

Final Exam Mon April 10

1 sheet of notes, 2-sided

"guess what's on the exam and study those"

Definitely: Dynamic Programming, Linear Programming, Flows

## Cuckoo Graph

Cycles are dangerous. Want to know how long the chains are.

$n$  elements hashed

Lemma: For  $c > 1$  and  $m \geq 2cn$ , probability cuckoo graph contains a shortest path of length  $l \geq 1$  between vertices  $i$ , and  $j$  is  $\leq \frac{1}{mc^l}$ .

(Assume  $h_1(A)$  and  $h_2(A)$  are random, uniform table slots).

Proof: by induction on  $l$

Base  $l = 1$  Edge  $(i, j)$  exists in graph with prob  $\leq n \frac{2}{m^2} = \frac{1}{c \cdot m}$

For  $l > 1$  shortest path from  $i$  to  $j$  has length  $l$  only if there exists  $p$  and

1. There exists a shortest path from  $i$  to  $p$  of length  $l-1$  (occurs with probability  $\leq \frac{1}{m \cdot c^{l-1}}$ ).
2. and, there exists the edge  $(p, j)$  (occurs with prob  $\leq \frac{1}{mc}$ ).

These two together occur with probability,  $\leq \frac{1}{m^2 \cdot c^l}$

Sum over possible nodes  $p \rightarrow n \leq \frac{ml}{m^2 c^l}$

Probability that  $k$  and  $k'$  hash to the same path ("bucket") of cuckoo graph is probability of a path from  $h_1(k)$  or  $h_2(k)$  to  $h_1(k')$  or  $h_2(k')$

$$\leq 4 \sum_{l=1}^{\infty} \frac{1}{m} mc^l = \frac{4}{m} \frac{1}{c-1} = O\left(\frac{1}{m}\right).$$

Rehash means choose new hash functions and rehash all keys.

Probability rehash occurs:

- $\leq$  probability hashing creates cuckoo graph with a cycle
- $\leq \sum_{i=1}^m \Pr[\text{cycle involving } i]$
- $\leq \sum_{i=1}^m \sum_{l=1}^{\infty} \frac{1}{mc^l} \leq \frac{1}{2}$  for  $c \geq 3$ .

Expected number of rehashes is  $\leq 2$ .