Authentic performance in opposition networks

Does opportunity meet merit in UFC ?

Pierrick Leroy, Marc Santolini

2023

Table of Contents

Table of Contents

Key Idea

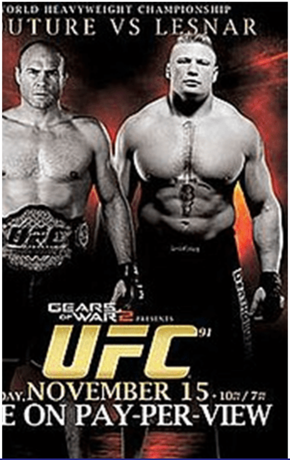
May the best win ?

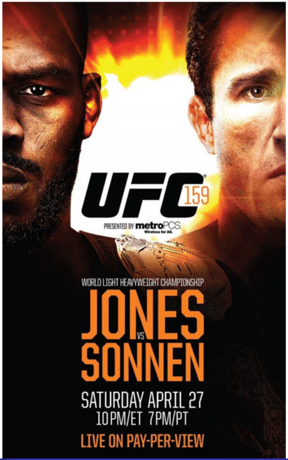
Opposition networks

Network modeling opposition

Key Idea

May the best win ?



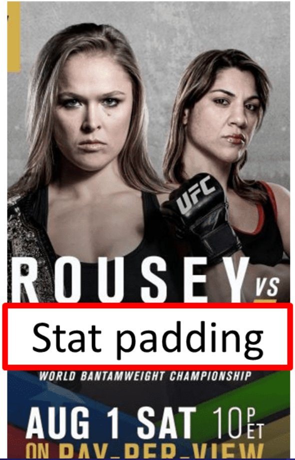


Opposition networks

Network modeling opposition

Key Idea

May the best win ?



Stat padding

Star priviledge

Claim to fame

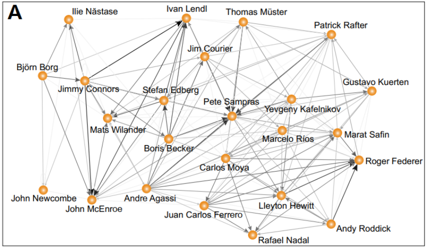
"Does opportunity meet merit ?"

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

"Network of contacts"[1] or "Network of adversaries"

Examples :

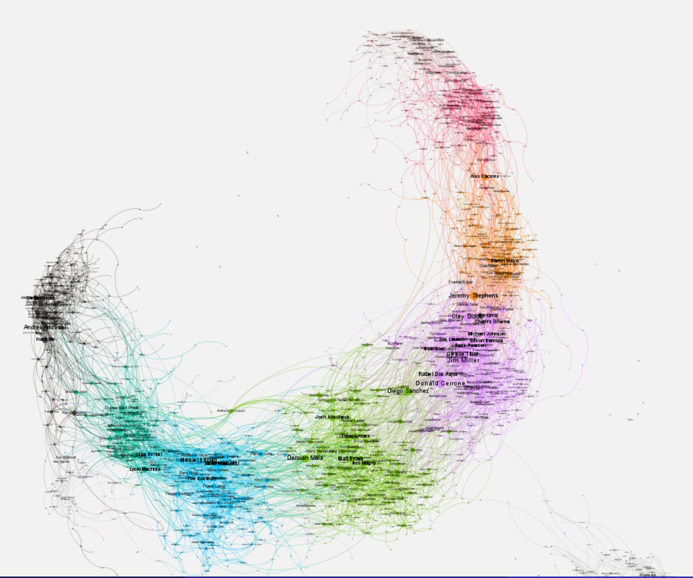
- Chess
- Tennis
- Fighting sport



[1] Filippo 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

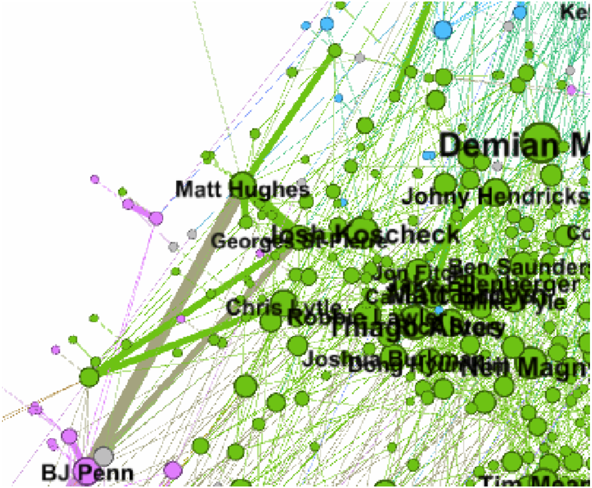


Figure: Zoomed in

Opposition networks
Contact networks - UFC case

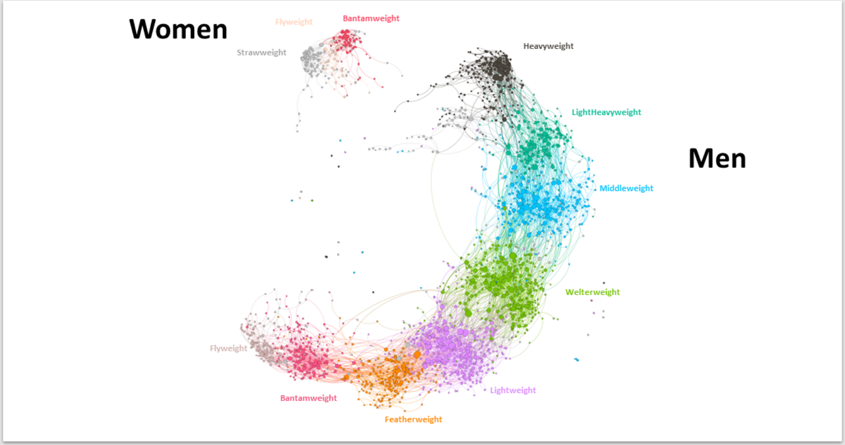


Figure: Clusters are weight classes

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*.

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

Literature

Traditional methods : ELO, Glicko but require a lot of matches
Contact networks : flavored/heuristic PageRanks in tennis[2][3], snooker[4] or NBA[5]
Actual rankings : too complicated[6] or judge based[7]

---

[2] Radicchi, "Who Is the Best Player Ever?"

[3] London, Németh, and Nemeth, "Time-Dependent Network Algorithm for Ranking in Sports".

[4] O'Brien and Gleeson, "A Complex Networks Approach to Ranking Professional Snooker Players".

[5] Shi and Tian, "Learning to Rank Sports Teams on a Graph".

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

[7] Staff, *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 Index[8] 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

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

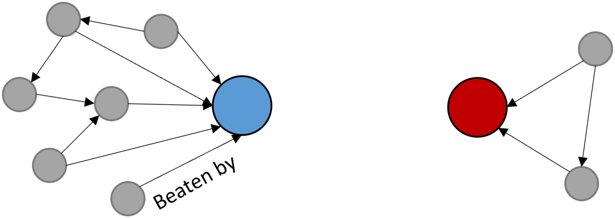
Key Idea

find *credible judges* before ranking

Proposed approaches
Problem illustration

Key Idea
find *credible judges* before ranking

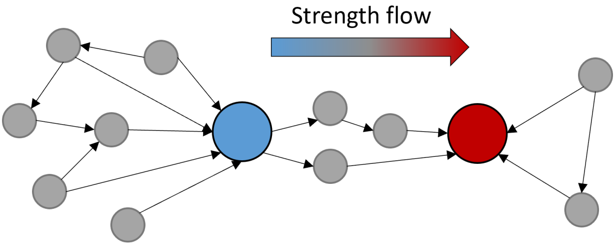
Blue is the best, but is it ?



Proposed approaches
Problem illustration

Key Idea

find *credible judges* before ranking



How to identify the judges ?

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

Baseline rating function

$f_{baseline} = w - l$ with:

- w : number of wins
- l : number of losses

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

Parameters α and t can be optimized on data.

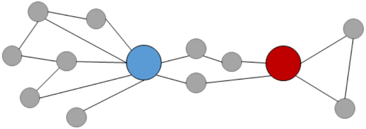
Triple pageranks

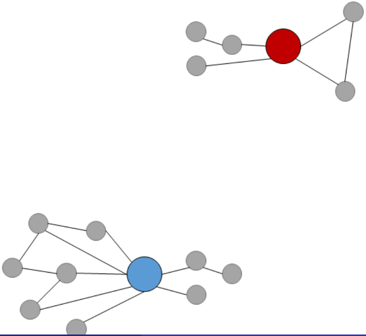
1. undirected PR from blue
2. undirected PR from red
3. directed PR from judges

Proposed approaches
Approach 1 : triple PageRank

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

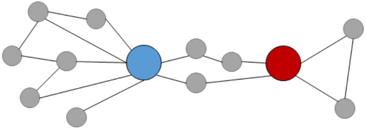
1. undirected PR from blue
2. undirected PR from red
3. directed PR from judges

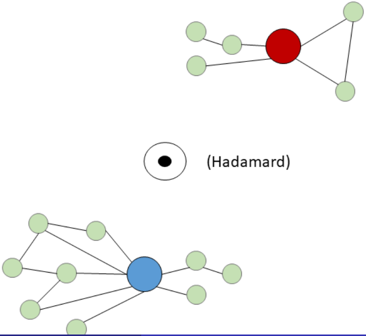




Triple pageranks

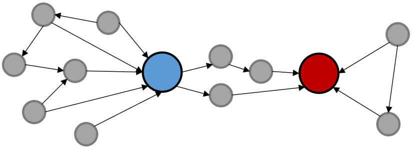
1. undirected PR from blue
2. undirected PR from red
3. directed PR from judges

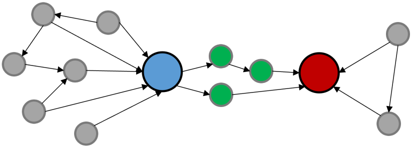




Triple pageranks

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)

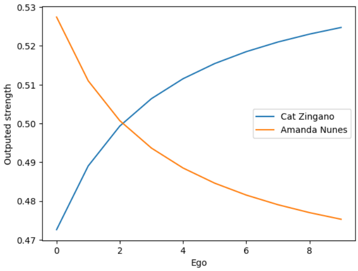
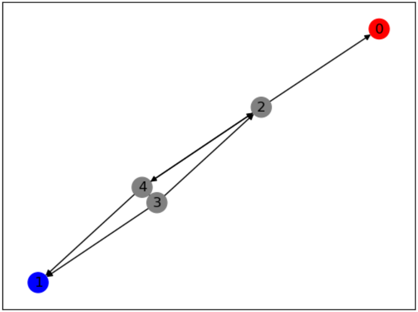


Figure: An upset was the victory of Zingano against Nunes. Ego parameter helps us adjusting the importance of direct confrontations.

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 node.



Prediction task - with temporality

Edge direction prediction (winner prediction)

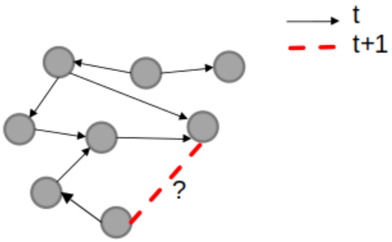
[label=prediction$_{w}ith_{t}emp$]

Given history, predict the outcome of matches

Method	*Acc*
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

Prediction task



What will be the direction of the red arrow ?

Prediction task - with temporality

Edge direction prediction (winner prediction)

The different methods 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

Prediction task

Edge direction prediction (winner prediction) - $f_{baseline}$

For every fight, rate each fighter with $f_{baseline}$ and predict the highest rated as the winner.

- $f_{baseline} = w - l$
- When rating is close, outcome is uncertain (diagonal)
- Gradient of win percentage orthogonal to the diagonal

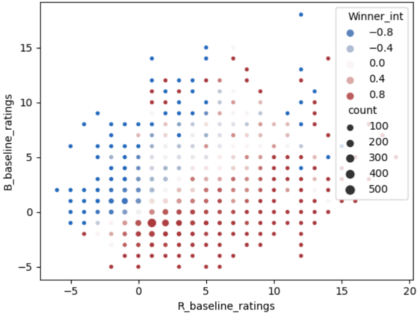


Figure: Average win percentage for pairs of baseline ratings

Prediction task

Edge direction prediction (winner prediction) - $f_{\alpha,t}$

For every fight, rate each fighter with $f_{\alpha,t}$ and predict the highest rated as the winner.

- $f_{\alpha,t}(w,l,d) = \frac{w-\alpha l+\frac{d}{2}(1-\alpha)}{w+l+d+t}$
- Parameters α and t are optimized on a train set using gradient descent and evaluated on a test set

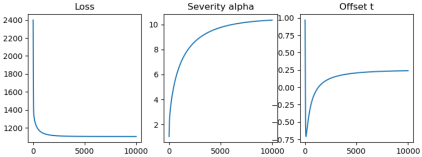


Figure: Convergence of parameters

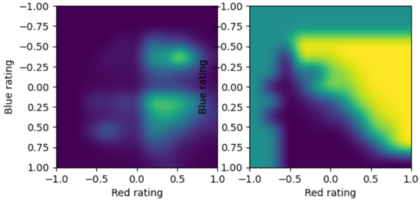


Figure: Density (left) and

- Input : 6 dimensions
- Optimized by gradient descent
- Not a rating function ie there is no map from the input space of fighters to $\mathbf{R}$ nor to a space with an order relation so it cannot be a ranking function

Prediction task

Edge direction prediction (winner prediction) - MLP_{rater}

An MLP adapted to rate fighters:

- Optimized by gradient descent
- Siamese network with one input for each fighter's features mapped to a scalar value
- Tradeoff : drop in prediction performance but this MLP rater is a rating function (maps input space to $\mathbf{R}$)

Prediction task

Edge direction prediction (winner prediction) - PR and CPR

Same process as on this slide but without the temporality ie all links are present

Prediction task

Edge direction prediction (winner prediction)

Recap

Method	*Acc*
$f_{baseline}$	0.800
$f_{\alpha,t}$	0.844
*MLP**	0.878
MLP_{rater}	0.847
PR	0.663
CPR	0.665

*not a rating function

Synthetic dataset
Network construction

Motivation : easier to analyze
Characteristics :

- 100 nodes
- 10 links per node on average
- hidden strength variable sampled from uniform distribution
- no ageing
- no draws
- no styles impact

Synthetic dataset

Network construction

2 matchmaking (link generation) schemes:

- random erdos-renyi model
- random with strength aware rewiring and degree preservation (nodes are matched against nodes with comparable hidden strength)

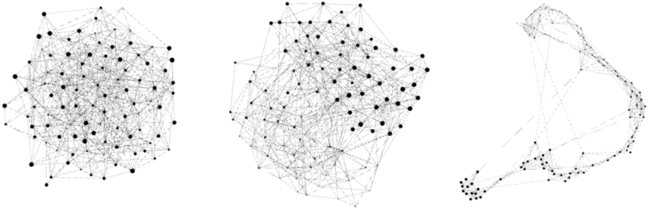
If there is a link between v_1 and v_2 its direction is from the weaker node (lowest hidden strength) to the strongest node (highest hidden strength)

Fraud is also introduced through 2 schemes:

- dumb fraudster: rearranges a proportion of its edges by selecting weaker opponents
- strength aware fraudster: rearranges a proportion of its edges by selecting slightly weaker opponents

Synthetic dataset

Strength rewiring



Random scheme | Strength rewiring (light) | Strength rewiring (strong)

Figure:

Synthetic dataset

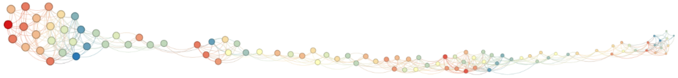


Figure: On strength aware network
Top : $f_{baseline}$ (delta degrees)
Bottom : PageRank

Table of Contents

Table of Contents

Table of Contents

Applications

Voting

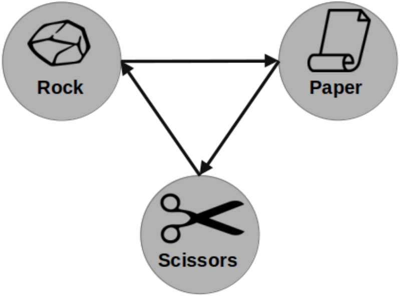
- Decide who has a claim for a given opportunity based on credible judges , e.g. a position in a company, a title shot in a championship...
- The structure of the network makes the process interpretable

Fraud detection

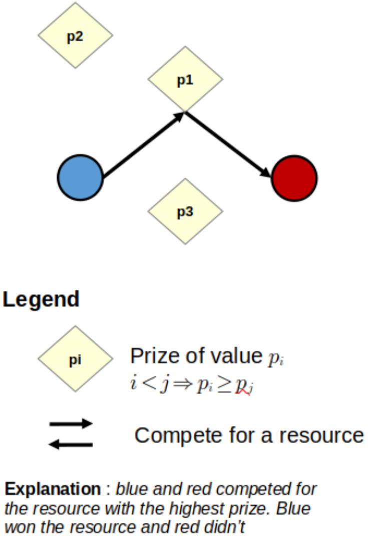
- By analysing discrepancies with another ranking
- Identify peculiarities in a network

Table of Contents

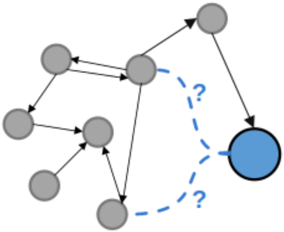
Characteristics of some nodes might give them an edge against some other type of nodes. In fighting sports, we can think of styles that are more or less effective against other styles. This can be modeled with node features.



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).



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

Extensions

Temporal aspect

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 dynamic[9]

A model incorporating a notion of time like in[10] would therefore be interesting to study.

[9] Wang, Song, and Barabási, "Quantifying Long-Term Scientific Impact".

[10] London, Németh, and Nemeth, "Time-Dependent Network Algorithm for Ranking in Sports".

GNN inspiration ?

Walk index : quantify the quality of the boundary