ORIE 633 Network Flows August 28, 2007

Lecture 2

Lecturer: Anke van Zuylen *Scribe: Kathleen King*

1 Applications of the maximum flow problem

This lecture discusses two applications of the maximum flow problem: baseball elimination and carpool fairness.

1.1 Baseball Elimination

Example. Consider the following baseball division with four teams.

Team	**Wins**	**Remaining Games**	**Games Against**			
			NY	Bos	Tor	Bal
New York Yankees	93	8	-	1	6	1
Boston Red Sox	89	4	1	-	0	3
Toronto Blue Jays	88	7	6	0	-	1
Baltimore Orioles	86	5	1	3	1	-

Definition 1 *A team* wins its division *if it wins more games than the other teams in the division. (Ties are okay; that is, more than one team may win its division.)*

Definition 2 *A team is* eliminated *if they can't finish first given any outcome of the remaining games.*

We want to know what teams have been eliminated already. Clearly Baltimore is eliminated because they will end the season with at most 91 wins while the Yankees already have 93 wins. It's less obvious that Boston has also been eliminated. It's true that Boston could still win 93 games, so it would seem that they could win provided that the Yankees lose the rest of their games. However, if the Yankees lose all of their games, this means that Toronto must win 6 games against the Yankees, giving Toronto the 94 wins they would need to beat Boston. Our goal is to determine a systematic way to solve problems like this.

Notation. Letting T denote the set of teams in the division, we adopt the following notation for each team $i \in T$:

w_i = number of wins for team i

g_i = number of games left to play for team i

g_{ij} = number of games left for team i to play team j.

For subsets $R \subseteq T$, we also define:

$$\begin{aligned}
w(R) &= \sum_{i \in R} w_i = \text{number of wins of the teams in } R \\
g(R) &= \sum_{i,j \in R, i<j} g_{ij} = \text{number of games to be played where both teams are in } R \\
a(R) &= \frac{w(R) + g(R)}{|R|}.
\end{aligned}$$

Claim 1 *Some team $i \in R$ wins at least $a(R)$ games.*

Proof: The total number of wins by all teams in R must be at least the total of their current wins plus the number of games played within set R, which is $w(R)+g(R)$. Therefore, the average number of wins by teams in R is $a(R)$, so some team must win at least $a(R)$ games. □

Corollary 2 *For a team $i \in T$ and any $R \subseteq T - \{i\}$, if $a(R) > w_i + g_i$, then team i is eliminated.*

Example. Let $R = \{\text{New York, Toronto}\}$ and $i = \text{Boston}$. Then $a(R) = \frac{(93+88)+6}{2} = 93.5 > 93 = 89 + 4 = w_i + g_i$. So Boston is eliminated, as we saw above.

Now let x_{ij} be the number of times team i defeats team j in the remaining games. Then team k is *not* eliminated if there exists $\{x_{ij}\}$ such that the following conditions hold:

$$\begin{aligned}
x_{ij} + x_{ji} &= g_{ij}, \quad \forall i, j \in T \\
w_k + \sum_{j \in T} x_{kj} &\geq w_i + \sum_{j \in T} x_{ij}, \quad \forall i \in T \\
x_{ij} &\geq 0, \quad x_{ij} \text{ integer}, \quad \forall i, j \in T
\end{aligned}$$

The first condition states that exactly one of team i or team j must win each game played between teams i and j. The second states that, at the end of the season, team k must have won at least as many games as any other team. The third simply guarantees that the x_{ij} must have nonnegative integer values.

If such x_{ij} exist, then team k is not eliminated, so it wins its division (possibly in a tie with another team). So then we could change some of the x_{kj} to make team k win all of its remaining games, and team k would still not be eliminated. This means we could find x'_{ij} to satisfy the following three criteria:

$$\begin{aligned}
x'_{ij} + x'_{ji} &= g_{ij}, \quad \forall i, j \in T & (1) \\
w_k + g_k &\geq w_i + \sum_{j \in T} x'_{ij}, \quad \forall i \in T & (2) \\
x'_{ij} &\geq 0, \quad x'_{ij} \text{ integer}, \quad \forall i, j \in T & (3)
\end{aligned}$$

We can create a network to determine whether team k is not eliminated. To do this, we create source and sink nodes, s and t; a node for every team $i \in T - \{k\}$; and a pair node for each $\{i, j\} \subseteq T - \{k\}$ with $i < j$ to avoid double counting. We make edges from the source s to the pair node $\{i, j\}$ and give these capacities of g_{ij}. We make edges from team node i to the sink t and give them capacities of $w_k + g_k - w_i$. We also create edges from each pair $\{i, j\}$ to teams i and j with infinite capacity. This is shown in Figure 1.

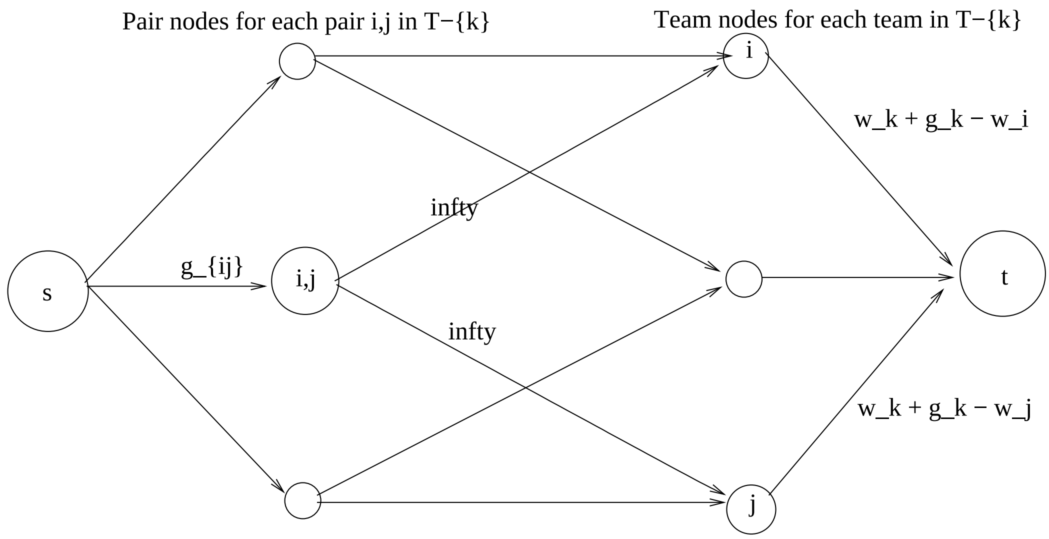

Figure 1: Flow instance for deciding if team k is not eliminated.

Note that we can assume that the capacities on the arcs going from the team nodes to the sink t are non-negative, since if $w_k + g_k - w_j < 0$, then $w_j > g_k + w_k$, and we know that team k is eliminated.

Now let $G = g(T - \{k\})$ be the sum of capacities on the arcs out of s, which gives the number of games to be played not involving team k. Then we can state the following lemma:

Lemma 3 *If a flow of value G exists, then team k is not eliminated.*

Proof: Notice that if a flow of value G exists, it must saturate all of the arcs out of s. Let x_{ij} be the flow from pair node $\{i, j\}$ to team node i. We want to show that the three conditions given above hold.

1. $x_{ij} + x_{ji} = g_{ij}$ is satisfied since the flow to pair node $\{i, j\}$ is g_{ij}, so flow conservation guarantees that the flow out, which is $x_{ij} + x_{ji}$ equals g_{ij}.

2. $\sum_{j \in T - \{k\}} x_{ij} \leq w_k + g_k - w_i$ is satisfied because of flow conservation and capacity constraints on arcs into t.

3. All of the x_{ij} are nonnegative, and we can assume they are all integers because of the integrality property of flow.

Hence we have found an appropriate set of x_{ij}, so team k is not eliminated. $\square$

Now we can show the opposite direction: if a flow of this value does not exist, then we can prove that the team is eliminated.

Lemma 4 *If a flow of value G does not exist, then team k is eliminated.*

Before proving this lemma, let us observe that a minimum cut may not contain a pair node $\{i,j\}$ unless both team nodes i and j are also in the cut. If either of these team nodes is not in the cut, then the cut has infinite capacity and so is not minimal. (For instance, the cut $\{s\}$ would have a lower capacity.)

Proof: Let S be a minimum $s-t$ cut. By our observation above, if a pair node $\{i,j\}$ is in S, then so are team nodes i and j. Let R be the set of team nodes in S and let P be the set of pair nodes in S. Then we can give the following expression for the capacity of S:

$$\begin{aligned} u(\delta^+(S)) &= \sum_{\{i,j\}\notin P} g_{ij} + \sum_{i\in R}(w_k + g_k - w_i) \\ &= G - \sum_{\{i,j\}\in P} g_{ij} + \sum_{i\in R}(w_k + g_k - w_i) \end{aligned}$$

We know that $\sum_{\{i,j\}\in P} g_{ij} \leq \sum_{i,j\in R, i<j} g_{ij} = g(R)$ since if $\{i,j\} \in P$ then both $i \in R$ and $j \in R$. Therefore

$$\begin{aligned} u(\delta^+(S)) &= G - \sum_{\{i,j\}\in P} g_{ij} + \sum_{i\in R}(w_k + g_k - w_i) \\ &\geq G - g(R) + |R|(w_k + g_k) - w(R) \end{aligned}$$

Also $u(\delta^+(S)) < G$, so we get

$$G - g(R) + |R|(w_k + g_k) - w(R) < G.$$

Subtracting G and rearranging gives

$$w_k + g_k < \frac{1}{|R|}(g(R) + w(R)) = a(R).$$

Hence k is eliminated by Corollary 2. $\square$

The following two lemmata will be proven in the first problem set.

Lemma 5 *If team k is eliminated, then for any team ℓ such that*

$$w_k + g_k \geq w_\ell + g_\ell$$

team ℓ is also eliminated.

Lemma 6 *$O(\log |T|)$ flow computations determine all eliminated teams.*

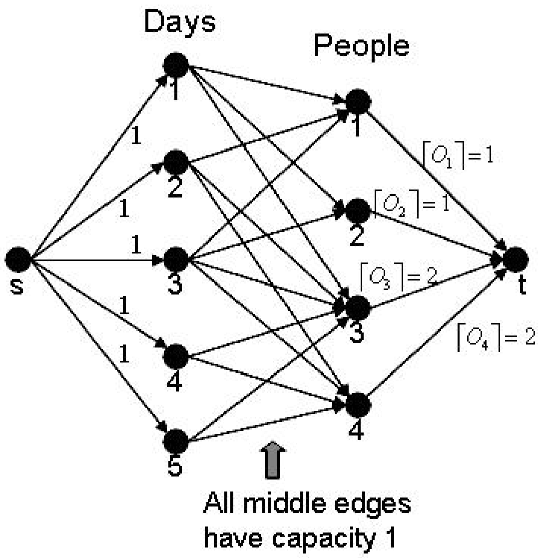

Figure 2: Flow instance for determining a fair carpool.

1.2 Carpool Fairness

Description: In this scenario, n people are sharing a carpool for m days. Each person may choose whether to participate in the carpool on each day.

Example. The following table describes a carpool in which 4 people share a carpool 5 days. X's indicate days when people participate in the carpool.

Person	**Days:** 1	2	3	4	5
1	X	X	X		
2	X		X		
3	X	X	X	X	X
4		X	X	X	X

Our goal is to allocate the daily driving responsibilities 'fairly.' One possible approach is to split the responsibilities based on how many people use the car. So, on a day when k people use the carpool, each person incurs a responsibility of $\frac{1}{k}$. That is, for each person i, we calculate his or her driving obligation O_i as shown below. We can then require that person i drives no more than $\lceil O_i \rceil$ times every m days. Table 1.2 shows the calculation of these O_i and their ceilings.

Person	**Days:** 1	2	3	4	5	O_i	$\lceil O_i \rceil$
1	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{4}$			1	1
2	$\frac{1}{3}$		$\frac{1}{4}$			$\frac{7}{12}$	1
3	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{7}{4}$	2
4		$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{19}{12}$	2
$\sum$	1	1	1	1	1	-	-

Table 1: Driver Responsibilities

To determine whether such an assignment is possible, we formulate the problem as a network, as shown in Figure 2.

We use this network to prove a claim for an m day carpool.

Claim 7 *If flow of value m exists, then a fair driving schedule exists.*

Proof: Note that all capacities are integer and if a flow of value m exists, then an integral flow of value m also exists. So, for each day, exactly one arc pointing outward has a flow of 1. This arc points to some person, and this is the person who should drive for the day. By flow conservation and the capacity of the arcs into t, no one will have to drive more than their obligation. $\square$

Note that we do not have to compute the maximum flow to conclude that there always exists a fair driving schedule.

Claim 8 *A flow of value m always exists.*

Proof: We can always give a fractional flow of value m, where each person present on a given day drives $\frac{1}{k}$ on a day when k people participate in the carpool. $\square$