

Authentic performance in opposition networks

Does opportunity meet merit in UFC ?

Pierrick Leroy, Marc Santolini

2023

Table of Contents

- 1 Introduction
- 2 Rating systems
- 3 Opportunity
- 4 Relation between rating and opportunities
- 5 Extensions

Table of Contents

- 1 Introduction
- 2 Rating systems
- 3 Opportunity
- 4 Relation between rating and opportunities
- 5 Extensions

Opposition networks

Network modeling opposition

Key Idea

May the best win ?

Opposition networks

Network modeling opposition

Key Idea

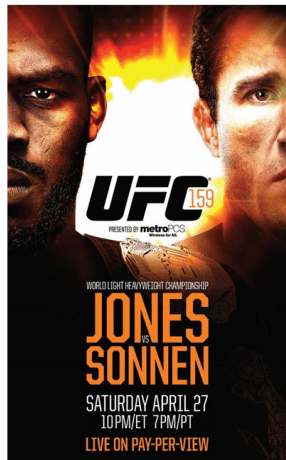
May the best win ?



Leroy, Santolini



Merit in opposition networks



2023

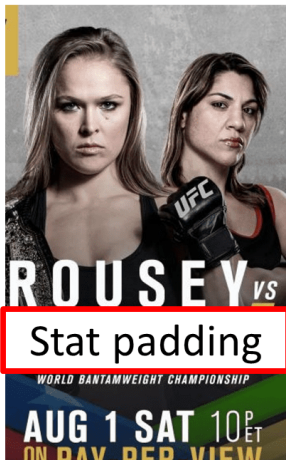
4 / 33

Opposition networks

Network modeling opposition

Key Idea

May the best win ?



Stat padding

Leroy, Santolini



Star priviledge

Merit in opposition networks



Claim to fame

2023

4 / 33

Opposition networks

Research questions

"Does opportunity meet merit ?"

- ① How worthy of each other are opponents ?
- ② How to disentangle performance and success ?

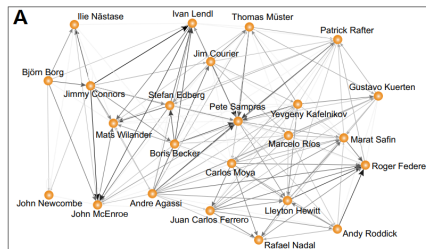
Opposition networks

Contact networks

"Network of contacts"¹ or "Network of adversaries"

Examples :

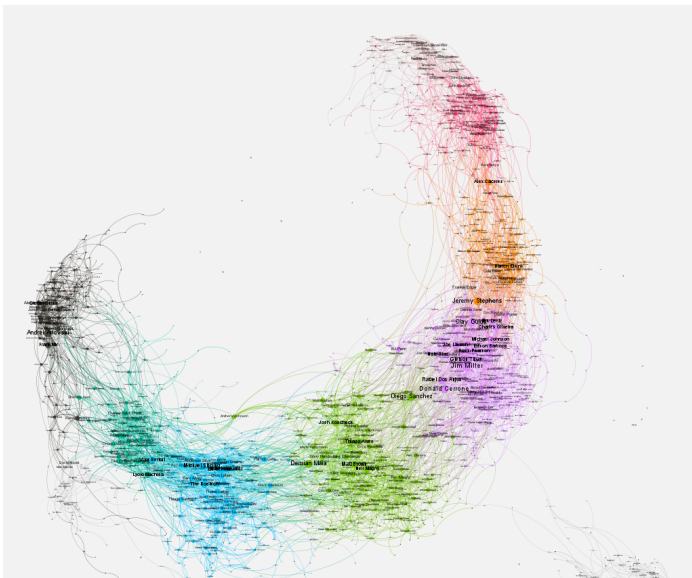
- Chess
- Tennis
- Fighting sport



¹Filippo Radicchi. "Who Is the Best Player Ever? A Complex Network Analysis of the History of Professional Tennis". In: *PLoS ONE* 6.2 (Feb. 9, 2011), e17249. arXiv: 1101.4028 [physics].

Opposition networks

Contact networks - UFC case



Opposition networks

Contact networks - UFC case

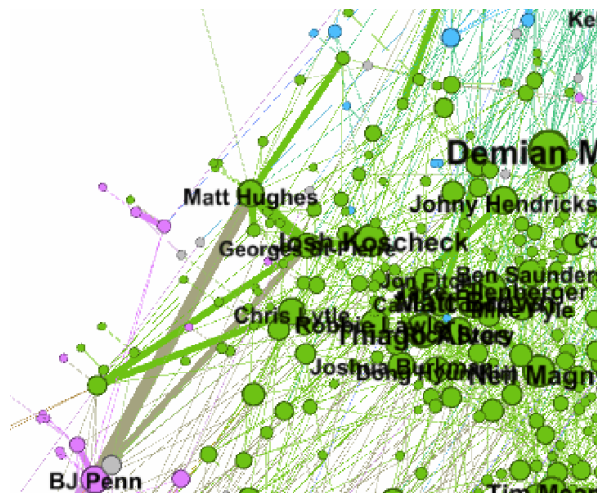


Figure: Zoomed in

Opposition networks

Contact networks - UFC case

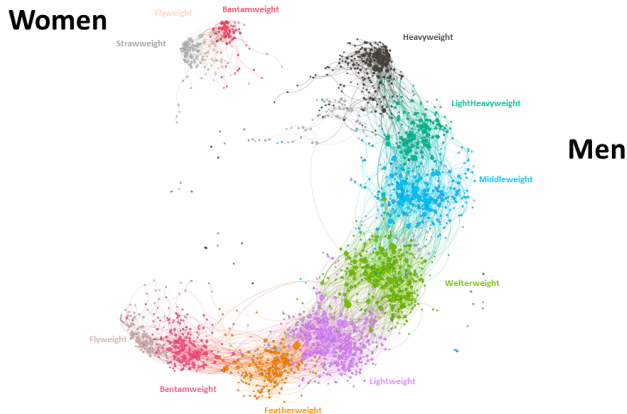


Figure: Clusters are weight classes

Opposition networks

Research questions

Research question 1.1

How worthy of each other are opponents ?

Contacts are created differently in different networks

Tennis

- Player chooses tournament
- Objective : win

Fighting sports

- Player chooses opponents
- Objective : get a title shot

Choosing one's opponent leads to confusing phenomena like *stat-padding* or *storytelling*.

Opposition networks

Research questions

Research question 1.2

How to disentangle performance and success ?

Performance can be relative.

Track-and-field

- Player are evaluated indiv.
- Performance doesn't depend on opponent
- Objective : produce best score

Tennis

- Players against each other
- Performance depends a lot on opponent
- Objective : beat opponent

Being evaluated within opposition complexifies the assessment of individual performance.

Table of Contents

- 1 Introduction
- 2 Rating systems
- 3 Opportunity
- 4 Relation between rating and opportunities
- 5 Extensions

Traditional methods : ELO, Glicko but require a lot of matches

Contact networks : flavored/heuristic PR in tennis²³, snooker⁴ or NBA⁵

Actual rankings : too complicated⁶ or judge based⁷

²Radicchi, “Who Is the Best Player Ever?”

³London, Németh, and Nemeth, “Time-Dependent Network Algorithm for Ranking in Sports”.

⁴O'Brien and Gleeson, “A Complex Networks Approach to Ranking Professional Snooker Players”.

⁵Shi and Tian, “Learning to Rank Sports Teams on a Graph”.

⁶*THE INFAMOUS BOXREC RANKING FORMULA EXPLAINED IN DETAIL* — Ringside Boxing News.

⁷Staff, *How Do the UFC Rankings Work?*

Methodology

- ① Study error cases of pagerank inspired by real life situations
- ② Formalize necessary conditions for a better solution
- ③ Design rating systems based on judges identification
 - Reverse PR on trimmed networks
 - Walk Index⁸ inspired (ongoing)
- ④ Evaluate the rating systems
 - On simple synthetic data
 - On real world datasets
 - Against each other and vs older approaches
 - On a prediction task : direction of link given history
- ⑤ Application : relation with success
 - Define an opportunity measure
 - Study the correlation between success and performance on UFC data

⁸Razin, Verbin, and Cohen, *On the Ability of Graph Neural Networks to Model Interactions Between Vertices*.

Proposed approaches

Problem illustration

Key Idea

find *credible judges* before ranking

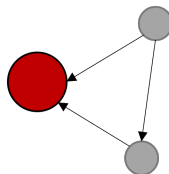
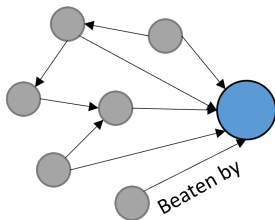
Proposed approaches

Problem illustration

Key Idea

find *credible judges* before ranking

Blue is the best, but is it ?

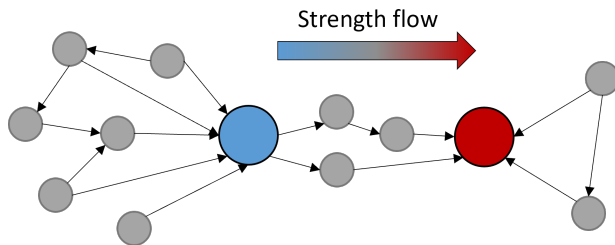


Proposed approaches

Problem illustration

Key Idea

find *credible judges* before ranking



Proposed approaches

2 approaches

How to identify the judges ?

- 1 Triple PageRank on *trimmed network*
- 2 Walk Index - model interaction levels

Proposed approaches

Baseline

A simple function based only on wins, losses and draws:

$$f_{\alpha,t}(w, l, d) = \frac{w - \alpha l + \frac{d}{2}(1 - \alpha)}{w + l + d + t}$$

with:

- t : offset ie experience-competence threshold
- α : severity of losses
- w : number of wins
- l : number of losses
- d : number of draws

Proposed approaches

Approach 1 : triple PageRank

Triple pageranks

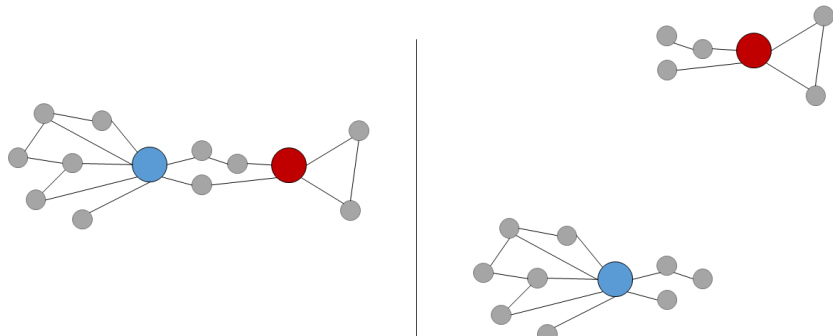
- 1 undirected PR from blue
- 2 undirected PR from red
- 3 directed PR from judges

Proposed approaches

Approach 1 : triple PageRank

Triple pageranks : decomposition into *trimmed networks* ie *networks where opponent has been removed*

- 1 undirected PR from blue
- 2 undirected PR from red
- 3 directed PR from judges

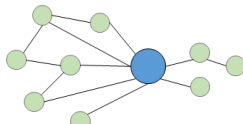
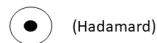
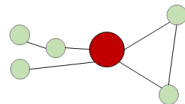
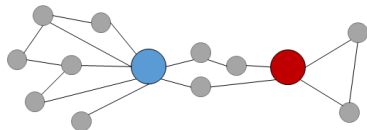


Proposed approaches

Approach 1 : triple PageRank

Triple pageranks

- 1 undirected PR from blue
- 2 undirected PR from red
- 3 directed PR from judges

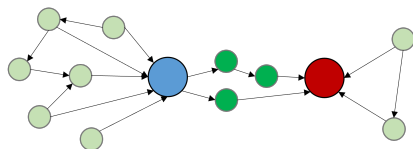
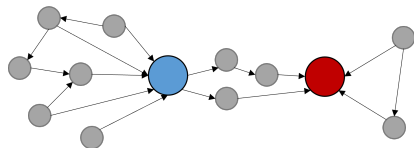


Proposed approaches

Approach 1 : triple PageRank

Triple pageranks

- 1 undirected PR from blue
- 2 undirected PR from red
- 3 directed PR from judges



Proposed approaches

Approach 1 : triple PageRank - parameters

Ego

Models the competing nodes "own consciousness" during final PR from judges

- Ego is scaled on the highest values of the judges
- Ego=1 means the competing nodes have the same decision power as the strongest judge (seeds are equal)
- Ego=0 means the competing nodes have no intrinsic decision power (seeds = 0)

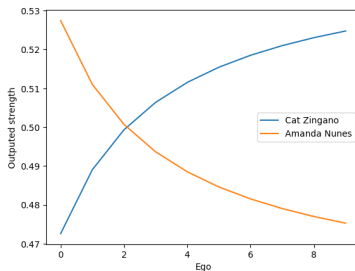


Figure: An upset was the victory of Zingano against Nunes. Ego parameter helps us adjusting the importance of direct confrontations.

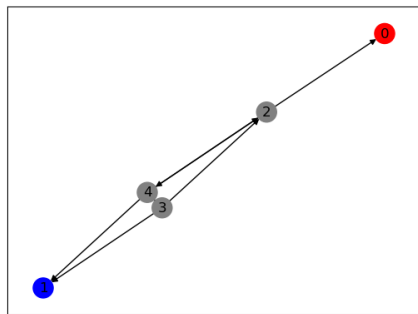
Proposed approaches

Approach 1 : triple PageRank - parameters

Experience-competence

Another parameter (ongoing) should allow to decide what to value more between *experience* and *competence*, where a node experience is related to the number of contacts and its competence is related to the strength of its opponents.

On the figure, it means we should be able to adjust a parameter to output 0 (experienced) or 1 (competent) as the best nodes.



Proposed approaches

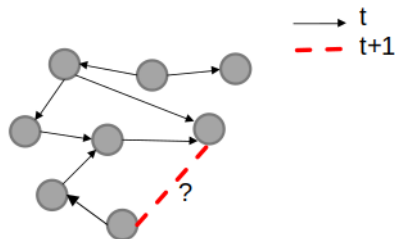
Prediction task

Given history, predict the outcome of matches

<i>Method</i>	<i>Acc</i>
PR	0.5062
PR (CPR subset)	0.5114
CPR	0.5263
PR with recency*	0.5238
PR with recency* (CPR subset)	0.5304
CPR with recency*	0.5426

*With recency, link weight w decrease over time as a strictly decreasing function (here $w \propto a^t$ with t in years and a a parameter with $a \leq 1$)

Prediction task



What will be the direction of the red arrow ?

Proposed approaches

Prediction task

The different method don't perform well. It may be because the system is dynamically complex :

- Ageing of nodes (strength increases and decreases)
- Autoregulative matchmaking (winners against winners)
- Multidimensional strength (styles)

If the results are not striking enough to prove scientific value:

- 1 Find another task were there is more contrast between PR and CPR
- 2 Design a synthetic problem

Table of Contents

- 1 Introduction
- 2 Rating systems
- 3 Opportunity**
- 4 Relation between rating and opportunities
- 5 Extensions

Table of Contents

- 1 Introduction
- 2 Rating systems
- 3 Opportunity
- 4 Relation between rating and opportunities
- 5 Extensions

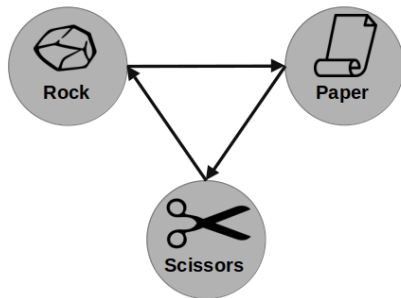
Table of Contents

- 1 Introduction
- 2 Rating systems
- 3 Opportunity
- 4 Relation between rating and opportunities
- 5 Extensions**

Extensions

Node characteristics

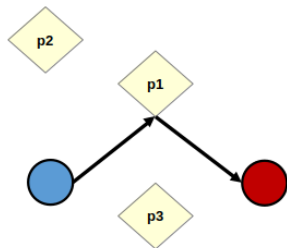
Characteristics of some nodes might give them an edge against some other type of nodes. In fighting sports, we can think of styles that are more or less effective against other styles. This can be modeled with node features.



Extensions

Resource allocation

Resource allocation might be framed as a bipartite graph where nodes compete for prizes of different values. This framework could be applied to domains such as tournaments (sports as in tennis or games like poker) or the job market of offers and candidates. Reinforcement learning might be introduced in a dynamic scenario. This scenario would probably have a lot in common with the prisoner dilemma (highest gain vs guaranteed gain).



Legend



Prize of value p_i

$i < j \Rightarrow p_i \geq p_j$



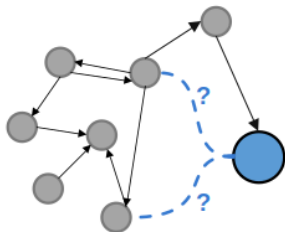
Compete for a resource

Explanation : blue and red competed for the resource with the highest prize. Blue won the resource and red didn't

Extensions

Reverse problem (RL)

Matchmaking role : considering the opposite point of view, an agent (node) may try to maximize a success metric by defeating the best opponents it can while avoiding losses. In this framework, the agent might or might not be aware of its own strength or other's.



*Which opponent should blue compete against next ?
What would be the best policy ?*

A natural question to ask when comparing two nodes in a network might be : "is there enough (shared) information to decide ?"

Idea : find a way to quantify uncertainty of judges

A natural extension to temporal graphs because recency is a factor that is particularly meaningful in numerous real world network contexts :

- Careers, be it in sports or not
- Scientific papers and their time dynamic⁹

A model incorporating a notion of time like in¹⁰ would therefore be interesting to study.

⁹Wang, Song, and Barabási, “Quantifying Long-Term Scientific Impact”.

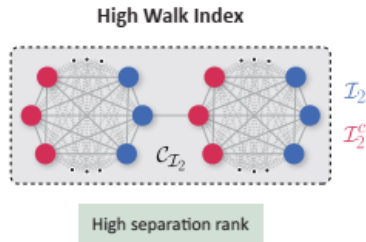
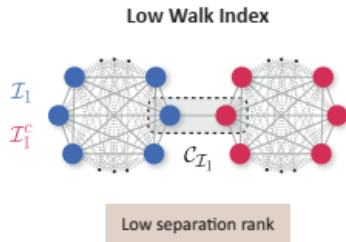
¹⁰London, Németh, and Nemeth, “Time-Dependent Network Algorithm for Ranking in Sports”.

GNN inspiration ?

Proposed approaches (Extension - ongoing)

Approach 2 : walk index

Walk index : quantify the quality of the boundary



Proposed approaches

Open questions

Experience-competence tradeoff (cf notebook) - the solution lacks a parameter