Quiz 8: Streaming(DGIM) and Web Advertising _ Solution

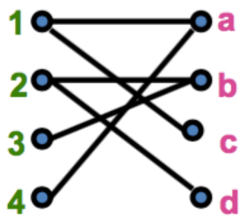
1) (1pts) Explain and write the definition of the competitive ratio.

Competitive ratio =

$$\min_{\text{all possible inputs } I} (|M_{greedy}|/|M_{opt}|)$$

(what is greedy's worst performance over all possible inputs I)

2) (1pts) Show a case using 4 ads and 4 queries to demonstrate the worst-case scenario in the greedy algorithm.

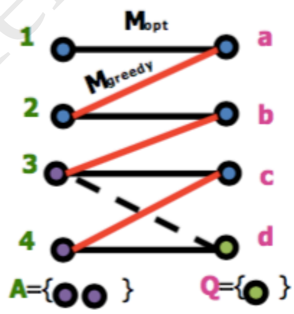


(1,a)
(2,b)

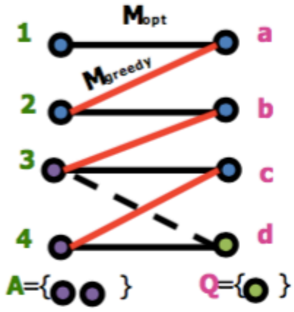
3) (2pts) Using the example below to show that the competitive ratio of the greedy algorithm is 1/2. (You need to use the sets Q and A in your answer)

(Refer to lecture "Web Advertising" page 32 and 33)

- Consider a case: $M_{greedy} \neq M_{opt}$
- Consider the set Q of queries matched in M_{opt} but not in M_{greedy}
- **A is the set of ads** that are adjacent (linked) to a non-matched query in Q, and A is already matched in M_{greedy}
  - If there exists such a non-matched (by M_{greedy}) ad adjacent to a non-matched query, then greedy would have matched them
- Since ads A are already matched in M_{greedy} then (1) $|M_{greedy}| \geq |A|$



- **Summary so far:**
  - Queries Q matched in M_{opt} but not in M_{greedy}
  - (1) $|M_{greedy}| \geq |A|$
- There are at least $|Q|$ such ads in A ($|Q| \leq |A|$) otherwise the optimal algorithm couldn't have matched all queries in Q
  - So: $|Q| \leq |A| \leq |M_{greedy}|$
- Q': matched in M_{opt} and also in M_{greedy}
  - $|M_{opt}| = |Q| + |Q'|$ and $|Q'| \leq |M_{greedy}|$
  - $|M_{opt}| \leq |M_{greedy}| + |Q|$
  - Worst case is when $|Q|$ is maximum $|Q| = |A| = |M_{greedy}|$
- $|M_{opt}| \leq 2|M_{greedy}|$ then $|M_{greedy}|/|M_{opt}| \geq 1/2$
- **Competitive Ratio = ½**
- **Greedy's worst performance over all possible inputs I**



33

4) (1pts) Fill up the table below with the Balance algorithm

- **Bidder A_1:** bid $x_1 = 20$ budget $b_1 = 40$
- **Bidder A_2:** bid $x_2 = 10$ budget $b_2 = 50$
- Assume ties are broken in favor of A_1

Query q	Assigned to Bidder (A_1, A_2 or No Ad)	Remaining Budget for A_1	Remaining Budget for A_2
At start	----	40	50
1st query q	A2	40	40
2nd query q	A1	20	40
3rd query q	A2	20	30
4th query q	A2	20	20
5th query q	A1	0	20
6th query q	A2	0	10
7th query q	A2	0	0
8th query q	No Ad	0	0

5) (2pts) Show that the storage requirement for the DGIM algorithm is $O(\log^2 N)$ bits (Refer to lecture “streaming” page 99-101)

- A ***bucket*** in the DGIM method is a record consisting of:
  - **(A) The timestamp of its end [$O(\log N)$ bits]**
  - **(B) The number of 1s between its beginning and end [$O(\log \log N)$ bits]**

- Timestamp: 0..N-1 (N is the window size)
  - So $\log_2 N$ bits

- Number of 1's: $2^j \leq N$, $j \leq \log_2 N$
  - So $\log_2(\log_2 N)$ for representing j
  - Can ignore; too small comparing to the timestamp requirement

- Each bucket requires $\approx O(\log_2 N)$

- At most 2*j buckets
  - of sizes: $2^j, 2^{j-1}, \ldots, 1$ (at most two for each size)

- Size of the largest bucket $2^j \leq N$
  - So $j \leq \log_2 N$ and $2j \leq 2\log_2 N$

- Total storage: $O(\log_2 N * \log_2 N)$ or $O(\log^2 N)$

6) (3pts) Show that the error rate for the DGIM algorithm is <= 50%

Suppose # of 1's in the last bucket $b = 2^j$

Case 1: estimate < actual value c

- Worst case: all 1's in bucket b are within range
- So estimate missed at most half of 2^j or 2^{j-1}
- $c \geq 2^j$
  - C has at least one 1 from b, plus at least one of buckets of lower powers: $2^{j-1} + 2^{j-2} \ldots + 1 = 2^j - 1$; $c >= 1 + 2^j - 1$; ***missed at most*** 2^{j-1}
  - $1 + 2 + 4 + .. + 2^{r-1} = 2^r - 1$
- So estimate missed ***at most 50% of c***
- That is, ***the estimate is at least 50% of c***

Case 2: estimate > actual value c

- Worst case: only rightmost bit of b is within range
- And only one bucket for each smaller power
- $c = 1 + 2^{j-1} + 2^{j-2} + \ldots + 1 = 2^j$
- Estimate $= 2^{j-1} + 2^{j-1} + \ldots + 1 = 2^j + 2^{j-1} - 1$
- So ***estimate is no more than 50% greater than c***