Convex Optimization Project

Malo Ferriol

May 2020

Abstract

This project aims to present a new ranking solution for soccer. The french league is organized as a round robin tournament. However this year in 2020 the season could not be completed due to the global pandemic caused by the Covid-19. We developed a new ranking system that will try to lessen the impact of the missed fixture.

1 Introduction

The ranking of a soccer league is based on an accumulative system. When a team win a game it is rewarded with 3 points. When two teams draw they are rewarded with 1 point. When a team looses it is not rewarded. Regardless of the content of a game the reward is always similar. Whether a team score 5 or 6 goals, the reward will only depends on the score. At the end of the season the ranking is based on the total of points. Their is fairness in this system since every teams plays against one another twice. One game is conducted on one team home ground and the second on the other team home ground. This year the french league Ligue 1 has decided to put the safety of everyone and to put an end to this season prematurely.

2 Problem statement

The accumulative ranking system is fair once the tournament is over. Before every teams has played against one another their is always a discrepancy since not all teams are equal. Based on this observation we decided to create a ranking system that would take the strength of the opponent in consideration in order to make a fairer ranking. We have decided to more feature that simply the outcome of the game. We want to take into consideration : the number of goals conceded and scored in a game, the number of loss, draw and games against one opponent.

3 Solution design

3.1 The problem representation

3.1.1 Graph representation

The tournament is an unfinished round robin tournament. The matrix representation of the tournament is not full rank. We wanted to avoid having to use a representation with missing values. We decided to make a graph representation of the tournament. This gave us more freedom for the implementation and the data structure could easily be adapted.

3.1.2 Pairwise comparison

Each team performance is measured by its performance during confrontation. We want to rank the teams their performance based on their result and their opponent. In our graph representation we have decided that each confrontation will be represented as a double directed edge. Each directed edge will represent the performance of the teams. The edges are assigned a weight. We use the concept of "source" and "target" nodes. The score of an edge from a source to a target represent the performance of the target team against the source team.

3.2 The methods

From the graph representation of our tournament we want to extract the information that will enable us to rank each team. The nodes of the graph are the teams and the edge their performance against one another.

3.2.1 Eigenvector centrality

The Eigenvector centrality is measure of the influence of each nodes on the graph (network). Each nodes has a score in relation with the rest of the nodes in the graph. A node with a higher has more influence than a node with a lower score. The scores are link to each other with the following concept, a connection to a node with a higher score will be more relevant than the same connection to a nodes with a lower score. Iteration after iteration we assign the to each node the sum of the score of its connected node until the result converge. This is equivalent to calculate the largest eigenvalue with its eigenvector of the adjacency matrix.

We use this concept in our ranking. The edges are measures of the performance of the target node.

3.2.2 PageRank

The PageRank algorithm is a variant of the Eigenvector centrality. This algorithm use the normalize influence of each node. It also adds a dumping factor, to "represent the random of a walker in the network".

At first the dumping factor seams like a wrong usage. Why add some link randomly between teams since this would not represent their performance.

Actually after some thought we came to the realisation that this dumping factor have a signification. Our intuition behind this factor is that the factor represent the randomness that is part of the game. In soccer since the result can be a victory for a team of a draw there is often better team that have unexpected result against smaller team.

For example the PSG which has a budget of more than 600 M in a leagues where more than half of the team have less than 50M is expected to win every game. However they actually did not win the league in 2016/2017.

3.3 The solution for Directed Graph

We implemented two graph representation. The first is a Directed graph. Each edge is a sum of the of the weight of the performance of team against another.

3.3.1 The network design

The network is composed of 20 nodes and 380 edges. This is the number of pairwise representation of teams $n * (n - 1)$ where n is the number of teams here 20.

3.3.2 Our weight algorithm

We decided to aggregate our feature with our own weight definition. The weight of a game is as follow :

$$w = (l/g) + (d/2) + c/(c + s) \tag{1}$$

where

w is the weight

l is the number of defeat by the team of the source node

g is the number of game played against this opponent

d is the number of draw between the teams

c is the number of goal conceded by the team of the source node

s is the number of goal scored by the team of the source node

3.4 Multi Directed Graph Representation

The second is a Multi Directed graph. Each edge is the representation of a the performance of team against another. This is new design allow us to use each game performance separately.

3.4.1 The network design

The network is composed of 20 nodes and 540 edges. This is the number of confrontation multiplied by two. Each game has a pair of edge.

3.4.2 Centrality in multi-graph

In this new design the notion of centrality is altered. The graph cannot be represented by a single matrix since a team can have multiple confrontation against the same opponent.

Our method use the centrality score of multiple feature separately and then perform an aggregation to obtain the new score for each team.

To use the sum of the weight of multiple feature to get a single matrix and calculate the centrality score is equivalent of the sum of the weight on a directed graph.

3.4.3 Our weight algorithm

The score of a team is as follow :

$$t = l + (d/2) + c/(c + s) \tag{2}$$

where

t is the centrality score of a node

l is the centrality score of a node for the graph representation of the defeat of a team

d is the centrality score of a node for the graph representation of the draw of a team

c is the centrality score of a node for the graph representation of the goal conceded of a team

s is the centrality score of a node for the graph representation of the goal scored by a team

4 Ranking result

In the following table we present the result from our models.

In order to be able to compare the result of the different model we normalized them using the min max normalization.

All our models are predefined by us to combine the different feature. There is human influence. The solution is an heuristic solution.

Solved by the Power method

Team	centrality_score
Paris St Germain	1.0000
Marseille	0.6457
Lille	0.5748
Rennes	0.5531
Reims	0.5336
Monaco	0.5105
Nantes	0.4485
Bordeaux	0.4424
Nice	0.4400
Lyon	0.4256
Montpellier	0.4187
Strasbourg	0.3726
Metz	0.3630
Brest	0.3549
Angers	0.3505
Dijon	0.3253
Saint Étienne	0.2768
Amiens	0.2682
Nîmes	0.2662
Toulouse	0.0000

Solved by the PageRank method

Team	centrality_score
Paris St Germain	1.0000
Reims	0.7357
Rennes	0.6424
Marseille	0.6128
Monaco	0.6122
Lille	0.5622
Nantes	0.5174
Dijon	0.4906
Bordeaux	0.4714
Nice	0.4702
Montpellier	0.4094
Lyon	0.4093
Amiens	0.3944
Metz	0.3810
Brest	0.3683
Strasbourg	0.3650
Angers	0.3261
Nîmes	0.2981
Saint Étienne	0.2724
Toulouse	0.0000

Figure 1: Result from our directed graph

The two methods yields different result.

The power methods has a great score difference between the first and second team. Where as the PageRank methods reduce that gap. The power methods obtains better result than the PageRank. The team Marseille has won 16 of its 27 game. They also did not obtain good result against the other good team. However their result are better than the one of a team such as Reims whose record is good when it comes to conceding goals with only 21 it is the best defense. The PageRank methods is giving too much importance to the goal difference compared to the result.

Solved by the Power method

Team	centrality_score
Paris St Germain	1.0000
Marseille	0.6481
Lyon	0.6229
Reims	0.6127
Monaco	0.5794
Rennes	0.5759
Bordeaux	0.5718
Montpellier	0.5411
Lille	0.5363
Nice	0.4824
Nantes	0.4513
Dijon	0.4331
Amiens	0.3971
Strasbourg	0.3860
Angers	0.3572
Brest	0.3437
Nîmes	0.3216
Metz	0.2769
Saint Étienne	0.1729
Toulouse	0.0000

Solved by the PageRank method

Team	centrality_score
Paris St Germain	1.0000
Marseille	0.7146
Reims	0.6410
Lyon	0.6065
Rennes	0.6030
Lille	0.5609
Bordeaux	0.5583
Monaco	0.5341
Montpellier	0.5328
Nice	0.4717
Nantes	0.4437
Strasbourg	0.4338
Dijon	0.4030
Brest	0.3909
Angers	0.3703
Metz	0.3369
Amiens	0.3030
Nîmes	0.2886
Saint Étienne	0.2576
Toulouse	0.0000

Figure 2: Result from our Multi Directed graph

The two methods yields different result.
Both methods well classified the team Marseille as second.

Using the centrality of each feature separately get more interesting feature. We can compare the performance of each team to the performance of their opponent on the desired feature. The scoring performance is not altered by the outcome of the game.

The Power methods is not as good as the PageRank who in my opinions obtains a result closer to the real ranking.

5 Conclusion

The ranking of sports team can be done by using the graph representation of their confrontation. We used different graph and solving algorithm to find heuristically a better ranking system. Those system suffer from the human influence. The next step would be to calculate the versatility of the multi directed graph to use as a centrality measure.