CMU 15-896
Matching 3: Online algorithms

Teacher: Ariel Procaccia
Display advertising

- Display advertising is the largest matching problem in the world
- Bipartite graph with advertisers and impressions
- Advertisers specify which impressions are acceptable — this defines the edges
- Impressions arrive online
The (simplest) model

• There is a bipartite graph $G = (U, V, E)$, $|U| = n$

• $U$ is known “offline”, the vertices of $V$ arrive online (with their incident edges)

• Objective: maximize size of matching

• ALG has competitive ratio $\alpha \leq 1$ if for every graph $G$ and every input order $\pi$ of $V$, $\frac{ALG(G, \pi)}{OPT(G)} \geq \alpha$
Algorithm GREEDY

- Algorithm GREEDY: match to an arbitrary unmatched neighbor (if one exists)

Poll 1: Competitive ratio of GREEDY?

1. $1/n$
2. $1/\sqrt{n}$
3. $1/\log n$
4. $1/2$
Upper bound

- **Observation:** The competitive ratio of any deterministic algorithm is at most $1/2$.
Take 2: Algorithm RANDOM

- Obvious idea: randomness
- Algorithm RANDOM: Match to an unmatched neighbor uniformly at random
- Achieves $\frac{3}{4}$ on previous example

Competitive ratio of RANDOM on graph on the right?
Take 3: Algorithm RANKING

- Algorithm RANKING:
  - Choose a random permutation \( \pi: U \to [n] \)
  - Match each vertex to its unmatched neighbor \( u \) with the lowest \( \pi(u) \)
- Looks like this is doing better than RANDOM on previous example!
- Theorem [Karp et al. 1990]: The competitive ratio of RANKING is \( 1 - 1/e \)
Proof of theorem

• Assume for ease of exposition that $\text{OPT} = n$
• Fix a perfect matching $M^*: U \cup V \to U \cup V$
• Fix $\pi$ and $u \in U$
• If $u$ is matched under $\pi$, $(\pi, u)$ is a match event at position $\pi(u)$, otherwise miss event
• ALG is the sum of probabilities of match events at all positions
Proof of theorem

• $\pi$ induces a matching $M^\pi$
• Consider a miss event $(\pi, u^*)$ with $\pi(u^*) = t$
• $v^* = M^*(u^*), u' = M^\pi(v^*)$
• Define $\pi_i$ by moving $u^*$ to position $i = 1, \ldots, n$
• Claim: for each $i$, $M^{\pi_i}(v^*) = u_i$
  with $\pi_i(u_i) \leq t$
Proof of theorem

• Proof of claim: by illustration

\[ u^* \quad v^* \quad u' \quad u' \]

\[ \pi_1 \quad \pi_2 \quad \pi_3 \quad \pi_4 \]
Proof of theorem

- We have a 1-to-1 mapping between miss events \((\pi, u^*)\) and match events \((\pi_i, u_i)\) where
  \[ M^{\pi_i}(u_i) = M^*(u^*) \quad \text{and} \quad \pi_i(u_i) \leq \pi(u^*) \]

- **Claim:** Each miss event at position \(t\) is mapped to \(n\) unique match events

- **Proof of claim:**
  - Fix miss events \((\pi, u)\) and \((\pi', u')\) such that \(\pi(u) = \pi'(u') = t\), and both are mapped to \((\hat{\pi}, \hat{u})\)
  - \(M^{\hat{\pi}}(\hat{u}) = M^*(u) = M^*(u') \Rightarrow u = u'\)
  - The map only moves \(u\) from position \(t\) in \(\pi\) and \(\pi'\), giving \(\hat{\pi}\) in both cases \(\Rightarrow \pi = \pi'\) ■
Proof of theorem

• We get the following set of equations for every $t = 1, ..., n$:

$$n \cdot \Pr[\text{Miss at } t] \leq \sum_{s \leq t} \Pr[\text{Match at } s]$$

• Setting $x_t = \Pr[\text{Match at } t]$, this is

$$1 - x_t \leq \frac{1}{n} \sum_{s \leq t} x_s$$

• By minimizing the objective function $\sum_t x_t$ over this polytope, we get $\sum_t x_t \geq \left(1 - \frac{1}{e}\right)n$  ■
**Upper bound**

- **Theorem [Karp et al. 1990]:** No randomized alg has competitive ratio better than $1 - \frac{1}{e} + o(1)$
- The proof below is due to Wajc [2015]
- Fractional algorithm: deterministically assign fractional weights to edges such that s.t.
  \[ \forall u \in U \cup V, \ f(u) = \sum_{(u,v) \in E} w_{uv} \leq 1 \]
- **Lemma [Wajc 2015]:** For any randomized alg there is a fractional alg with at least the same competitive ratio
Proof of theorem

- First online vertex $v_1$ is connected to all $U$
- Let $u_1 \in \arg\min_{u \in U} f(u)$, in particular $f(u_1) \leq 1/n$
- $u_1$ will not be connected to any future online vertex
Proof of theorem

- $t$-th online vertex $v_t$ is connected to all $U \setminus \{u_1, \ldots, u_{t-1}\}$
- $u_t \in \text{argmin}_{u \in U \setminus \{u_1, \ldots, u_{t-1}\}} f(u)$
- $u_t$ will not be connected to any future online vertex
Proof of theorem

Poll 2: What is OPT?

1. $n/2$
2. $n \left(1 - \frac{1}{e}\right)$
3. $3n/4$
4. $n$
Proof of theorem

• After step $t$, offline vertices that continue to be matched are matched to an average of at least

$$f(u) = \sum_{k=1}^{t} \frac{1}{n-k+1}$$

• Following the arrival of the $t$-th online vertex with $t = n \left(1 - \frac{1}{e}\right) + 1$, it holds that offline vertices that will neighbor future online vertices are matched to an average of

$$f(u) = \sum_{k=1}^{n\left(1 - \frac{1}{e}\right) + 1} \frac{1}{n-k+1} = \sum_{k=\frac{n}{e}}^{n} \frac{1}{k} \geq \ln(n) - \ln\left(\frac{n}{e}\right) = 1$$
Proof of theorem

• So at step $t$, $\frac{1}{n-t} \sum_{k=t+1}^{n} f(u_k) \geq 1$, but because $f(u) \leq 1$ for all $u \in U$, this means that $f(u_k) = 1$ for all $k = t + 1, \ldots, n$

• That is, the algorithm cannot match the vertices $v_{t+1}, \ldots, v_n$

• $\text{ALG} \leq n \left( 1 - \frac{1}{e} \right) + 1$ \qed