Graph-Based Modeling and Prediction of Match Outcomes in International Football Tournaments

Author: Francis Osei
Supervisor: Prof. Dr. Christophe Ley

Faculty of Science and Technology
Department of Mathematics
oseifrancis633@gmail.com

Table of Contents

1. Motivation
2. Ranking Football Teams
3. Rating of Football Matches
4. Model Comparison
5. Conclusion

Motivation

- Machine learning and statistical models have distinct methodologies.
- They often encounter similar challenges When applied to football prediction

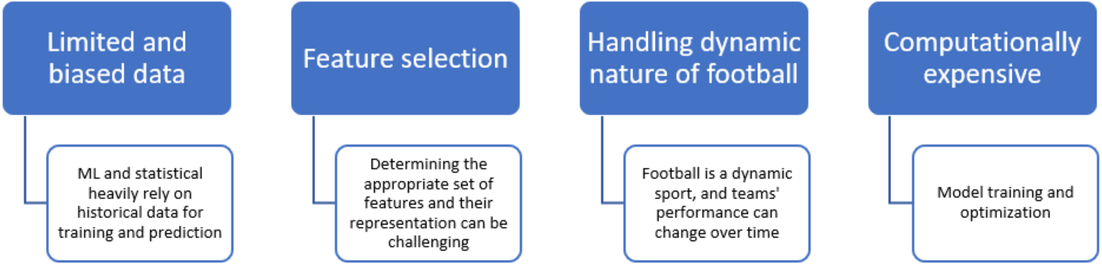


Figure 1: Challenges in ML and statistical models

Motivation

- Hybrid model demonstrated superior performance in forecasting football match outcomes.

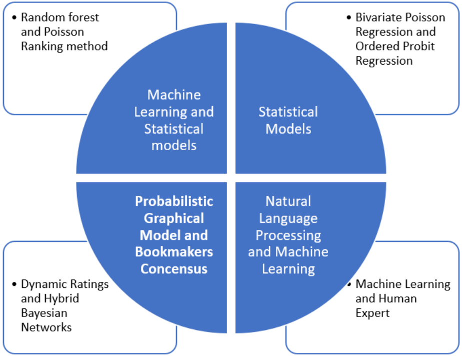


Figure 2: Examples of hybrid models

Objective

- To address the potential problems involved in football predictions, a network-based method using

  1. Eigenvector centrality.
  2. Random walk.

- The model consider **FIFA club ranking score** of each football club as a key factor

Preliminary

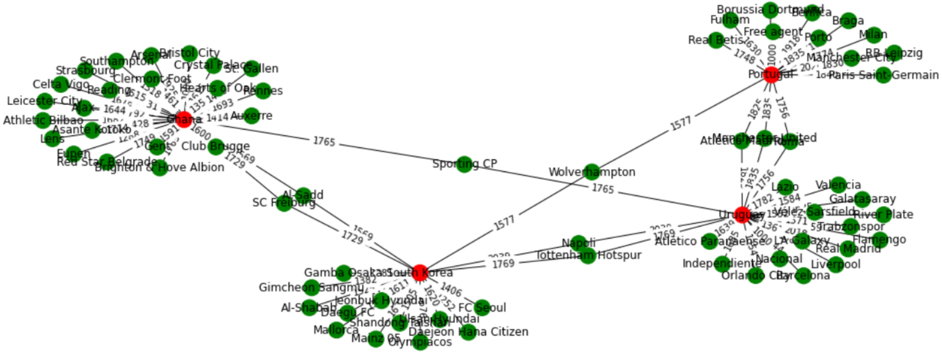


Figure 3: Graphical Representation of the Interconnections Among Group F Teams in the 2022 World Cup

An Overview of Graphs in Football

- A graph $G = (V, E)$ consists of two sets of information:
  1. Set of nodes $V = \{v_1, v_2, \cdots v_l\}$
  2. Set of vertices $E = \{e_1, e_2, \cdots e_n\}$
- In this framework, the nodes are partitioned into two disjoint sets
  1. $\Gamma = \{x_1, x_2, \cdots, x_{l_1}\}$ (football clubs)
  2. $\Omega = \{y_1, y_2, \cdots, y_{l_2}\}$ (national teams)

  such that $\Gamma \cup \Omega = V$ and $l_1 + l_2 = l$.
- Edges represent relationship between national teams and football clubs

Node Degree

- Mathematically, the degree of a node v_i is computed as

$$d(v_i) = |\{v_j : (v_i, v_j) \text{ is an edge}\}, \ i,j = 1,2,3\cdots,l| \quad (1)$$

- Node degree analyzes **diversity's** impact on football team performance in tournaments.

- Once the node degree $d(v_i)$ surpasses a specific threshold, it starts to negatively impact the team's performance (Ignacio (2022)) .

UNIVERSITÉ DU LUXEMBOURG

Node Degree

Teams	Year	Position	Node degree
Italy	2006	winner	10
France	2006	runner-up	16
Spain	2010	winner	9
Netherlands	2010	runner-up	17
Germany	2014	winner	11
Argentine	2014	runner-up	15
France	2018	winner	15
Croatia	2018	runner-up	21
Argentine	2022	winner	18
France	2022	runner-up	16

Table 1: Comparison of Winners and Runners-up in the 2006 to 2022 World Cups

Drawbacks of Node Degree for Teams Ranking

- The node degree may not be enough to fully describe the connection of the network in a weighted network.

  1. The strength of the connections is often represented by weights.
  2. We only consider the amount of connections a national team has with football clubs

Eigenvector Centrality

- Eigenvector centrality measures node importance based on connections to other significant nodes.
- In the context of football prediction, eigenvector centrality can be used to determine the most influential teams in a tournament.
- Mathematically, eigenvector centrality assigns a score x_i to each team $i \in V$ denoted as:

$$x_i = \beta \sum_{j \in V:(i,j) \in E} x_j = \beta \sum_{j \in V} A_{ji} x_j \tag{2}$$

where A_{ij} represent the adjacency matrix of a weighted graph, where the entries correspond to the edge weights.

Eigenvector Centrality

- Let us consider a tournament comprising three teams, where each team is represented by a total of six players.

Clubs	Score
A	300
A	300
B	500
B	500
C	700
G	350

(a) Nation Team X

Clubs	Score
C	700
C	700
B	500
B	500
B	500
H	670

(b) National Team Y

Clubs	Score
E	310
E	310
F	270
G	350
A	300
D	250

(c) National Team Z

Eigenvector Centrality

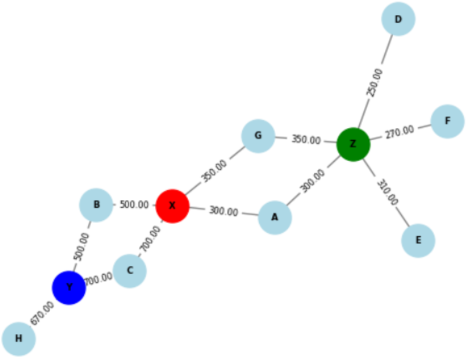


Figure 4: Figure

Teams	Eigenvector centrality score
X	0.50
Y	0.50
Z	0.40

Power Method

- To compute the eigenvector centrality vector x_i, we can utilize an iterative process known as the power method.

1. Choose an initial vector $\mathbf{x}^{(0)}$ of length n.
2. Perform the following steps for a specified number of iterations or until convergence is reached:

   2a. Update the vector $\mathbf{x}^{(t)}$ at iteration t using the equation $\mathbf{x}^{(t)} = A\mathbf{x}^{(t-1)}$, where A is the adjacency matrix.

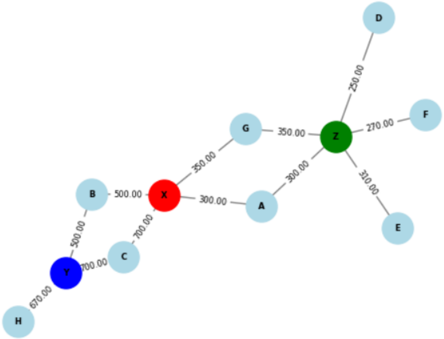
   2b. Normalize the vector $\mathbf{x}^{(t)}$ by dividing it by its largest element to prevent numerical instability and ensure that the magnitudes of the vector remain in a reasonable range.
3. Check for convergence by comparing the difference between $\mathbf{x}^{(t)}$ and $\mathbf{x}^{(t-1)}$.

Assessing the Probability of Match Outcomes

- A random walk on a graph is a process that begins at some vertex, and at each time step moves to another vertex.
- We can calculate the likelihood of different teams winning a match by using random walk approach.
- Random walks on graphs can be considered as specific instances of **Markov chains**.

Assessing the Probability of Match Outcomes

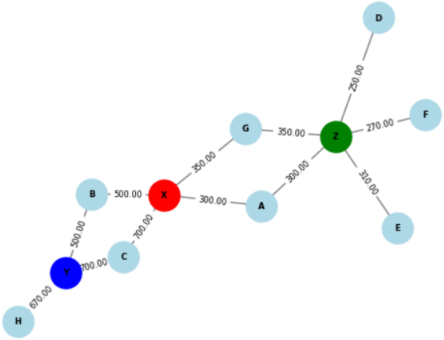
- Let's $w(e)$ be a weight function that assigns a weight to each edge $e \in E$,
- To calculate the likelihood of a match outcome Team X and any other national team ξ,
  1. We can perform a random walk starting from team X
  2. Compute the probability of reaching team ξ within a certain number of steps.



Assessing the Probability of Match Outcomes

- The initial distribution, $P_0(X) = 1$,
- $P_0(\xi) = 0$ for all other nodes ξ
- The transition probabilities from team X to its neighboring nodes v is given as

$$p(X, v) = \frac{w(e)}{\sum_{e \in V_X} w(e)} \tag{3}$$

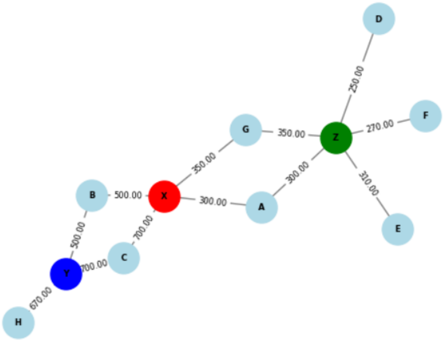


Assessing the Probability of Match Outcomes

- Let $P(X)$ represent the probability of team X winning a match
- The probability can be estimated by solving the following system of equation:

$$P(X) = C_X \cdot p(X, u_0) \cdot \prod_{k=1}^{n} p(u_{k-1}, u_k) \cdot p(u_n, \xi) \tag{4}$$

Here, C_X represents the centrality score of team X,

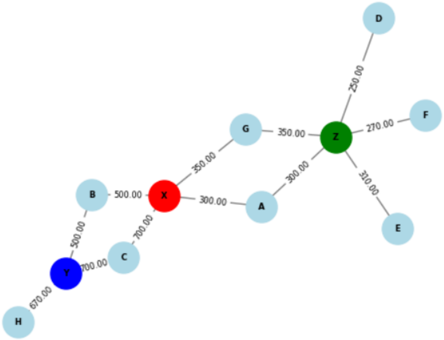


Assessing the Probability of Match Outcomes

Teams	Eigenvector centrality score
X	0.50
Y	0.50
Z	0.40

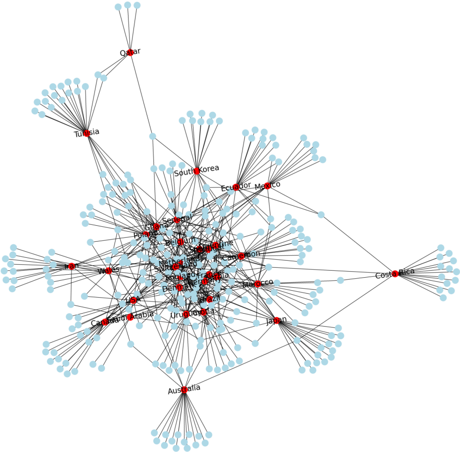
Teams	X	Y	Z
X	-	0.48	0.62
Y	0.52	-	0.64
Z	0.38	0.36	-

Table 3: Estimating the probabilities of each team's victory in matches against one another



Analysis of the Proposed Model against Other Models

- FIFA world cup 2022 network



A Comparative Analysis of the Proposed Model against Other Models

FIFA	Bookmakers	Opta	Hybrid-model	Graph-Based	Actual Ranking
BRA	BRA	BRA	BRA	FRA	ARG
BEL	FRA	ARG	ARG	BRA	FRA
ARG	ARG	FRA	GER	CRO	CRO
FRA	ENG	ESP	NED	ARG	MOR

Table 4: Top 4 Teams Ranking for FIFA World CUP 2022

A Comparative Analysis of the Proposed Model against Other Models

- Comparing the Network-Based Method to
  - Mathematical Modelling Approach by Oxford University's modeller Joshua Bull
  - Random forest

Stages	Joshua Bull	Graph-Based Method	Machine Learning
Round 16	0.56	0.75	0.75
Quater final	0.75	0.75	0.75
Semi final	0.50	0.75	0.50
Final	0.00	0.50	0.00

Table 5: Percentage of Correctly Predicted Teams for Each Stage of the Tournament

Conclusion

- We introduced a novel approach based on graph theory to rank and predict the outcomes of international football competitions.
- We applied the model to the FIFA World Cup 2022.
  1. A total of 27 out of the 48 group stage matches were accurately predicted.
  2. Out of the 16 matches in knockout stage, 11 were correctly predicted.
  3. The model achieved a probability of 63.54% for correctly predicting match outcomes.

Thank you

References I

1. Dixon, M. J., & Coles, S. G. (1997). *Modelling association football scores and inefficiencies in the football betting market*. Journal of the Royal Statistical Society

2. Borgatti, S. P. (2006). *Identifying sets of key players in a social network*. Computational & Mathematical Organization Theory, 12, 21-34.

3. Beal, R., Middleton, S. E., Norman, T. J., & Ramchurn, S. D. (2021). *Combining machine learning and human experts to predict match outcomes in football: A baseline model*. In Proceedings of the AAAI Conference on Artificial Intelligence.

4. Groll, A., Ley, C., Schauberger, G., & Van Eetvelde, H. (2019). *A hybrid random forest to predict soccer matches in international tournaments*. Journal of Quantitative Analysis in Sports, 15(4), 271-287.

References II

5. Constantinou, A.C., (2019) *Dolores: a model that predicts football match outcomes from all over the world.* Mach Learn 108, 49–75. https://doi.org/10.1007/s10994-018-5703-7
6. https://www.zeileis.org/news/fifa2022/, Access date (July 09, 2023).
7. https://www.linkedin.com/pulse/predicting-2022-fifa-world-cup-using-machine-learning-ameya-diwan/, Access date (July 09, 2023)
8. https://online-bookmakers.com/world-cup/, Access date (July 09, 2023)
9. https://www.maths.ox.ac.uk/node/61756, Access date (July 05, 2023)