

Online Algorithms

Anil Maheshwari

anil@scs.carleton.ca
School of Computer Science
Carleton University
Canada

Bipartite Matching

Linear Program and Dual Linear Program

Primal-Dual LP

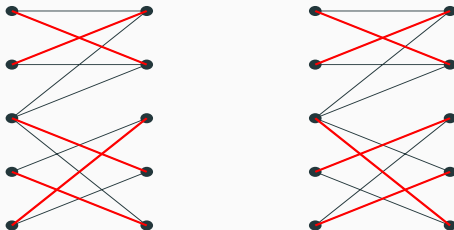
Fractional Matching

Bipartite Matching

Bipartite Matching

Let $G = (V = L \cup R, E)$ be a bipartite graph where the vertex set V consists of the sets L and R (referred to as 'left' and 'right' sets) and a set E of edges (v, w) where $v \in L$ and $w \in R$.

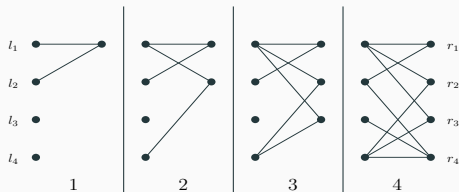
The set $M \subseteq E$ is a matching in G if no two edges in M share a vertex.



Online Matching Problem

Input: All the vertices in the set L are known in advance, but the vertices in R and the edges are presented over time. At each time instant $t \in \{1, 2, 3, \dots\}$, a new vertex $r_t \in R$ and all its incident edges arrive.

Task: The online matching algorithm needs to decide among all the currently unmatched neighbors of r_t in the set L to which vertex (if any) r_t should be matched. The vertex r_t remains matched to that vertex for the rest of the algorithm.



Output: Find a matching M of the largest possible size. Find M such that the ratio $\frac{|M|}{|M^*|}$ is as large as possible, where M^* is (offline) maximum matching in G .

Competitive Ratio

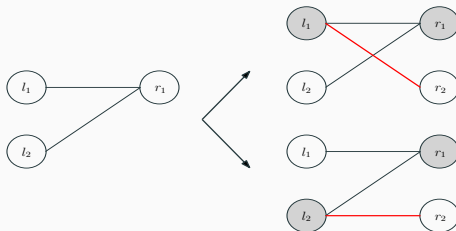
Ratio of the value returned by an Online Algorithm in comparison to the (best) Offline algorithm.

What is the largest value of $c \leq 1$, such that

$$\frac{\text{Value (online algorithm)}}{\text{Value (off-line algorithm)}} \geq c$$

Lower Bound on Deterministic Algorithms

Consider a bipartite graph on 4 vertices, where $L = \{l_1, l_2\}$ and $R = \{r_1, r_2\}$. At the first time step the algorithm is presented with the vertex r_1 and the two incident edges (r_1, l_1) and (r_1, l_2) .



Adversary chooses what to do in the next time stamp and therefore the Competitive Ratio = $\frac{1}{2}$

Greedy Bipartite Matching Algorithm

Greedy Online Matching Algorithm:

At time step t :

Match r_t to any of the unmatched neighbors in the set L

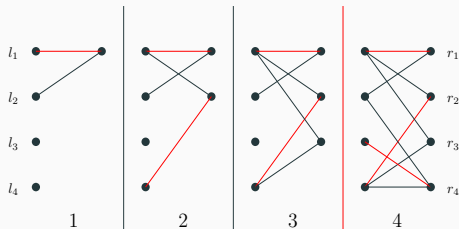


Figure 1: Competitive Ratio: $\frac{|M|}{|M^*|} = \frac{3}{4}$

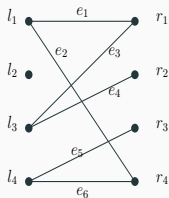
Exercise: Show that the Greedy Online Matching Algorithm has a competitive ratio $\geq \frac{1}{2}$.

Linear Program and Dual Linear Program

Linear Programs and Dual Linear Programs

Definitions

1. For each edge $e \in E$, let $x_e \geq 0$. Let $x = [x_1, \dots, x_{|E|}]$ be the vector of variables corresponding to the edges.
2. $\mathcal{N}(v)$, the neighborhood set of v , is the set of edges incident to the vertex v .
3. $c = [1, \dots, 1]$ is a vector of all 1s of length $|E|$
4. $b = [1, \dots, 1]$ is a vector of all 1s of length $|V|$
5. A is a $|V| \times |E|$ matrix and its ij -th entry is 1 if the edge corresponding to the column j is incident on the vertex corresponding to the row i , otherwise 0.



$$A = \begin{matrix} & \begin{matrix} e_1 & e_2 & e_3 & e_4 & e_5 & e_6 \end{matrix} \\ \begin{matrix} l_1 \\ l_2 \\ l_3 \\ l_4 \\ r_1 \\ r_2 \\ r_3 \\ r_4 \end{matrix} & \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \end{matrix}$$

The Primal Linear Program can be stated as follows:

Primal LP:

Objective function:

$$\max \sum_{e \in E} x_e \quad \Bigg| \quad \max c^T x$$

Subject to:

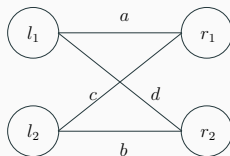
$$\text{For all } v \in L \cup R : \sum_{e \in \mathcal{N}(v)} x_e \leq 1 \quad \Bigg| \quad Ax \leq b$$

$$\text{For all } e \in E : x_e \geq 0 \quad \Bigg| \quad x \geq 0.$$

Objective function

The size of the maximum matching is the optimal value of the objective function of the LP

Example: Primal LP



Bipartite Matching LP:

$\max a + b + c + d,$

Subject to:

$l_1 : a + d \leq 1$

$l_2 : b + c \leq 1$

$r_1 : a + c \leq 1$

$r_2 : b + d \leq 1$

$a, b, c, d \geq 0$

Maximum value of the objective function of LP is 2.

For example, set $a = b = c = d = \frac{1}{2}$

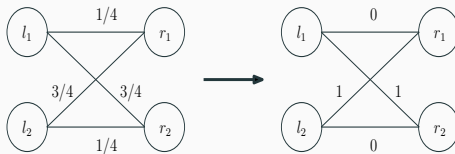
Alternatively, set $a = b = 1$ and $c = d = 0$

Integral Solution

For the bipartite matching LP

$$\begin{aligned} \max \quad & \sum_{e \in E} x_e, \\ \forall v \in L \cup R: \quad & \sum_{e \in \mathcal{N}(v)} x_e \leq 1, \\ \forall e \in E: \quad & x_e \geq 0 \end{aligned}$$

there is always an integral solution (i.e. each variable taking integral values) that achieves an optimal value.



Primal-Dual LP

Consider the following Linear Program:

$$\begin{aligned} \max \quad & x_1 + x_2 \\ \text{s.t.} \quad & x_1 + 2x_2 \leq 4 \\ & 2x_1 + x_2 \leq 6 \\ & x_1, x_2 \geq 0 \end{aligned}$$

Since $x_1 \geq 0$ and $x_2 \geq 0$: we have

Observation 1: Value of objective function $x_1 + x_2 \leq 2x_1 + x_2 \leq 6$

Observation 2: Value of objective function $x_1 + x_2 \leq x_1 + 2x_2 \leq 4$

Observation 3: Value of objective function

$$x_1 + x_2 \leq \frac{1}{2}(x_1 + 2x_2) + \frac{1}{4}(2x_1 + x_2) = x_1 + \frac{5}{4}x_2 \leq \frac{1}{2}4 + \frac{1}{4}6 = \frac{7}{2}$$

Observation 4: Value of objective function is upper bounded by

$x_1 + x_2 \leq y_1(x_1 + 2x_2) + y_2(2x_1 + x_2) = x_1(y_1 + 2y_2) + x_2(2y_1 + y_2) \leq 4y_1 + 6y_2$,
provided

1. $y_1, y_2 \geq 0$
2. $y_1 + 2y_2 \geq 1$ (corresponding to x_1)
3. $2y_1 + y_2 \geq 1$ (corresponding to x_2)

Finding the right upper bound can also be expressed as a **(dual)** linear program:

$$\min 4y_1 + 6y_2$$

$$y_1 + 2y_2 \geq 1$$

$$2y_1 + y_2 \geq 1$$

$$y_1, y_2 \geq 0$$

Example (contd.)

	Primal	Dual
Objective Value	$\max x_1 + x_2$	$\min 4y_1 + 6y_2$
Constraints	$x_1 + 2x_2 \leq 4$	$y_1 + 2y_2 \geq 1$
	$2x_1 + x_2 \leq 6$	$2y_1 + y_2 \geq 1$
	$x_1, x_2 \geq 0$	$y_1, y_2 \geq 0$

- Set $y_1 = y_2 = 1/3$. This choice satisfies Dual LP constraints and results in an upper bound of $\frac{10}{3} < \frac{7}{2}$
 - Set $x_1 = \frac{8}{3}$ and $x_2 = \frac{2}{3}$ satisfies Primal LP constraints and results in the objective value of $\frac{10}{3}$
- ⇒ The upper bound using the linear combination that we obtained is the optimal value

Primal-Dual LP - Standard Form

Primal-Dual LP Pair

Primal LP: $\max c^T x$ subject to $Ax \leq b, x \geq 0$

Dual LP: $\min b^T y$, subject to $A^T y \geq c, y \geq 0$

Observations:

1. For each variable in the Primal we have a constraint in the Dual.
2. For each constraint in the Primal we have a variable in the Dual.
3. Maximization becomes a Minimization problem.
4. (Weak Duality) If x and y are feasible solutions to the Primal and Dual LPs: $c^T x \leq (A^T y)^T x = y^T (Ax) \leq y^T b = b^T y$

Strong Duality Theorem

If x and y are optimal values for the Primal and Dual LPs, then $c^T x = b^T y$

Primal-Dual LP for Matching

Primal LP:

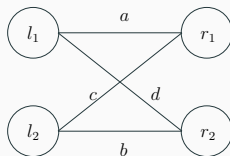
$$\begin{array}{ll} \max \sum_{e \in E} x_e & \left| \max c^T x \right. \\ \text{Subject to:} & \\ \text{For all } v \in \{L \cup R\} : \sum_{e \in \mathcal{N}(v)} x_e \leq 1 & \left| Ax \leq b \right. \\ \text{For all } e \in E : x_e \geq 0 & \left| x \geq 0 \right. \end{array}$$

For Dual LP: Introduce $|V|$ variables corresponding to each vertex constraint of the primal. We label them $p_1, \dots, p_{|V|}$ and let $p = (p_1, \dots, p_{|V|})^T$

Dual LP:

$$\begin{array}{ll} \min \sum_{v \in V} p_v & \left| \min b^T p \right. \\ \text{Subject to:} & \\ \text{For all edges } e = (v, w) \in E : p_v + p_w \geq 1 & \left| A^T p \geq c \right. \\ p_v \geq 0, \text{ for all } v \in V & \left| p \geq 0 \right. \end{array}$$

Example: Primal-Dual Bipartite Matching LP



	Primal LP	Dual LP
Objective Value	$\max a + b + c + d$	$\min l_1 + l_2 + r_1 + r_2$
Subject to:	$l_1 : a + d \leq 1$	$a : l_1 + r_1 \geq 1$
	$l_2 : b + c \leq 1$	$b : l_2 + r_2 \geq 1$
	$r_1 : a + c \leq 1$	$c : l_2 + r_1 \geq 1$
	$r_2 : b + d \leq 1$	$d : l_1 + r_2 \geq 1$
	$a, b, c, d \geq 0$	$l_1, l_2, r_1, r_2 \geq 0$

Greedy Bipartite Matching Algorithm

Greedy Online Matching Algorithm:

At time step t :

Match r_t to any of the unmatched neighbors in the set L

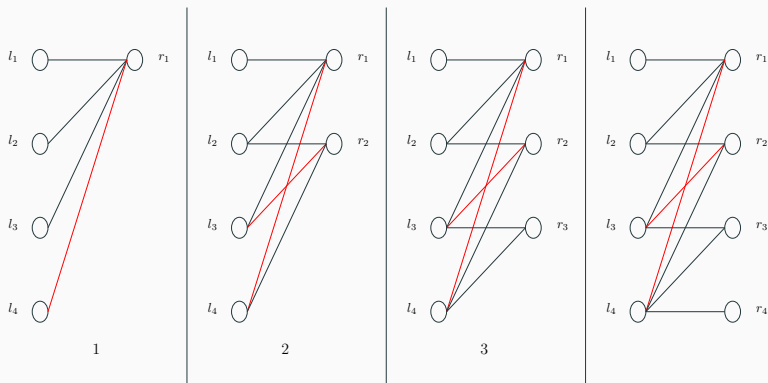


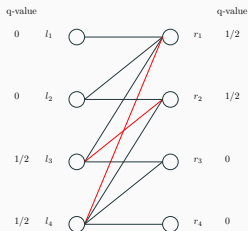
Figure 2: Red edges in greedy matching.

Analysis of Greedy Algorithm using Primal-Dual LP

For analysis, for each vertex $v \in V$ define the quantity $q_v \geq 0$ as follows.

Initialize: For all $v \in V$: $q_v = 0$

Update: If the greedy algorithm decides to add edge $e = (v, w)$ to the matching, set $q_v = \frac{1}{2}$ and $q_w = \frac{1}{2}$.



Observation: The size of the matching M reported by the greedy algorithm is given by $|M| = \sum_{v \in L \cup R} q_v$

Analysis of Greedy Algorithm (contd.)

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

Subject to:

For all edges $e = (v, w) \in E : p_v + p_w \geq 1$

$p_v \geq 0$, for all $v \in L \cup R$

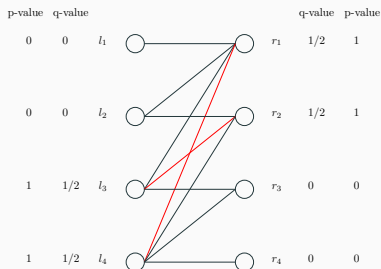
$$\min b^T p$$

$$A^T p \geq c$$

$$p \geq 0.$$

For all $v \in V$: set $p_v = 2q_v$

Observation: For each edge $e = (v, w) \in E$, $p_v + p_w = 2q_v + 2q_w \geq 1$ (as the greedy algorithm doesn't leave both v and w unmatched)



Analysis of Greedy Algorithm (contd.)

Competitiveness of Greedy

Greedy algorithm is $\frac{1}{2}$ -competitive.

Proof:

1. The value of the objective function of the dual LP is given by

$$\sum_{v \in V} p_v = 2 \sum_{v \in V} q_v = 2|M|.$$

2. From the weak/strong duality, the objective value of dual LP ($= 2|M|$) is an upper bound to the objective value of Primal LP ($= |M^*|$)
3. $2|M| \geq |M^*|$
4. Thus, Competitive Ratio $\frac{|M|}{|M^*|} \geq \frac{1}{2}$

□

NOTE: Primal-Dual LP is used only for proving that the Greedy Online Bipartite Matching Algorithm is $\frac{1}{2}$ -competitive.

In the algorithm, we **don't** solve an LP.

Fractional Matching

Fractional Bipartite Matching

Input: Bipartite graph $G = (V = L \cup R, E)$

- Vertices in L are known in advance and each has a unit capacity.
- Vertices in R come in an online fashion along with its incident edges.
- Each vertex in R has a unit amount of information to handout.
- At each time instant t , we transmit the information from the current vertex r_t to its neighboring vertices in the set L if:
 1. Sum total of the information transmitted from r_t to its neighbors in L is at most 1.
 2. One or more neighbors of r_t may receive the information provided they do not exceed their capacity of 1.
 3. Once the information is transmitted it cannot be reversed.

An Observation

Let x_e = Amount of information that travels on edge e .

For any vertex $v \in \{L \cup R\}$, $\text{Level}(v) = \sum_{w \in N(v)} x_{vw} \leq 1$

Fractional Matching Problem

Maximize the total information received by the vertices in the set L , i.e.

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

Note: For the (static) graph G the Bipartite Matching LP,

$$\max \sum_{e \in E} x_e, \forall v \in L \cup R : \sum_{e \in N(v)} x_e \leq 1, \forall e \in E : x_e \geq 0,$$

applies to this problem formulation as x_e 's can take fractional values. Value of the objective function is the size of the maximum fractional matching in G .

Waterlevel Algorithm

WATERLEVEL Algorithm

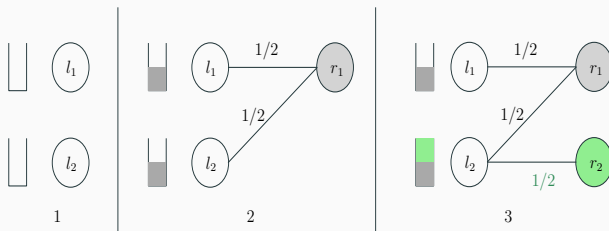
At any time step t :

Drain the water (information) from r_t to its neighbors where the preference is always given to the neighbor with the largest residual capacity remaining till

Case 1: All neighbors of r_t are saturated, or

Case 2: r_t transmits all its information

Example 1: $G = (V = L \cup R, E)$, where $L = \{l_1, l_2\}$, $R = \{r_1, r_2\}$ and $E = \{r_1 l_1, r_1 l_2, r_2 l_2\}$



Example 2

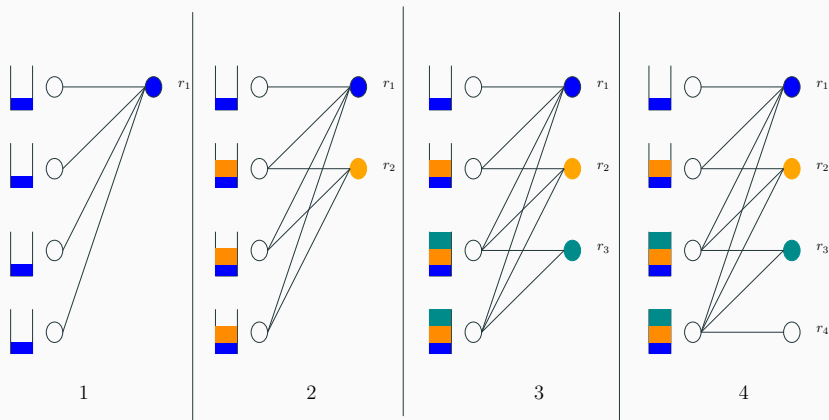
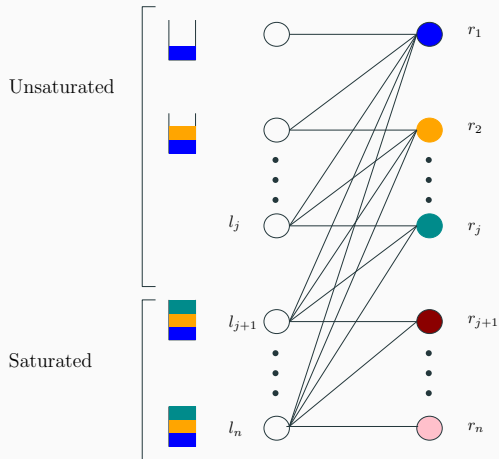


Figure 3: Fractional Flow Value = $4 \times \frac{1}{4} + 3 \times \frac{1}{3} + 2 \times \frac{5}{12} = \frac{17}{6} > 2$. Total flow value received at vertices in L are $l_1 = \frac{1}{4}$; $l_2 = \frac{1}{4} + \frac{1}{3} = \frac{7}{12}$; and $l_3 = l_4 = \frac{1}{4} + \frac{1}{3} + \frac{5}{12} = 1$.

Example 3

Let $L = \{l_1, \dots, l_n\}$ and $R = \{r_1, \dots, r_n\}$

Let $E = \{(l_i, r_j) | i \geq j, \text{ for all } i, j \in \{1, \dots, n\}\}$



Example 3 (contd.)

Let j be the first index at which there is no further flow of information from vertices in $r_t \in R$ for $t > j$.

$\implies$ All the vertices $l_{j+1}, \dots, l_n$ are saturated

The index j must satisfy $\frac{1}{n} + \frac{1}{n-1} + \dots + \frac{1}{n-j+1} \geq 1$

For what value of j , $\frac{1}{n} + \frac{1}{n-1} + \dots + \frac{1}{n-j+1} \geq 1$?

Recall that $H_n = \sum_{i=1}^n \frac{1}{i} \approx \ln n$

Thus, $\frac{1}{n} + \frac{1}{n-1} + \dots + \frac{1}{n-j+1} = H_n - H_{n-j} \approx \ln n - \ln(n-j) = \ln \frac{n}{n-j}$

If $j = n(1 - \frac{1}{e})$,

$$\ln \frac{n}{n-j} = \ln \frac{n}{n - n(1 - \frac{1}{e})} = \ln e = 1$$

Example 3 (contd.)

Competitive Ratio

The competitive ratio of the Waterlevel Algorithm on Example 3 is $\approx (1 - \frac{1}{e})$

Proof:

1. Vertices $r_1, \dots, r_{j-1}$ are able to send all of their information to vertices of L
2. Vertices $r_{j+1}, \dots, r_n$ aren't able to send any information
3. Total weight of the fractional matching computed by the WATERLEVEL algorithm is $\approx j \approx n(1 - \frac{1}{e})$
4. G has a perfect matching (match l_i to r_i) $\implies$ optimal offline matching has size n .
5. Competitive ratio $\approx \frac{n(1 - \frac{1}{e})}{n} = (1 - \frac{1}{e}) \approx 0.63$

□

Primal Dual Analysis of WATERLEVEL Algorithm

	$\min \sum_{v \in V} p_v$	$\min b^T p$
Dual LP:	Subject to:	
	For all edges $e = (v, w) \in E : p_v + p_w \geq 1$	$A^T p \geq c$
	For all $v \in V : p_v \geq 0$	$p \geq 0$

Recall the analysis of Greedy Algorithm:

By setting $p_v = 2q_v$, we had

1. For all vertices v , $p_v \geq 0$
2. For all edges $e = (v, w) \in E$, $p_v + p_w \geq 1$
3. $\sum_{v \in V} p_v = 2 \sum_{v \in V} q_v = 2|M|$
4. Upper Bound: $2|M| \geq |M^*| \implies \frac{|M|}{|M^*|} \geq \frac{1}{2}$

Question: Can we follow the same analysis?

1st Approach

For all $v \in V$, initialize $q_v = 0$

After the execution of the WATERLEVEL algorithm, for each edge $e = vw \in E$, set $q_v = q_v + \frac{1}{2}x_{vw}$ and $q_w = q_w + \frac{1}{2}x_{vw}$, where x_{vw} is the amount of flow on edge vw .

Size of the fractional matching: $|M| = \sum_{v \in V} q_v$

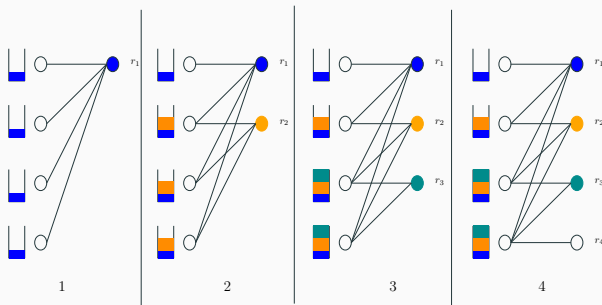
What about the constraint $p_v + p_w \geq 1$ for all edges $e = (v, w) \in E$ in the Dual LP?

Can we get away by setting $p_v = cq_v$, where $c < 2$?

If so, competitive ratio will be $\frac{1}{c}$

1st Approach (contd.)

Consider the edge $l_4 r_4$ in Example 2. Edges incident to r_4 have $x_e = 0$



For any of those edges the sum total of the q values of their end points is at most $\frac{1}{2}$

$\implies$ to satisfy the Dual LP constraints we need to set $p_v = 2q_v$

2nd Approach

Idea: Uneven Split

Instead of splitting the value of the flow x_e on each edge $e = vw \in E$ between its endpoints evenly, split in such a way that $q_v + q_w \geq 1 - \frac{1}{e} = \frac{e-1}{e}$

$$p_v = \frac{e}{e-1} q_v$$

All constraints of the Dual LP are satisfied

The competitive ratio will be $\geq 1 - \frac{1}{e} \approx 0.63$

Observation

Assume that after the algorithm has terminated, the vertex $v \in L$ isn't saturated. Let the information content that v has received during the entire execution of the algorithm equals $\text{Level}(v) < 1$. Moreover, assume that $vw \in E$. Now consider the step in the online algorithm when $w \in R$ was revealed. In that step w routed the information to its neighbors (including v) in L whose Level's were at most $\text{Level}(v)$.

Proof: Follows from the water-filling analogy since v finished with $\text{Level}(v)$ at the termination and w can only send information to its neighbors up to $\text{Level}(v)$ upon its arrival.

Partition Function $f(x) = e^{x-1}$

Let $f(x) = e^{x-1}$ for $0 \leq x \leq 1$

Let $w \in R$ and let $vw \in E$.

Let a 'small' amount dx of information flows on the edge vw when w arrives.

Uneven Split

Partition the increase $x_{vw} = dx$ among q_v and q_w by the function f as follows:

$$q_v = q_v + f(\text{Level}(v))dx$$

$$q_w = q_w + (1 - f(\text{Level}(v)))dx$$

Observation: The increase in the value of $q_v + q_w$ is dx .

If $\text{Level}(v) \approx 1$ then a large proportion of dx is assigned to q_v as

$$f(\text{Level}(v)) = e^{\text{Level}(v)-1} \approx 1$$

$\implies$ function f doesn't split x_{vw} evenly among q_v and q_w

Objective: Evaluate q values for all vertices in $L \cup R$ after the Waterlevel Algorithm has terminated.

Processing of $w \in R$ resulted in

Case 1: $\text{Level}(v) = \sum_{z \in R} x_{vz} = 1$ (v gets saturated)

Case 2: $\sum_{v \in L} x_{vw} = 1$ (w is completely drained)

Next we analyze the sum $q_v + q_w$ for both the cases:

Case 1: v is saturated

We know that $v \in L$ is saturated on the termination of the algorithm, i.e.

$$\text{Level}(v) = 1$$

During the course of the algorithm Level of vertex v increased from 0 to 1

$$\text{Thus for the edge } vw: q_v + q_w \geq q_v = \int_0^1 f(x)dx = \int_0^1 e^{x-1}dx = 1 - \frac{1}{e}$$

Case 2: w is drained out

- Vertex w has sent all of its information to its neighbors including v
- Suppose $\text{Level}(v) = X$, where $0 \leq X \leq 1$, at the termination of the algorithm.
- By the Key Observation we know that when w was sending information to its neighbors, their Level's were at most X .
- Thus using the fact that f is increasing (therefore, $1 - f$ is decreasing), we have

$$q_w \geq \int_0^1 (1 - f(X)) dx = (1 - e^{X-1}) \int_0^1 dx = 1 - e^{X-1}$$

$$\text{-Therefore, } q_v + q_w \geq \int_0^X f(x) dx + 1 - e^{X-1} = e^{X-1} - \frac{1}{e} + 1 - e^{X-1} = 1 - \frac{1}{e}$$

Analysis of Waterlevel Algorithm

Claim

Waterlevel algorithm is $1 - \frac{1}{e}$ -competitive for fractional bipartite matching.

Proof: For any edge $e = (vw) \in E$, $q_v + q_w \geq 1 - \frac{1}{e}$

Set $p_v = \frac{e}{e-1} q_v$ for all $v \in L \cup R$

All the constraints of the Dual LP are satisfied.

We know that $\sum_{e=vw \in E} (q_v + q_w) = |M|$ and the objective value of the Dual LP is an upper bound to the objective value of the Primal LP.

Optimal value of the Primal is the size of the optimal fractional matching M^* .

Thus, $\sum_{e=vw \in E} p_v + p_w = \frac{e}{e-1} \sum_{e=vw \in E} q_v + q_w = \frac{e}{e-1} |M| \geq |M^*|$

Equivalently, $\frac{|M|}{|M^*|} \geq 1 - \frac{1}{e}$

□

1. Richard Karp, Umesh Vazirani, and Vijay Vazirani, An optimal algorithm for on-line bipartite matching. ACM-STOC, 1990.
2. Nikhil Devanur, Kamal Jain, and Robert Kleinberg, Randomized primal-dual analysis of RANKING for online bipartite matching, ACM-SIAM SODA 2013.
3. Aranyak Mehta, Amin Saberi, Umesh V. Vazirani, Vijay V. Vazirani: AdWords and generalized online matching. Journal of ACM 54(5): 22 (2007)
4. Bala Kalyanasundaram, Kirk Pruhs: An optimal deterministic algorithm for online b-matching. Theoretical Computer Science 233(1-2): 319-325 (2000)