Baseball Elimination

(hockey and basketball too)

Kevin Wayne

Princeton University

www.cs.princeton.edu/~wayne

The Problem

team i	wins w_i	losses l_i	to play r_i	against $= r_{ij}$ Atl	Phi	NY	Mon
Atlanta	83	71	8	–	1	6	1
Philly	80	79	3	1	–	0	2
New York	78	78	6	6	0	–	0
Montreal	77	82	3	1	2	0	–

Which teams have a chance of finishing season with most wins?

- Montreal eliminated since it can finish with at most 80 wins, but Atlanta already has 83
  - $w_i + r_i < w_j \Rightarrow$ team i eliminated
  - only reason sports writers appear aware of
  - sufficient reason for elimination, but not necessary

The Problem

team i	wins w_i	losses l_i	to play r_i	against $= r_{ij}$ Atl	Phi	NY	Mon
Atlanta	83	71	8	–	1	6	1
Philly	80	79	3	1	–	0	2
New York	78	78	6	6	0	–	0
Montreal	77	82	3	1	2	0	–

Which teams have a chance of finishing season with most wins?

- Philly can win 83, but still eliminated . . .
  - if Atlanta loses a game, then some other teams wins one

Answer depends not just on how many games already won and left to play, but also on who they're against.

History

- first popularized by Alan Hoffman in 1960's
- standard textbook application of optimization

Schwartz '66 determine whether one particular team is eliminated using 1 max flow

Hoffman-Rivlin '67 necessary and sufficient conditions for team to be eliminated from t-th place

Gusfield-Martel '92, McCormick '96 first-place elimination number for one particular team using 1 parametric max flow

Adler, Erera, Hochbaum, Olinick all first-place and playoff elimination numbers using LP – `riot.ieor.berkeley.edu`

Main Results

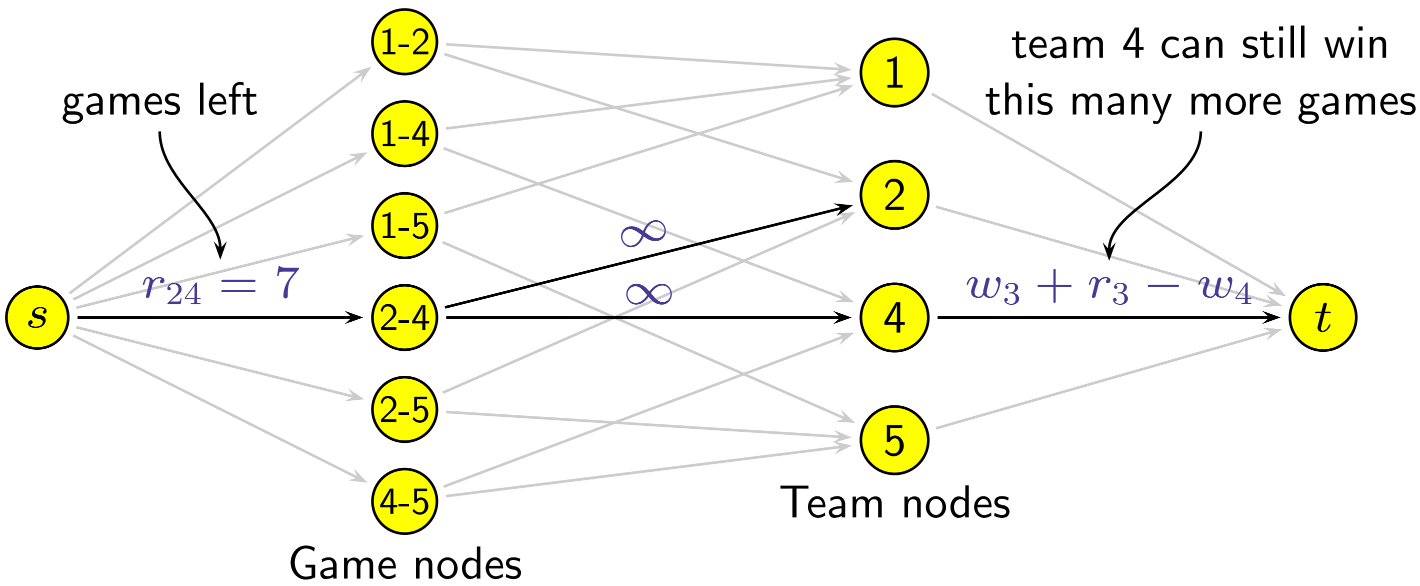
- Surprising structural property
  - problem not as difficult as many mathematicians would have you believe

- Find all eliminated teams and 1st-place elimination numbers
  - same asymptotic complexity as 1 max flow
  - previous approach: 1 separate max flow for each team
  - factor n speedup ($n = \#$ teams)

Classic Max Flow Formulation

Can team 3 finish with most wins?

- Assume team 3 wins all remaining games $\Rightarrow w_3 + r_3$ wins
- Divvy remaining games so all teams have $\leq w_3 + r_3$ wins



Theorem [Schwartz '66]: Team 3 not eliminated iff max flow saturates all arcs leaving source.

Reason for Elimination (for sports writers)

team i	wins w_i	losses l_i	to play r_i	against $= r_{ij}$ NY	Bal	Bos	Tor	Det
NY	75	59	28	–	3	8	7	3
Baltimore	71	63	28	3	–	2	7	4
Boston	69	66	27	8	2	–	0	0
Toronto	63	72	27	7	7	0	–	0
Detroit	49	86	27	3	4	0	0	–

AL East: August 30, 1996

- Detroit could finish season with $49 + 27 = 76$ wins
- Consider $R = \{$NY, Bal, Bos, Tor$\}$
  - have already won $w(R) = 278$ games
  - must win at least $r(R) = 27$ more

  $\Rightarrow$ average team in R wins at least $305/4 = 76\frac{1}{4}$ games

Always a Certificate of Elimination

$$R \subseteq N, \quad w(R) := \overbrace{\sum_{i \in R} w_i}^{\#\text{ wins}}, \quad r(R) := \overbrace{\tfrac{1}{2} \sum_{i,j \in R} r_{ij}}^{\#\text{ remaining games}}$$

$$a(R) := \frac{\overbrace{w(R) + r(R)}^{\text{LB on avg \# games won}}}{|R|}$$

If $a(R) > w_i + r_i$ then i is eliminated (by R)

Theorem [Hoffman-Rivlin '67]: Team i is eliminated iff $\exists\, R$ that eliminates i.

$\Rightarrow$ $R =$ team nodes on sink side of min cut

Surprising New Structural Property

team i	wins w_i	losses l_i	to play r_i	win % $w_i/(w_i+l_i)$	order $w_i + r_i$
Atlanta	83	71	8	.539	91
New York	78	78	6	.500	84
Philly	80	79	3	.503	83
Montreal	77	82	3	.481	80

Property: team j eliminated and $w_i + r_i \leq w_j + r_j \quad \Rightarrow$ i also eliminated.

[proved independently by Adler, Erera, Hochbaum, Olinick '98 using LP]

- we prove using flows and cuts $\Rightarrow$ faster algorithm
- extended by Gusfield-Martel to soccer, but no algorithm

Analysis and Consequences

Property: team j eliminated and $w_i + r_i \le w_j + r_j \quad \Rightarrow$ i also eliminated.

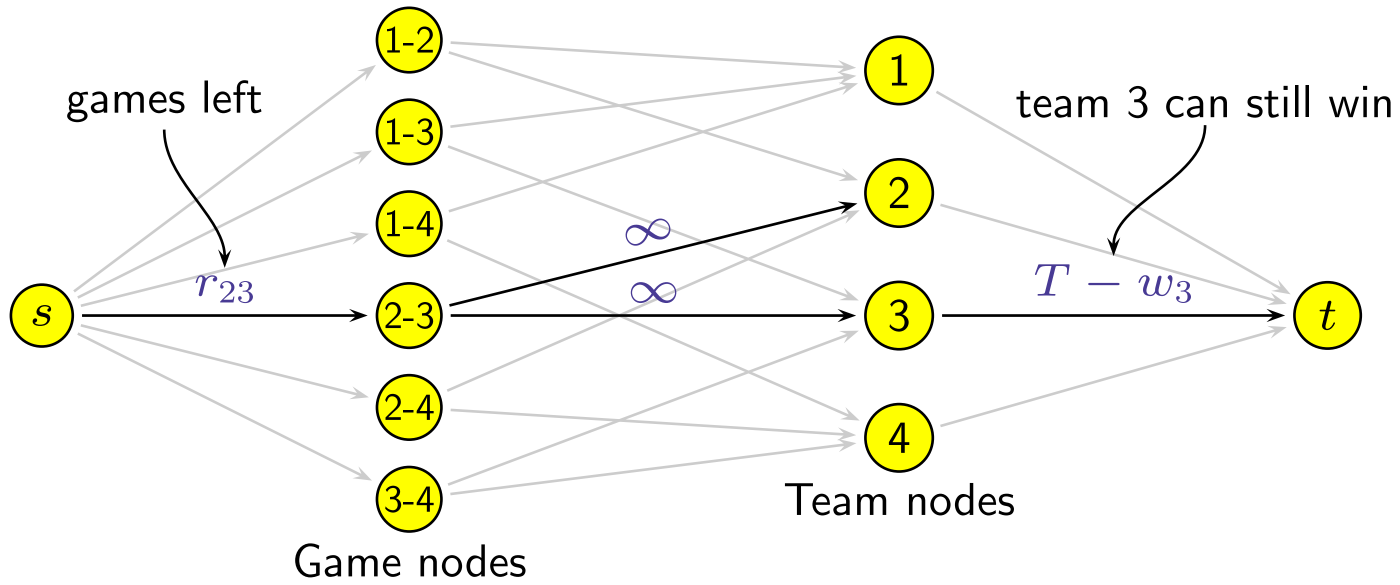
- j eliminated $\Rightarrow\ \exists\, R$ s.t. $a(R) > w_j + r_j \ge w_i + r_i$
- $i \notin R \Rightarrow R$ eliminates i trivially
- $i \in R \Rightarrow R \setminus \{i\}$ eliminates i

Corollary: $\exists\, T$ s.t. i eliminated $\iff w_i + r_i < T$.

- threshold T determines all eliminated teams
- can binary search among n values for threshold $T \Rightarrow$ can determine all eliminated teams with $\log n$ max flows
- use parametric max flow to find $T \Rightarrow$ same asymptotic complexity as 1 preflow-push max flow . . .

Parametric Max Flow Formulation

Find outcome in which no team finishes with more than T wins

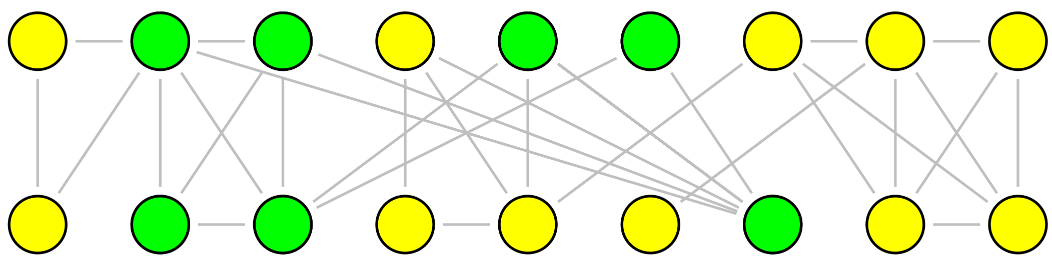


Goal: find smallest (fractional) value T for which max flow saturates all source arcs

- use [GGT '93] to find min cut for all values of T
- min cut for T^* gives R that maximizes $a(R) = \frac{w(R)+r(R)}{|R|}$

Special Case: Max Density Subgraph

Find subgraph whose ratio of number of internal arc to nodes is maximum.



- create baseball league with $w \equiv 0,\ r \equiv 1$
  $\Rightarrow$ max density subgraph problem

$$a(R^*) := \max_{R \subseteq N} \left\{ \frac{w(R) + r(R)}{|R|} \right\}, \quad w(R) = \sum_{i \in R} w_i,\ r(R) = \tfrac{1}{2} \sum_{i,j \in R} r_{ij}$$

Conclusions

- Find all eliminated teams in same complexity as 1 max flow
  - factor n speedup
- Threshold property for baseball elimination
  - elimination ordering: wins + remaining games
  - if teams have played same number of games, then "sportswriter ordering" is same as "elimination ordering"
- Dual of our problem is generalization of max density subgraph problem