
- Για $n < 50$, γνωρίζουμε ότι ο χρόνος εκτέλεσης είναι $T(n) = \Theta(1)$.

Προσδιορισμός αναδρομικής σχέσης

$T(n)$ — SELECT(i, n)

$\Theta(n)$

1. Διαίρεσε τα n στοιχεία σε ομάδες των 5. Βρες το μέσο καθεμίας ομάδας με ενθετική ταξινόμηση.

$T(n/5)$

2. Βρες το μέσο x των μέσων των ομάδων με αναδρομική εκτέλεση της SELECT

$\Theta(n)$

3. Διαμέρισε γύρω από τον οδηγό x. Έστω $k=rank(x)$

$T(3n/4)$

4. **if** $i = k$ **then return** x
   elseif $i < k$
   then αναδρομικά βρες το i-μικρότερο στοιχείο στο χαμηλότερο υποπίνακα
   else αναδρομικά βρες το $(i–k)$- μικρότερο στοιχείο στον υψηλότερο υποπίνακα

Επίλυση αναδρομής

$$T(n) = T\left(\frac{1}{5}n\right) + T\left(\frac{3}{4}n\right) + \Theta(n)$$

Αντικατάσταση: $T(n) \le \frac{1}{5}cn + \frac{3}{4}cn + \Theta(n)$

$T(n) \le cn$

- $$= \frac{19}{20}cn + \Theta(n)$$

$$= cn - \left(\frac{1}{20}cn - \Theta(n)\right)$$

$$\le cn \quad ,$$

αν το c επιλεχθεί μεγάλο αρκετά για να χειριστούμε και το $\Theta(n)$ και τις εναρκτήριες περιπτώσεις.

Συμπεράσματα

- Αφού η δουλειά σε κάθε επίπεδο της αναδρομής είναι ένα σταθερό ποσοστό (19/20) μικρότερο, η δουλειά ανά επίπεδο είναι μία γεωμετρική πρόοδος η οποία κυριαρχείται από τη γραμμική δουλειά στη ρίζα.
- Στην πράξη, ο αλγόριθμος τρέχει αργά, διότι η σταθερά μπροστά στο n είναι μεγάλη.
- Ο τυχαιοκρατικός αλγόριθμος είναι πολύ πιο πρακτικός.