Temporal Graph Machine Learning for Outcome Prediction in Football Matches Using Passing Networks

Ties Bloemen Marc Delgado Szymon Gołębiowski Olaf Heijl Oriol Monge

September 26, 2025

Introduction

Context

A lot of football prediction algorithms are used widely throughout the world, due to the massive popularity of the game. Whether it is for betting, or for predicting if your favorite team will win, there is a lot of interest in correctly predicting the outcome of a match, from football fans to management teams.
While traditional machine and deep learning methods can be effective for predicting a match, we want to exploit the graph-like nature of football during matches to predict the match winner of the game, given the statistics of the match that is happening live.

Graph ML formulation of an application

We take a lot of inspiration from the architecture decisions in the *"We Know Who Wins"* paper [1]. In a match, we can model a graph per team. In this graph, the nodes are players, and the directed edges are weighted by the number of passes among players. For the nodes, we will additionally embed some features of the players, subdivided into two categories: static features and dynamic features. For the static features, we will add the height and weight of a player. For the dynamic features, we will have the mean position of a player (as seen from above), the standard deviation for the position, and the passing accuracy per player. These will get updated as the game goes on. Additionally, team-specific global statistics will get aggregated, such as shots on goal, number of fouls, etc. These statistics will be combined with the graph information to create one embedded vector. Then the two embedded vectors (one per team) will be used to predict who will win.

Problem statement

In this project, our goal is to develop a graph-based model that predicts the outcome of a football match at fixed intervals throughout the game, by leveraging temporal information from passing networks. We aim to enhance an existing live match prediction system by integrating temporal graph learning techniques, allowing the model to capture the evolving dynamics of team interactions during the game.

Previous Work

Our project is primarily inspired by the paper *"We Know Who Wins"* [1], which employs passing networks in combination with Graph Attention Networks (GATs) to predict match outcomes. Similarly, [2] applies GATs and Graph Convolutional Networks (GCNs) to outcome prediction in the domains of e-sports and American football, where game states are represented as graphs. Both works emphasize a key limitation of current approaches: the lack of explicit modeling of temporal dependencies in passing networks or game states. Addressing this gap is the central motivation of our project, as we aim to incorporate time-varying characteristics directly into the graph-based predictive models.

Hypotheses

We believe that our model, which uses the temporal dependencies, will outperform the baseline of the previous work [1]. That is because we will add temporal data, and have more meaningful features per node. We also think that the Product-graph method will outperform the RNN model and the disjoint model, as there are more intricacies. It will depend however on the amount of data we train the model on.

Methodology

Our objective is to develop an outcome predictor for soccer matches using passing networks as the primary data representation. As a starting point, we will implement a Graph Attention Network (GAT) based on the approach proposed by [1], which leverages passing networks and in-game features accumulated over a predefined time window. This model will serve as our baseline.
Building on this, we plan to incorporate the temporal dynamics of passing networks and in-game features by exploring the following approaches:

- **Disjoint model**: process each snapshot of the passing network and associated features independently, and aggregate the resulting graph embeddings using a recurrent neural network (RNN).
- **Graph-RNN model**: directly extend recurrent architectures to operate over graph-structured inputs.
- **Product-graph approach**: construct a higher-order graph that encodes both temporal and structural information jointly.

We plan to collect the data independently using the `SoccerData` SDK [3], which extracts structured information from *WhoScored.com* [4] into `pandas` dataframes. The data for every match is event-based and provides detailed play-by-play information including passes, shots, and other in-game events. This will allow us to build the required passing networks and associated in-game features for our models.

Beyond model development, we aim to investigate the **scalability** and **generalization** of our approach. Specifically, we will train our models on data from a single league (e.g., the English Premier League) and evaluate their performance when applied to matches from other leagues. This will allow us to assess the extent to which the learned representations transfer across different competitive contexts.

Experiments

In general, a good experiment is to vary the window in which we aggregate: Our baseline will use windows of 5 minutes to aggregate the score, but perhaps this is too fine, or too coarse. We can run a histogram for each chosen time window. Additionally, we can measure the effect of also including the global game statistics as mentioned in the Graph-ML formulation. If we take these out, does the accuracy drop, or increase? Finally, we can evaluate which product-graph embedding is optimal for this problem (Cartesian, Kronecker, etc).

Potential Setbacks

A potential challenge lies in the construction of our dataset. We will need to collect the data through the `SoccerData` SDK, which extracts information from *WhoScored.com*. While this approach ensures flexibility and reproducibility, it may also be time-consuming, particularly in terms of handling data scraping limitations, cleaning raw event data, and transforming it into passing networks and feature representations suitable for our models. This additional preprocessing effort could reduce the time available for model development and experimentation.

References

[1] Jinmo Lee, Eunil Park, and Angel P del Pobil. "We know who wins: graph-oriented approaches of passing networks for predictive football match outcomes". In: *Journal of Big Data* 12.1 (2025), p. 147.

[2] Peter Xenopoulos and Claudio Silva. "Graph neural networks to predict sports outcomes". In: *2021 IEEE international conference on big data (big data)*. IEEE. 2021, pp. 1757–1763.

[3] *WhoScored - soccerdata 1.8.7 documentation — soccerdata.readthedocs.io*. `https://soccerdata.readthedocs.io/en/latest/datasources/WhoScored.html`. [Accessed 26-09-2025].

[4] *Football Statistics | Football Live Scores | WhoScored.com — whoscored.com*. `https://www.whoscored.com/`. [Accessed 26-09-2025].