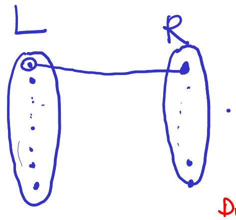
Online bipartite matching



P:

$$\max \sum_{e \in E} x_e$$

s.t. $\sum_{v \in e} x_e \leq 1 \quad \forall v \in V$

Dual variable y_v

$x_e \geq 0.$

D:

$$\min \sum_{v \in V} y_v$$

s.t. $y_u + y_v \geq 1 \quad \forall (u,v) \in E.$

$y_v \geq 0. \quad \forall v.$

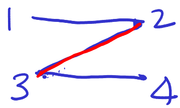
Analysis of the greedy algorithm using dual-fitting.

- Initialize $y_v \leftarrow 0 \quad \forall v \in L \cup R.$
- whenever (u,v) is added to the matching,

  $y_u \leftarrow \frac{1}{2}, \quad y_v \leftarrow \frac{1}{2}.$

$\sum_{v \in L \cup R} y_v$ = size of the computed matching.

This is because, everytime an edge is included, the dual objective by 1 (no dual variable is set more than once).



$y_2 = y_3 = \frac{1}{2}, \quad y_1 = y_4 = 0$

Dual constraint violated for the edge $(1,2)$.

Claim: $(2y_v : v \in L \cup R)$ is dual feasible.

Proof: Consider $(u,v) \in E$. If (u,v) is in the matching, then $y_u = \frac{1}{2}, y_v = \frac{1}{2}$.

$$\Rightarrow 2y_u + 2y_v = 2 > 1. \checkmark$$

If (u,v) is not in the matching. Either u or v must be matched. So, the corresponding y variable must be $\frac{1}{2}$.

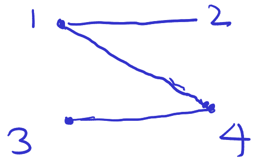
$$\Rightarrow 2y_u + 2y_v \geq 1 \checkmark.$$

$$\underset{\text{computed soln.}}{2 \cdot \text{Cost of}} = \sum 2 \cdot y_v \underset{\text{weak duality}}{\geq} \text{Val}(P) \geq \text{OPT}$$

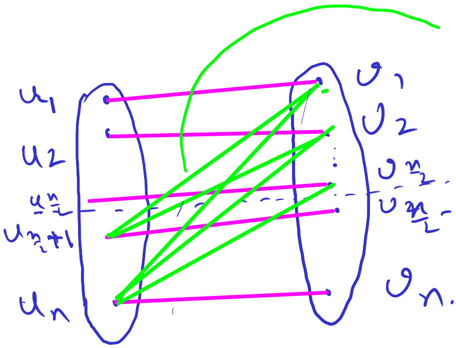
$\Rightarrow$ The greedy algorithm is 2-competitive.

Randomized algorithms for bipartite matching

Idea 1: Match an incoming vertex with a uniform random unmatched neighbor.



Idea 1 produces a matching of expected size $\frac{3}{2} = \frac{3}{4} \cdot OPT$.



Complete bipartite graph.

— : u_i : $i \le \frac{n}{2}$

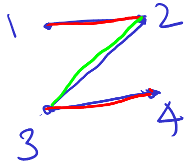
— : u_i : $i > \frac{n}{2}$

Let the vertices on the right arrive in the order $v_1, v_2, \ldots, v_{\frac{n}{2}}, v_{\frac{n}{2}+1}, \ldots, v_n$.

Exercise: Idea 1 in expectation produces a matching of size $\frac{n}{2} + O(\log n)$.

Idea 2 (RANKING algorithm)

- choose a random permutation σ of vertices on the left.
- Match each incoming vertex to its first (with respect to σ) unmatched neighbor.

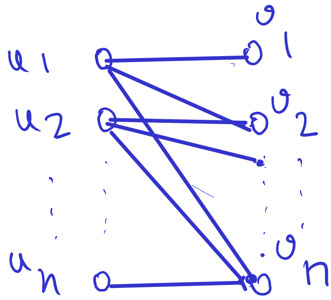


$\delta_1 = 1\ 3 \quad \{(1,2),(3,4)\}.$

$\delta_2 = 3\ 1 \quad \{(2,3)\}$

$y_1 = y_2 = \frac{1}{2}.$ $\quad y_1 = y_4 = 0.$

$y_3 = y_4 = \frac{1}{2}.$ $\quad y_2 = y_3 = 1.$



$n \quad n-1 \cdots 1$

$\underbrace{y_{u_1} = \frac{1}{2}, \quad y_{v_1} = 0.}$

Idea 2:

$L = \{u_1, \dots, u_n\}.$

- For each $i = 1, \dots, n$ sample a uniform random real number x_i in the interval $[0,1]$ independently.
- In each step, match the incoming vertex with its unmatched neighbour with minimum x value.

Setting the dual variables

Plan: For each edge (u,v) included in the matching, distribute the cost of 1 between y_u & y_v depending on x_u.

Aim: $\exists c_{<2}$, $\forall (u,v) \in E$, $c \cdot (y_u + y_v) \geq 1$.

Note that $\forall (u,v)$ in the computed matching

$$y_u + y_v = 1.$$

$$y_u := g(x_u) \quad \text{and} \quad y_v = 1 - g(x_u)$$

Want: There exists a 'small' constant c (<2) s.t., $\forall (u,v) \in E$

$$c \cdot (y_u + y_v) \geq 1.$$

Fix an edge $(u,v) \in E$. Fix (condition on) x_w for all $w \in L \setminus \{u\}$.

Consider running RANKING on the graph G' obtained by removing u from G.

Case 1: v is not matched.

Claim: u is matched by RANKING run on G.

Proof: Consider the time step when v arrives. At this point, if u is matched then the proof follows. Otherwise the run of

RANKING on G so far has been identical to the run of RANKING on G'. Thus v does not have any unmatched neighbor other than u. So v will be matched with u.

$$y_u = g(x_u). \quad \therefore \mathbb{E}[y_u] = \mathbb{E}[g(x_u)] = \int_0^1 g(y)\,dy$$

It is not clean if v is matched in G. $\to y_v \geq 0.$

$\boxed{y_v = 0}$ ($\because$ v not matched).

$$c \cdot \mathbb{E}[y_u + y_v] \geq 1 \Rightarrow \underbrace{\int_0^1 g(y)\,dy}_{\text{"large"}} \geq \frac{1}{c}$$

We want $\int_0^1 g(y)\,dy$ to be large. $\geq \frac{1}{c}$. — (1)

<u>Case 2</u>: v is matched to a vertex u'.

<u>Claim</u>: Assume that $x_u < x_{u'}$. Then u is matched when RANKING is run on G.

<u>Proof</u>: Assume that u is not matched when v arrives. (otherwise the claim follows). Then, the run of RANKING on G and G' are identical. Thus the min. x value of unmatched neighbors of v other than u is $x_{u'}$. Since $x_u < x_{u'}$, v is matched to u. ▨

If $x_u < x_{u'}$, then u is matched (by the above claim) and $y_u = g(x_u)$.

Claim: Let $x_u < x_{u'}$. Then v is matched to some vertex with x value at most $x_{u'}$, when Ranking is run on G.

Proof: Imagine running RANKING on G & G' parallely. At each time step, the

(G-run) (G'-run)

set of unmatched vertices in G-run is a superset of the set of unmatched vertices in G'-run (verify. Do an induction on t). In the iteration in which v comes, all the unmatched neighbors of v in G'-run are also unmatched in G. Thus in the G-run v is matched with a vertex with x value at most $x_{u'}$.

g is a monotone fn. $g : [0,1] \to [0,1]$ — 2

If $x_u < x_{u'}$, the above claim implies that

$$y_v \geq 1 - g(x_{u'}).$$

$$y_u = \begin{cases} g(x_u) & \text{if } u \text{ is matched} \\ 0 & \text{otherwise.} \end{cases}$$

Not clean if u is matched if $x_u > x_{u'}$.

$$\mathbb{E}[y_u] \overset{\geq}{=} \int_0^{x_{u'}} g(y)\,dy$$

$$y_u = g(x_u) \cdot \mathbb{1}\{x_u \leq x_{u'}\}.$$

$$\mathbb{E}[y_u] = \int_0^1 g(y) \cdot \mathbb{1}\{y \leq x_{u'}\}$$

$$\mathbb{E}[y_v] \geq 1 - g(x_{u'})$$

$$c \cdot \mathbb{E}[y_u + y_v] \geq c \cdot \left[\int_0^{x_{u'}} g(y)\,dy + 1 - g(x_{u'})\right]$$

$$\underset{\text{want}}{\geq} 1$$

Sufficient to ensure that

$$c\left[\int_0^{x_{u'}} g(y)\,dy + 1 - g(x_{u'})\right] \geq 1$$

$$\int_0^{x_{u'}} g(y)\,dy + 1 - g(x_{u'}) \text{ is "large"} \geq \frac{1}{c}$$

$$\underbrace{\int_0^{\lambda} g(y)\,dy + 1 - g(\lambda) > \frac{1}{c} \quad \forall \lambda \in [0,1].}_{③}$$

We want a g that satisfies ①, ②, ③ for as small a c as possible.

Let us choose a g s.t. the function

$$f(\lambda) := \int_0^{\lambda} g(y)\,dy + 1 - g(\lambda) \quad \text{is}$$

constant.

$$0 = f'(\lambda) = g(\lambda) - g'(\lambda)$$

$$\Rightarrow g(\lambda) = g'(\lambda).$$

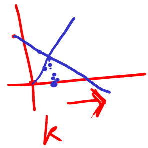
$$\Rightarrow g(y) = ke^{y}$$

Monotone. Satisfies requirement 2.

$$f(\lambda) = \int_0^{\lambda} g(y)\,dy + 1 - g(\lambda)$$

$$= \cancel{ke^{\lambda}} - k + 1 - \cancel{ke^{\lambda}} = 1 - k.$$

$$1 - k \geq \frac{1}{c}. \quad — (4)$$



$$\int_0^1 g(y)\,dy = ke - k \geq \frac{1}{c} \quad — (5)$$

We want to choose k s.t. (4) & (5) are satisfied with the least possible value of c. The optimal value of k is when $1-k = ke - k \Rightarrow k = \frac{1}{e}$.

$$g(y) = e^{y-1}.$$

set c to $\dfrac{1}{1-\frac{1}{e}}$.

Competitive ratio $= 1 - \frac{1}{e}$. $\left(> \frac{1}{2}\right)$.