

Introduction

Developing an accurate and robust ranking system that is applicable for all sports would be ideal, but due to the differences in sports it's a hard task to achieve.

- Using PageRank as a base algorithm, we are able to develop a ranking system that allows us to predict future results across all sports.
- However, this will only provide us with a low prediction accuracy due to it not being sport specific.
- PageRank will need to be modified to include sport (in this case netball) specific features to provide accurate predictions.

The Base Model

Using an adapted PageRank model, I created a base model to form a basic rating system for all sports.

- First we construct a network showing the outcomes of matches (see figure 1 below). A directed graph is used to show the outcomes of the matches, with the arrow indicating who won.
- The base model is then constructed by forming an adjacency matrix summarising the outcomes of the matches over the course of a season.
- In the base model adjacency matrix each team is awarded a value of α (where $\alpha \in (0.5, 1)$) for each match that they win, a value of 0.5 for each match they draw and a value of $1 - \alpha$ for each match they lose.
- This matrix is then adjusted to make it a stochastic transition probability matrix so that it can now be solved using the power method.
- Mathematically, given the transition probability matrix P, the rating of team j can be written as:

$$\pi_j = \sum_{i=1}^n p_{ij} \pi_i$$

- From this equation, it can be seen that wins over stronger teams will increase your rating more than wins over weaker teams. [1]

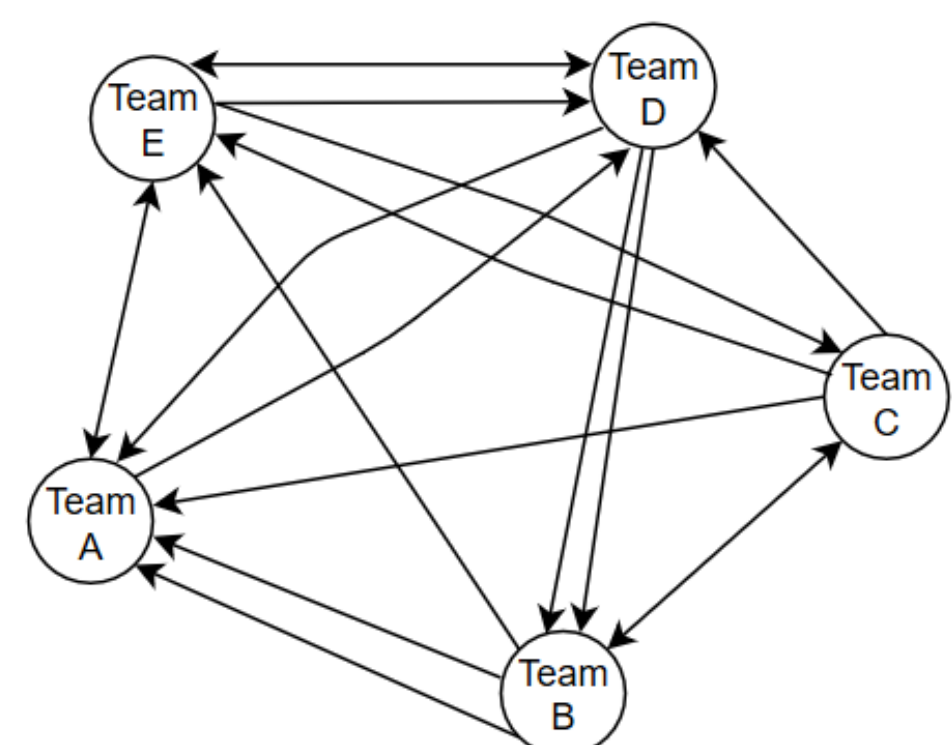


Figure 1: Network Example

Training the Model

Once a general framework for the PageRank model had been established, I then trained it on netball and football data:

- Vitality Netball 2018 & 2019
- 2018-19 Football National League

These data sets provide a good base level for the prediction accuracy of the PageRank algorithm, before I then develop the base model to include sport specific features to see if I can improve the accuracy of the PageRank algorithm for that sport (in this case netball).

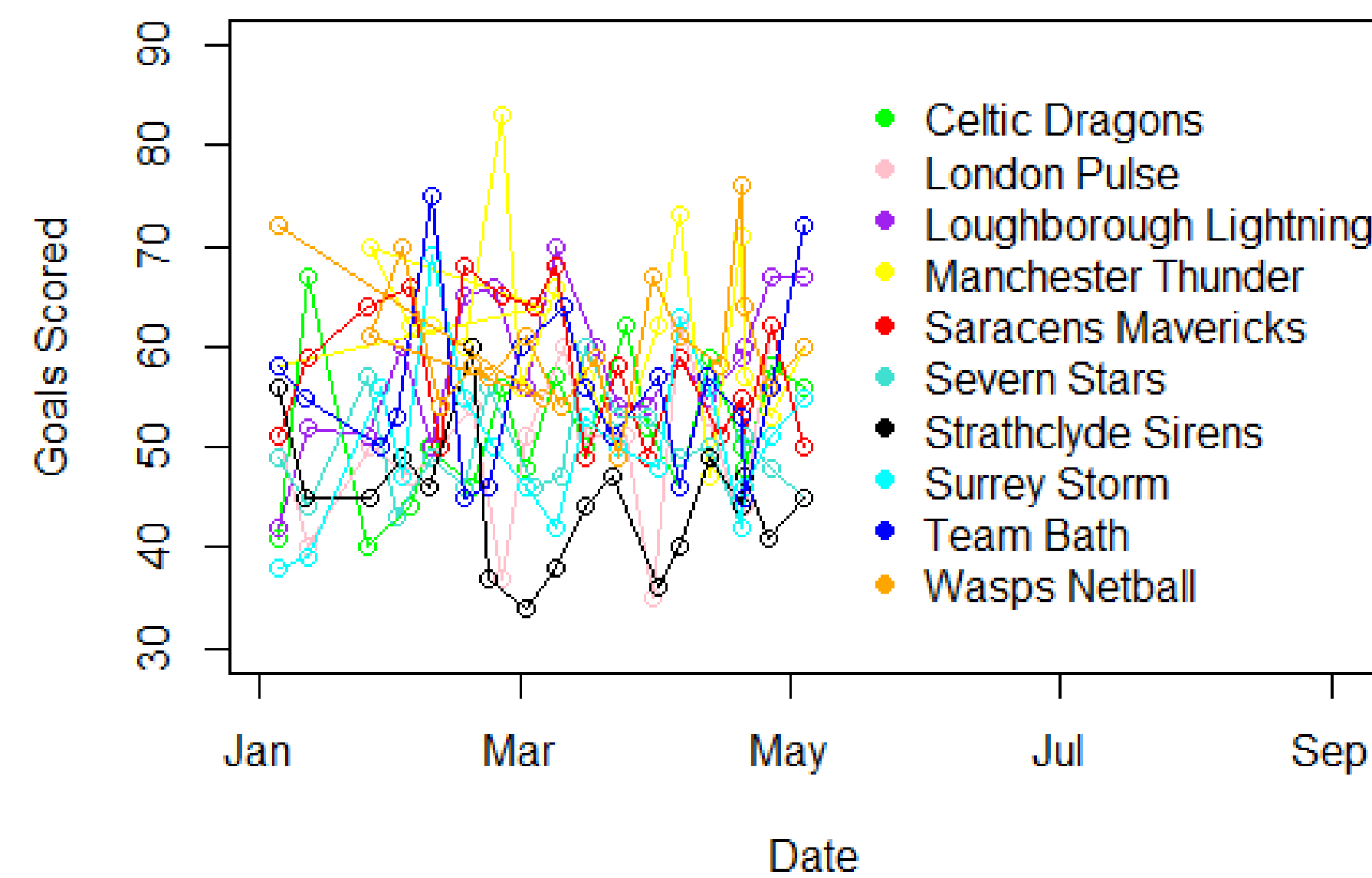


Figure 2: Vitality Netball 2019 Season

Idea Refinement

To develop the PageRank I began looking at netball's sport specific features, such as

- Home advantage
- Incentive to win
- Player injuries
- Margin of victory
- Winning streak

To incorporate margin of victory, for example, an indicator function was used in the base model adjacency matrix. The equation used to implement this was:

$$\lfloor (|G1 - G2| - \delta) \beta \rfloor + \zeta + 1$$

where goal difference is $|G1 - G2|$, δ is the lower bound for the range of goals, β is the weight given to that range of scores and ζ is the number of points already awarded prior to that goal range.

$$\begin{aligned} \text{For } 10 \leq |G1 - G2| \leq 26, & \quad \beta = \frac{1}{4} \quad \delta = 10 \quad \zeta = 1 \\ \text{For } 26 < |G1 - G2| \leq 32, & \quad \beta = \frac{1}{3} \quad \delta = 26 \quad \zeta = 5 \\ \text{For } 32 < |G1 - G2|, & \quad \beta = \frac{1}{2} \quad \delta = 32 \quad \zeta = 7 \end{aligned}$$

Results

Comparing the PageRank's predictive performance against existing benchmarks and bookmakers odds we can see how accurate the PageRank is.

- If the home team is taken as the favourite to win every match, the accuracy of this provides a good base level to see how the accuracy of the PageRank compares.
- The accuracy of taking the bookmakers favourite every time provides a good upper bound for the data.
- With regards to the football data, 44% the home team wins, the bookmakers accuracy is about 48% and the PageRank is at 46% so in the middle, where expected.
- Accuracy for the netball data was a lot higher then anticipated (see table 1 below), both with and without sport specific features.

	Vitality, 2018	Vitality, 2019	Australia, 2019
Without	0.622	0.667	0.577
With	0.656	0.656	0.596

Table 1: Accuracy of PageRank

After training the model, I put it to the test against the Australian Suncorp Super Netball 2019 League.

- For 4 consecutive weeks I ran my code against the bookmakers as a comparison. Out of the 4 matches a week, invariably I wouldn't agree with the bookmakers on at least one, PageRank was always correct on these occasions.

Conclusion and Future Enhancement

- It appears that the bookmakers have many more variables being taken into account for football data then they do for netball data. This explains why their accuracy for predicting the football is much better compared to PageRank then it is for the netball.
- Currently, the PageRank algorithm is only designed to work for one-on-one matches (such as netball, tennis or football) but I would like to adapt it such that it can be applicable to sports that are not one-on-one, for example rowing where a crew races a multitude of teams.

References

- [1] B. Vaziri. Markov-based ranking methods. 2016.