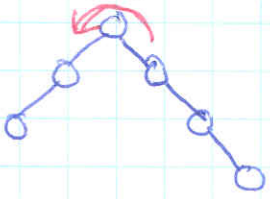
# Binary Search Trees (BSTs)

6.897
L3.1
Feb. 8, 2005

- standard: $O(\lg n)$ worst case / access
- when can we do better?
  - ↳ for what access sequences?
- Stochastic model: element $i$ accessed with prob. $p_i$
  - e.g. $p_j \approx 1$ (nearly all accesses are to $j$)
    - $\Rightarrow$ store $j$ at root
    - $\Rightarrow$ cost $\approx 1$ per access
  - Huffman tree (store data at leaves) ] but not a search tree
    - $\Rightarrow$ entropy bound $\sum_i p_i \lg \frac{1}{p_i} + O(1)$
- general access sequence $x = \langle x_1, x_2, \ldots, x_m \rangle$ allows correlation (not i.i.d.), whatever
  - e.g. $\langle 1, 2, \ldots, n \rangle$ (all elements in order)
    - $\Rightarrow$ treat tree like linked list — next element available in 1 rotation
    - $\Rightarrow$ cost = $O(1)$ per access



- (dynamic) OPT($x$) = min. cost to service $x$, over all BSTs
  - can tune to $x$!
- BST is $\alpha$-competitive if its $\text{cost}_{(x)} \leq \alpha \cdot \text{OPT}(x)$ for all $x$
- BST is dynamically optimal if it's $O(1)$-competitive

OPEN: is any BST dynamically optimal?
(is there a single "best" BST?)
↳ up to constant factors

- latest progress: $O(\lg \lg n)$-competitive (next lecture)
  [Demaine, Harmon, Iacono, Pătraşcu – FOCS 2004]