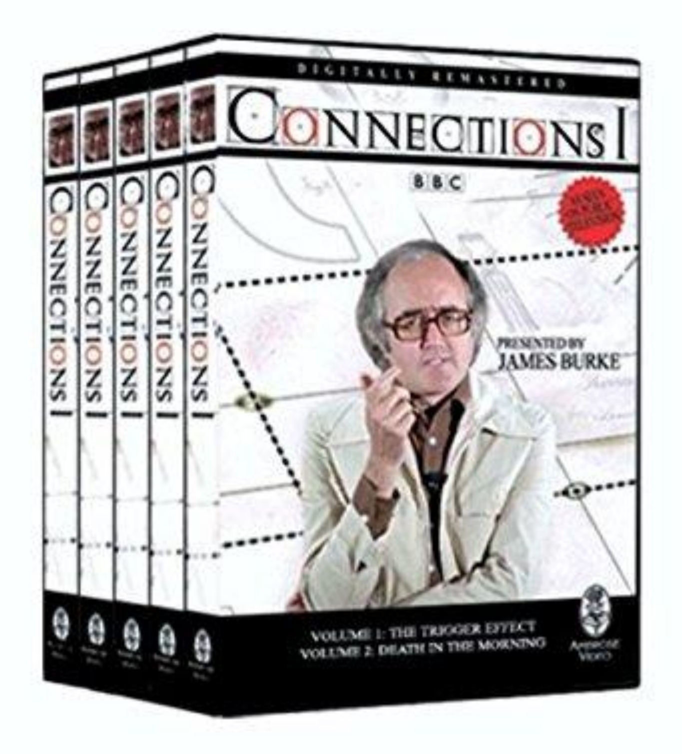
Ranking Algorithms!

Graphing Connections


DIGITALLY REMASTERED
CONNECTIONS I
BBC
PRESENTED BY JAMES BURKE
VOLUME 1: THE TRIGGER EFFECT
VOLUME 2: DEATH IN THE MORNING

Start: Senegal!

Senegal national football team

World Cup

MATCHES NEWS STANDINGS PLAYERS

Group H

	Team	MP	W	D	L	GF	GA	GD	Pts
1	Colombia	3	2	0	1	5	2	3	6
2	Japan	3	1	1	1	4	4	0	4
3	Senegal	3	1	1	1	4	4	0	4
4	Poland	3	1	0	2	2	5	-3	3

FIFA Tiebreakers

- Highest number of points
- Goal difference
- Goals scored
- Points obtained in group games between teams concerned
- Goal difference from games involving teams concerned
- Number of goals scored in games between teams concerned
- Fair play points
- Drawing of lots by FIFA

Spoiler - No good solution

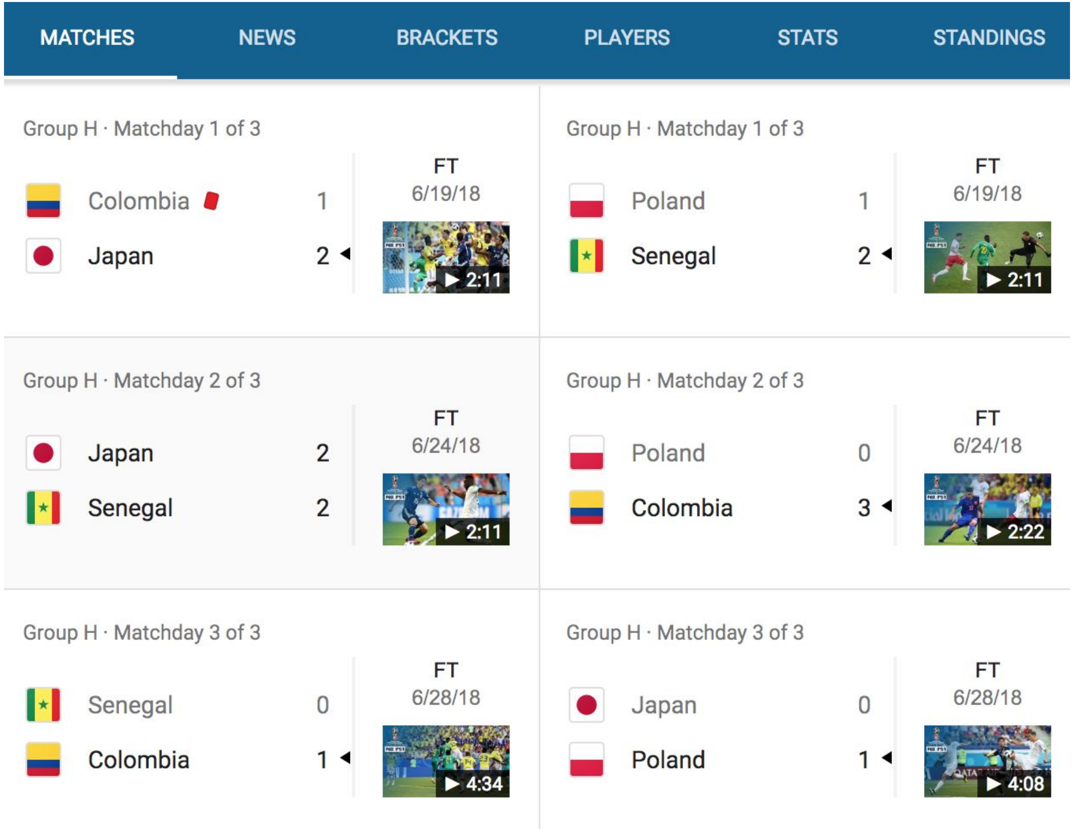
Arrow's impossibility theorem:

No ranked voting scheme meets a surprisingly modest set of criteria:

1. Universal (always works, regardless of number of voters/choices)
2. Fair and consistent (if everyone perfers a to b, ranking prefers a to b)
3. Non-dictatorial

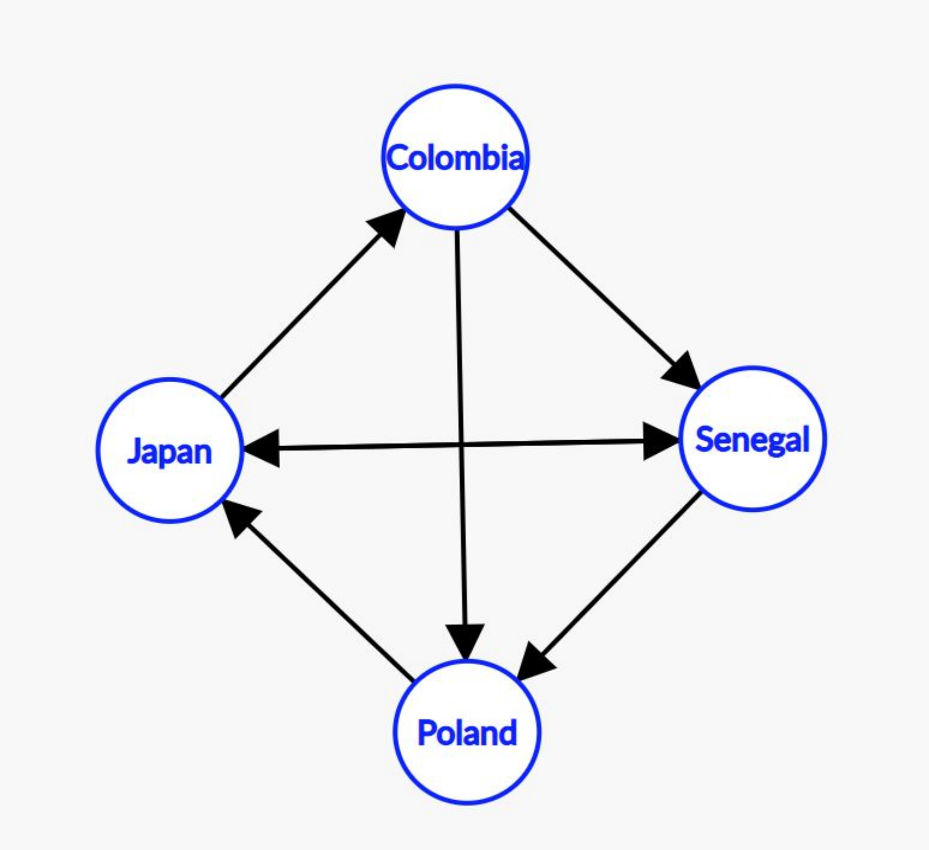
https://plato.stanford.edu/entries/arrows-theorem/

Mathematical representation: graph!



4 Teams
6 Games

Mathematical representation: graph!



Mathematical representation: Matrix!

	Colombia	Senegal	Poland	Japan
Colombia	-	1	1	-1
Senegal	-1	-	1	0
Poland	-1	-1	-	1
Japan	1	0	-1	-

Many applications

- Sports -> Tournament
  - Nodes = Teams, Vertices = Wins
- Google -> Rank pages
  - Nodes = Pages, Vertices = Inbound links
- Twitter -> Rank influential people in social graph
  - Nodes = Users, Vertices = Follow (Facebook=undirected graph)
- Biology
  - Nodes = Patients, Vertices = Who caught an STD from whom

2 Ranking Schemes!

ELO - Arpad Elo:

$$E_a = \frac{1}{1 + 10^{(R_b - R_a)/400}}$$

$$E_b = 1 - E_a$$

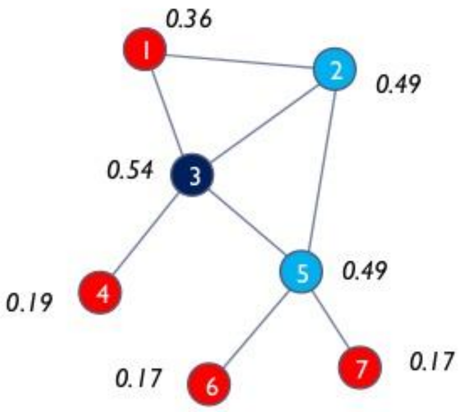
http://stephenwan.net/thoughts/2012/10/02/elo-rating-system.html
https://en.wikipedia.org/wiki/Elo_rating_system

“The Social Network”

Eigenvector centrality -> PageRank

Eigenvector centrality

- A node's eigenvector centrality is proportional to the sum of the eigenvector centralities of all nodes directly connected to it
- In other words, a node with a high eigenvector centrality is connected to other nodes with high eigenvector centrality
- This is similar to how Google ranks web pages: links from highly linked-to pages count more
- Useful in determining who is connected to the most connected nodes



Node 3 has the highest eigenvector centrality, closely followed by 2 and 5

Note: The term 'eigenvector' comes from mathematics (matrix algebra), but it is not necessary for understanding how to interpret this measure

Values computed with the sna package in the R programming environment. Definitions of centrality measures may vary slightly in other software.

PageRank: The $100 Billion Eigenvector

You have a billion pages to rank: it's a very large matrix.

Enter the damping factor:

Damping factor [edit]

The PageRank theory holds that an imaginary surfer who is randomly clicking on links will eventually stop clicking. The probability, at any step, that the person will continue is a damping factor *d*. Various studies have tested different damping factors, but it is generally assumed that the damping factor will be set around 0.85.[5] In applications of PageRank to biological data, a Bayesian analysis finds the optimal value of *d* to be 0.31.[23]

https://www.rose-hulman.edu/~bryan/googleFinalVersionFixed.pdf

https://en.wikipedia.org/wiki/PageRank

Questions?