Analysis of Team Passing Networks Considering Expected Threat

Sports Analytics: Final Project

Professor

Javier Martín Buldú

Team Members

Joe McMullen
joe.p.mcmullen@gmail.com
Lihong Ji
jilihong1998@gmail.com

Fútbol Club Barcelona, FCB
Universitat Politècnica de Catalunya, UPC
Postgraduate in Sports Analytics
Barcelona, 2023

Contents

Analysis of Team Passing Networks Considering Expected Threat

By Lihong Ji and Joe McMullen

Abstract

Network science is increasingly being applied to analyse the complex interactions which emerge in football matches, most often considering networks with the players of a team as nodes, and the passes between players as edges. These passing networks can be analysed to understand the structure and behaviour of teams when in possession of the ball. This project proposes the extension of passing networks in football considering an Expected Threat(xT) model which gives an expected value to possession based on location on the pitch. Directional passing networks are proposed considering accumulation of xT due to passes between players as edge weights in the network. The Eigenvector Centrality(EC) of players in the proposed networks is compared with previously established methods of constructing passing networks weighted by pass count. Pass count networks show strong correlation between EC of players as passers and receivers across all positions. xT weighted networks result in a sparser distribution between the EC of players as passers and receivers. Additionally, the comparison of player importance in networks which are weighted with pass count and those weighted with xT is analysed. The results indicate that xT weighted network analysis gives new insight into a player's role in a team's passing structure, more clearly distinguishing their importance to attacking play as passers and receivers. The metrics discussed may be applicable as key performance indicators in player profiling for recruitment and tactical deployment in matches.

Keywords: Passing Network, Expected Threat, Eigenvector Centrality, Key Performance Indicator.

1 Introduction

1.1 Background

Football is an invasion sport in which two teams of 11 players aim to score by moving the ball into the opposing team's goal. Given that each team consists of multiple agents working together to overcome the other, (in addition to input from coaches, referees, and supporters), the performance of each team that emerges is highly complex. Therefore, we can consider football in terms of hierarchies of complex systems. For example, a team is a complex system made up of players who are also complex systems themselves [9].

As a result, analysis and prediction of football performance can consider the approaches developed in the field of complex systems science[4]. These approaches were developed by considering the interactions, structures and behaviours of complex systems such as the internet, biological organisms, and social networks. Research in the field has been conducted to consider whether an approaches use to analyse these phenomena are applicable to sporting performance[11].

Network science is one branch of complex systems science which considers a system as nodes which represents agents and edges which represent an interaction between the nodes[6]. In the case of football, individual players can be considered as nodes and the edges between them can represent some actions among the players[2].

In football, passing between teammates has emerged as the most effective method to move the ball to a desired area whilst preventing the opposing team from claiming possession. As a result, analysis of a team's passing can be beneficial to gain insight into team performance. Directed networks in which the nodes represent the players of a team and the edges represent the volume of passes from one player to another have become popular as a method of performance analysis. Visualising this passing network of a team during a match with nodes plotted at average player location can give insight into team structure as well as passing patterns[3].

1.2 Motivation

One notable measure that arises in networks is the eigenvector centrality which is a measure of the transitive influence of nodes[7]. It suggests node connected to other well-connected nodes is important in the context of that network. In directed graphs, this is the importance of the node to the network as a receiver. One example is Google's page ranking system which uses the eigenvector centrality based on links to each webpage to determine the most important and provide this as the first search result[1].

Football analysis also represents a challenge as it is a game with sparse rewards; the average number of goals in each LaLiga game in the 2022/23 season was 2.51. With this complexity and goal sparsity, football results can be highly unpredictable and not always reflective of each team's performance. Attributing value to individual player performance is also desirable to determine areas for potential improvement through training or recruitment.

This has led to the development of additional metrics which give a more granular indication of team performance and player contribution. One such metric is expected goals (xG) which can be defined as the probability that a shot results in a goal and can be considered as a chance creation indicator[8].

However, xG can only assign value when a shot is taken. Shots occur more often than goals but are still sparse in the context of football. Additionally, xG gives more insight into the chance creation of the team as a whole and the performance of the shooter, but does not give context on how the other 10 players on the field contributed to the creation of that chance. Therefore, the field of sports analytics continues to work backwards from goals with the aim of quantifying how all actions affect the probability of each team scoring. Expected threat is one further development along this methodology.

Expected threat (xT) is a metric that provides the expectation of scoring within a possession, based on the location of the player in possession of the ball[10]. Areas of the pitch are considered as nodes in a Markov chain and theoretically action is the probability distribution in that area. Using xT, we can assign a value to each area of the pitch and determine how the likelihood of scoring changes as the ball moves between different areas. We can then calculate how the player in possession's actions increase or decrease the likelihood of their team scoring, and subsequently analyse the contribution of players to the threat of a team over a match or season.

1.3 Objectives

The aim of the project is to extend the analysis of passing networks in football to include consideration of the change in scoring expectation that passing combinations between players create.

By filtering the pass with different criterions, we can get distinct passing dynamics. In this project, we will leverage xT as a filtering criterion to refine the passing network, producing a more significant attacking-weighted network. This refined network will consist of only those passing connections that contribute positively to the accumulation of xT, representing the combinations of players that actively increase the team's scoring expectation through their collective passing sequences.

Then the eigenvector centralities of pass count matrix and pass xT matrix will be studied on different positions. Through the comparison we can get some insight of intricate passing patterns based on the eigenvector centrality difference of players.

2 Theoretical Methodology

2.1 Expected Threat Model

During the attack phase high synergy is performed by professional football players and each player plays an implacable role in the team. Even though there are already lots of models quantifying player's performance, high bias to the player who scores still exists in output results. Therefore, giving the proper credits to those linker players before the last pass is fairly necessary. In order to solve this, Expected Threat(xT) model is proposed based on the passing tiles[10]. Generally, Expected Threat is the goal expectation of a predefined tile given you have possession of the ball in that area. Concatenating all the tiles and its corresponding xT values, the xT surface is generated. Any related actions, such as pass, shoot and so on, can be evaluated by computing the difference of starting tile xT and ending tile xT.

Here, the algorithm of Expected Threat will be explained. As mentioned above that xT is the expectation of goal, the connection between goal and xT needs to be specified firstly. When a player has the ball at arbitrary tile, there is a probability distribution of possible actions for this player, in xT algorithm the actions are concluded as: pass, dribble and shoot. Each possible action deserves a reward, and obviously the reward of shot action is the goal. Subsequently, the reward of dribble and pass is supposed to be homogeneous to the goal, and in the algorithm the reward is set up as xT of last time point. Consequently the analytic expression of xT becomes a recursive function, which means the reasonable xT surface can be achieved after several iterations.

The complete analytic expression of Expected Threat is shown below. Some configurations need to be declared: the pitch is blocked down to 16*12 scale in this project. Alternative pitch partitions can also be considered, this particular configuration was selected following the original proposal of xT[10]. Pass and dribble are categorized together as the "Move" action.

$$\begin{aligned} xT_{x,y} &= P_{shot_{x,y}} * R_{goal} + P_{move_{x,y}} * R_{move} \\ R_{goal} &= P_{goal_{x,y}} * 1 + P_{nogoal_{x,y}} * 0 \\ R_{move} &= \sum_{z=1}^{16} \sum_{w=1}^{12} T_{(x,y)\to(z,w)} xT_{z,w} \end{aligned} \tag{1}$$

where (x, y) and (z, w) denote the tile coordinate, $T_{(x,y)\to(z,w)}$ denotes the transition probability of the ball from (x, y) to (z, w).

The reward of a goal is the goal expectation. Since xT is derived from event data, the number of goals arising from shot events provides the goal expectation. I a given tile, the ratio of goals to shots is defined as P_{goal}, then $P_{nogoal} = 1 - P_{goal}$.

The reward of a move action is the expectation of scoring later in the possession as a result of moving possession to a different tile. With the assumption that a player can pass or dribble the ball to any other tile on the pitch, the reward of move is the point-wise product of the current tile transition matrix and destination tile xT at the current iteration.

2.2 Passing Network

Passing is the fundamental action that underpins the dynamics of professional football. It serves as the crucial link that connects all the players on a team, enabling the emergence of various synergistic patterns and collective movements. Each pass involves two players – one who initiates the pass (passer) and another who receives it (receiver). This directional relationship between the passer and receiver is analogous to the representation of nodes and edges in graph theory, therefore, the concept of a passing network in football is a natural development.

Consider a team comprising 11 players, each represented as a node in a network. Between these nodes, there exists 121 potential directed edges, reflecting the possible passing connections between every pair of players. This network structure forms a directed graph, which serves as the framework for modeling and analyzing the team's passing dynamics.

After each match, the passing network graph can be computed with the data of pass events that occurred during the game. By taking the sum of passes between players as the edge weight, the passing network for a team is produced. This can be subjected to various analytical techniques to gain insights into the team's tactics, the specific roles and contributions of individual players, and the overarching strategies governing the movement and distribution of the ball across the team.

2.3 Eigenvector Centrality

Considering the passes between players as directional information flow, the passing network can be considered to be a Markov Random Field(MRF). This has a stationary distribution and satisfies the conditions of aperiodicity and irreducibility as the ball can be passed to any teammate in any location and travel via any path. Therefore, is it possible to analyse the stationary distribution of the passing network.

The adjacency matrix of a passing network is typically referred to as the "pass matrix," with columns representing the passers and rows representing the receivers. An 11-dimensional vector is initialized to represent the starting probabilities of each of the 11 players receiving possession. This player probability vector is then multiplied iteratively by the pass matrix until the probability distribution converges to a steady state

The mathematical expression of the stationary distribution is demonstrated below,

$$\lambda_{stationary} = \frac{A^C \lambda_{initial}}{Z} \tag{2}$$

where $\lambda_{statrionary}$ and $\lambda_{initial}$ denote the stationary probability distribution and the random initial probability distribution. A denotes the pass matrix. C denotes the iterations until convergence. Z denotes the normalizing constant.

When analyzing team passing networks, eigenvector centrality can be calculated using the same approach outlined in Equation (2). The resulting eigenvector centrality quantifies the relative transitive importance of each player in the teams passing structure. The classical eigenvector centrality computation utilizes the receiving pass matrix, where columns represent passers and rows represent receivers. This yields an eigenvector centrality metric reflecting players' ball-receiving contributions or their capacity to receive information from neighboring nodes. Alternatively, by transposing the pass matrix and computing the corresponding eigenvector centrality, a measure of the transitive importance of each players' passing contribution is obtained. This can be interpreted as quantifying how much information or influence the current node exerts on its neighboring nodes through outgoing passes.

3 Experimental Analysis

3.1 Data Management

3.1.1 Data Description

The data used in this study consists of match information and event data for 192 games in the first half of the 2023/24 LaLiga season provided by Opta(now StatsPerform). For each match, we have 4 files:

- Team

The "Team" collection contains team id and name for both teams in each match, the teams which featured can be seen below as Figure 1:

Figure 1 – 20 Teams in 2023/24 LaLiga season.

- All Events

The "All Events" collection contains data for all events that occurred during a match, and includes the match code in which the current event occurred, the event time information including the code of the half of the game, minute and second, the outcome of the event. It also includes the information of the event initiator: team, name. Also the x, y coordinates of event initiating position and ending position.

- Pass

The "Pass" collection contains additional detail on pass events not which is not in "All Events" including receiving player information and location, the sequence and possession number that each pass was part of and whether the pass was offside.

- Player

The "Player" collection is the summary of player information for each game, including identifier, name and shirt number, what position they started the game in and how many minutes they played.

The 4 files were obtained from the original f24.xml file of Opta, which contains all events of a match.

3.1.2 Database

In order to manage the data and facilitate remote data manipulation, a MongoDB Atlas cloud server was used for the database. 5 collections were built up: event, match, pass, player, team extracting the necessary data from the original dataset for network analysis. The structure of the database generated is detailed below.

- Event

In this collection, documents align with the structure of the all-event files provided by Opta. The specific data structure can be referred to Table 4 in Appendix. There are 317,575 event documents in total in this collection.

Additionally, a field for xT was created to store the xT value created by each action and a cummulative "time" field was created integrating"period","minute" and "second" fields.

- Match

This collection consists of documents for each match, including home and away team information and player information. For player information, a nested dictionary contains player name, shirt number, position, start information and minutes played. The data structure is shown below in Table 1. There are 192 match documents in the collection.

Table 1 – Match Collection Data Structure.

Attribute	**Data Structure**
home_team_id	String
away_team_id	String
home_team	String
away_team	String
match_id	String
home_players	Dict[String, Dict[String | Integer]]
away_players	Dict[String, Dict[String | Integer]]

- Pass

The collection contains the key information of pass events. This matches the structure of the provided pass files with the inclusion of time and xT fields as with the event collection. The specific data structure can be referred to Table 4 in Appendix. There are 183,899 pass event documents in total in the collection.

- Player

The Player collection contains information of the players who featured in any squad in the games featured in the dataset. Considering that one player may play different positions in different matches following tactical instructions, the attribute position is defined as a dictionary where the key is the position that the player has played and the value is the counter of that position. Given a player may transfer team, the team_id contains a list where the identifier of all teams the player features for is stored. The data structure in this collection is shown in Table 2. There are 646 player documents in the collection.

Table 2 – Player Data Types.

Attributes	Data Structure
player_id	String
first_name	String
last_name	String
known_name	String
team_id	List[int]
position	Dict[String, Integer]
starts	Integer
apps	Integer
minutes	Integer

- Team

This collection is derived from the other collections by establishing a dictionary to record team ids and the corresponding team name. The data structure of this collection is shown in Table 3. There are 20 team documents in this collection.

Table 3 – Team Data.

Attribute	Data Structure
team_id	String
team_name	String

3.2 Algorithm Implementation

Python is the programming language used to implement the algorithms and search methods in this project

There are 2 main classes in the project programming: xTheat and MatchAdvancedPassingStats.

The first class is used to collect information from seasonal event data and construct the seasonal xThreat surface. We also provide an API to users that it is available to create arbitrary team's xThreat surface by querying with team name. The idea of the algorithm is to block down the pitch into tiles first and then create a pitch graph using the tiles as the nodes. All the tile statistical data can be stored as the node features, which makes it easy to implement Expected Threat algorithm. In practice, we take advantage of NetworkX to create the pitch graph.

The second class "MatchAdvancedPassingStats" is an extending class of class "MatchInfoRetriever". The base class only extracts basic information of a match. Home and away team information are collected up and two graphs of both teams are created in the base class constructor. The maximum node number is set up as 11 in convenience for establishing passing network, and the 11 players of each team are the ones who play together for the longest time. Subsequently, in "MatchAdvancedPassingStats" the pass count and pass xThreat are then retrieved from the database. There are four member functions in this advanced class, which provide ways of getting the pass count matrix, pass xThreat matrix, and in&out eigenvector centrality. It is worth noting that the passing matrix is in(receiving) matrix, if out matrix needed the passing matrix is supposed to be transposed.

More details can be checked in the project repository.

3.3 Analysis of Results

3.3.1 Expected Threat

Using the method outlined above, an expected threat grid was produced using all events in the database as shown in Fig 2.

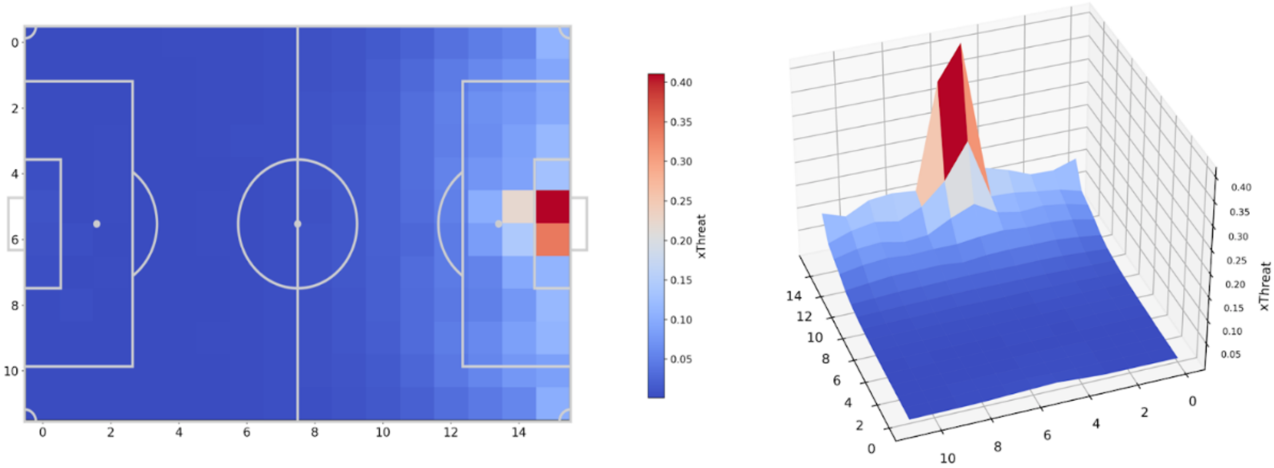

Figure 2 – The xT surface of all the teams.

Given our methodology of iteratively running the algorithm detailed in section 3.1, convergence is not achieved and the result is an approximation of the convergent state that might be expected from a markov chain. Increased iterations are equivalent to increasing the number of actions in a possession chain. As the discussed method does not include possession loss as a possible action, given enough iterations, all possession chains will result in shot actions and so the xT of the areas further from the goal will increase with further iterations.

In general, areas close to opposing team goals have high xT values. This is expected and would be similar to an xG model derived from the same data given that the expectation of a goal in this area is largely due to shooting and this is the most probable action. However, it can be observed in the corners that the xThreat is around 0.2, this is higher than would be seen in an equivalent xG model. From here, passing to the box gives the area its value and this highlights the deeper context regarding possession chains that the xT model provides.

This grid in Figure 2 is derived from data of the first 192 matches of the 2023/24 LaLiga season and provides a representation of the areas that were most dangerous in these matches. In this way, the xThreat grid represents a merging of attacking methodologies of the teams over this period. The grid indicates a slight bias in goals scored from the left side of the box indicating that teams scored more often from these zones. From this, the potential for an xT model to be used as a representation of team attacking tendencies can be concluded. The xT surfaces of top 4 teams in LaLiga when this data was collected are displayed in Figure 3. It can be observed that each team has a distinct xT surface suggesting different attacking tactics are implemented different teams.

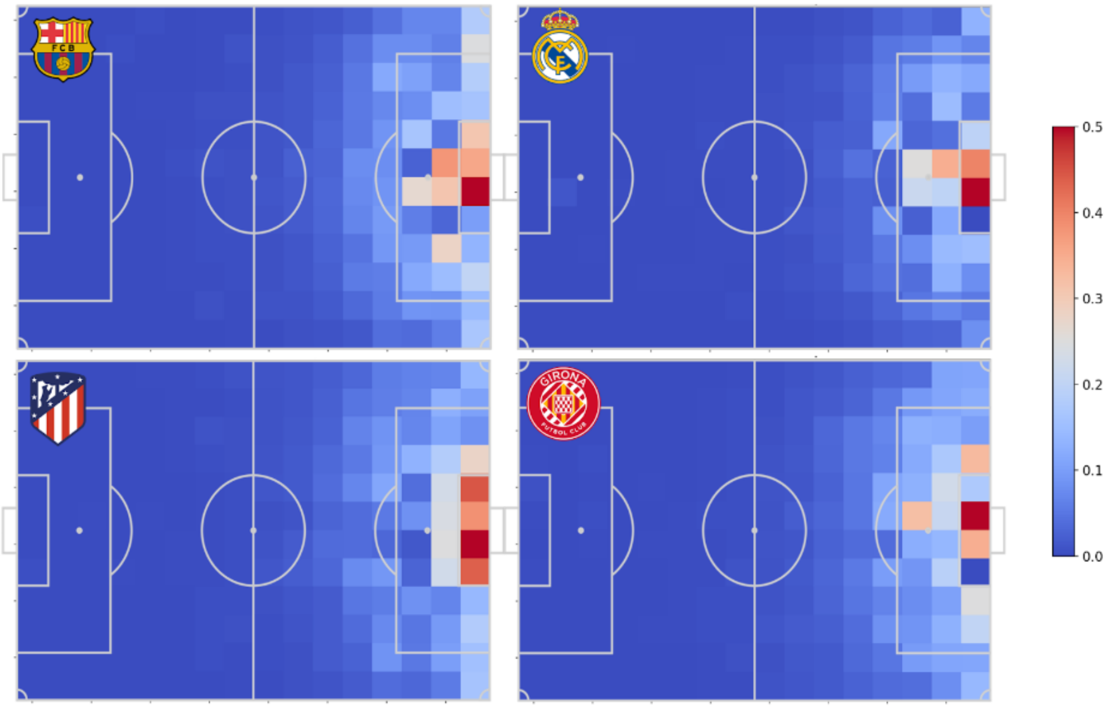

Figure 3 – the threat profiles for each team differs, Real Madrid (top right) have a high expected threat focused centrally in the penalty area whilst Atletico Madrid have a wider and more shallow profile of high xT zones indicating chances were created from closer to the opposing goal line. Barcelona (top left) and Girona (bottom right) have more varied profiles indicating chances were created in more varied patterns.

3.3.2 Passing Network

Difference between pass count network and xT network can be seen in Figure 4 and Figure 5 below showing the networks for the LaLiga match between Barcelona (red) and Sevilla (blue) on September 23r̂d 2023, the final result was Barcelona 1-0 Sevilla.

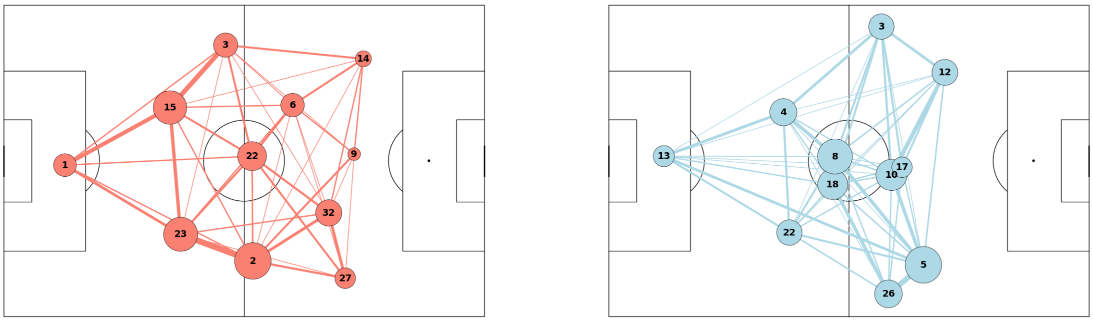

Figure 4 – Passing Networks of Barcelona(Left) and Sevilla(Right) with node size and edge width weighted by pass count. Node labels represent player shirt numbers.

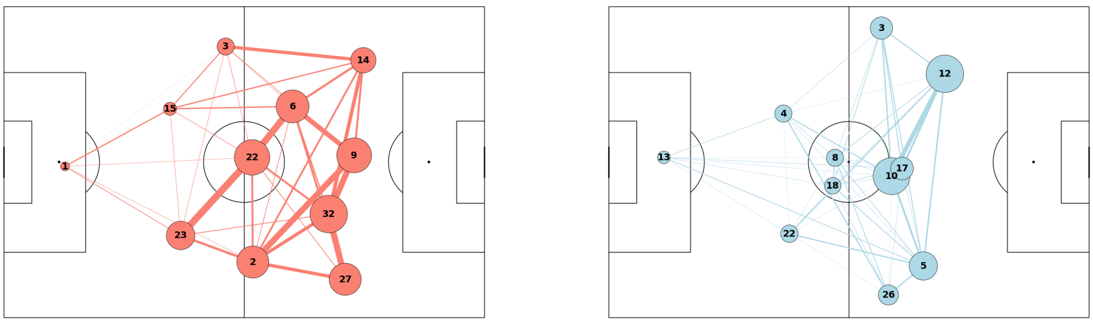

Figure 5 – xT Passing Networks of Barcelona(Left) and Sevilla(Right) with node size and edge width weighted by positive xT accumulation. Node labels represent player shirt numbers.

The node size and edge width can be observed to vary between the same teams when considering pass count and xT. The xT network distills the significance of players in the attacking phase. Barcelona player's attacking contribution is more homogeneous in contrast to the Sevilla xT network in which fewer players generated the majority of threat.

3.3.3 Networks and Eigenvector Centrality

Directed networks for each match were produced according to the methodology described in section 3.3 and the resulting adjacency matrix considering the edges of the network as the total count of passes between players was used to calculate the eigenvector centrality for each player. For all players who played over 200 minutes, the mean of their eigenvector centralities was taken across the networks they featured in. As previously discussed, eigenvector centrality gives the importance of the player as a receiver and is referred to as the "In Pass" eigenvector centrality. The transposed adjacency matrix is used to determine the importance of a player as a passer and is referred to as the "Out Pass" eigenvector centrality. The mean was again taken over all games.

Figure 6 shows the correlation between this mean in and out eigenvector centrality for all players of all the positions with over 200 minutes in the dataset. Goalkeepers(Green) had lower relative importance to the network and were more important to passing than receiving, indicating they are often originators of possession. Strikers(Red) show the opposite trend with greater importance as receivers suggesting they act as sinks for passing moves. Defenders(Blue) tended to be of high importance to both networks which suggests they are heavily involved in passing moves in intermediary stages. Midfielders(Orange) can be seen to vary in profile with some showing a correlation similar to strikers whilst others are within the defenders cluster. This indicates more complexity within the midfield class, with some midfielders performing a different role in a team's passing than others. However, the average in and out pass eigenvector centrality are highly correlated, the intuition for this is that it is highly likely when a player receives a pass, they continue the possession and give another pass regardless of position.

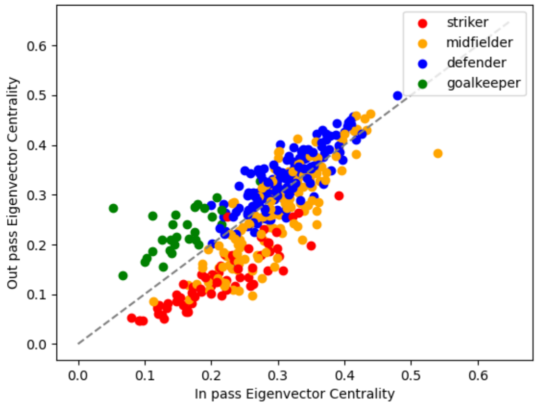

Figure 6 – In&out pass eigenvector centrality comparison.

Subsequently, networks were produced for each match considering positive expected threat contributions as the edge weights. The eigenvector centrality of the networks was computed on the adjacency matrix and transposed adjacency matrix and averaged for each player. This provides the mean "In xT" eigenvector centrality and the mean "Out xT" eigenvector centrality for each player which can be considered as each player's importance to the attacking play of a team as a receiver and passer respectively. Figure 7 shows weaker correlation between players contribution when considering positive xT contribution than when solely considering pass count. In addition the grouping of positions among the outfield roles is much sparser. Considering expected threat, greater distinction can be made between players than considering pass count alone.

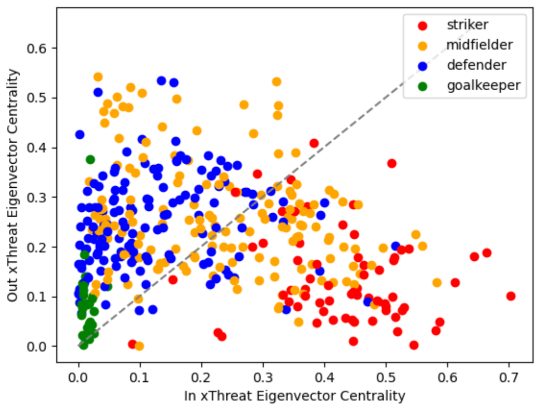

Figure 7 – In&out xT eigenvector centrality comparison.

The difference in mean “In xT” eigenvector centrality and “In Pass” eigenvector centrality was calculated which will be referred to as the “In eigenvector Centrality Difference”. A more positive value in this metric indicates a large attacking importance with a lower pass count importance as a receiver, and can be understood as a measure of attacking efficiency for a player as more threat is created with less passes received. The same process was followed to create the “Out eigenvector Centrality Difference” measuring attacking efficiency as a passer.

As shown in Figure 8 and Figure 9, most defenders have a negative value for both in and out eigenvector centrality difference. It can be concluded that teams mainly use defenders as maintainers of possession who play large numbers of safe passes and rarely play or receive aggressive penetrating passes. Some players are exceptions and it can be observed that they perform a more attacking role, with some examples of this for both passers and receivers.

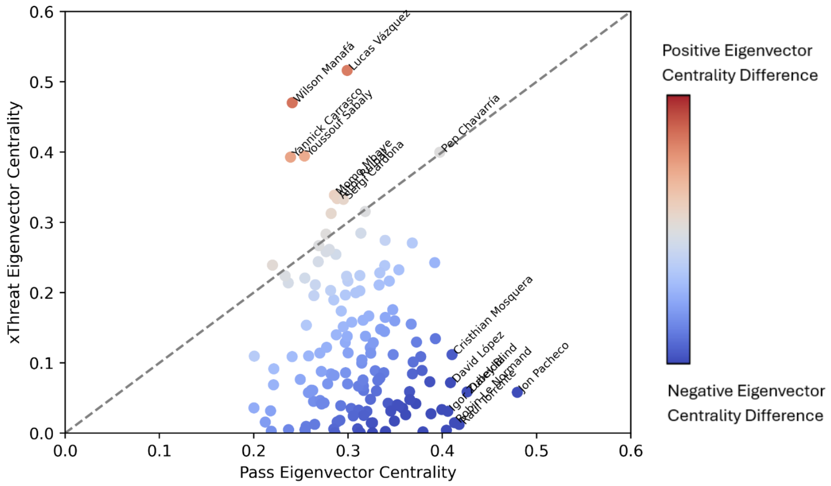

Figure 8 – Defender - In(ball receiving) pass and xT eigenvector centrality.

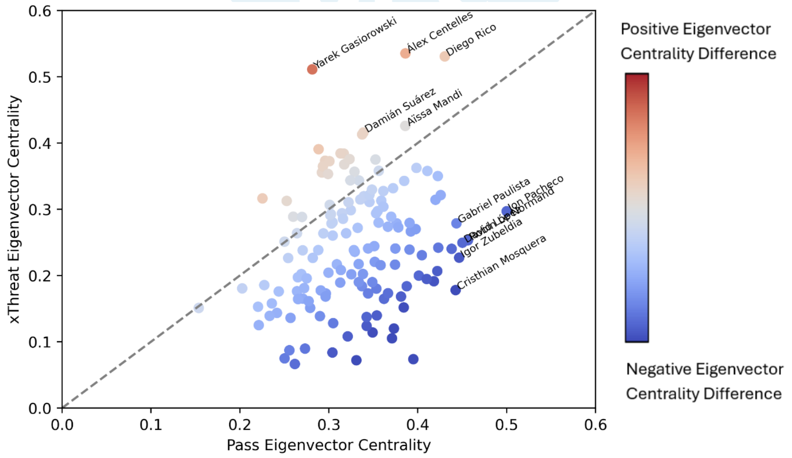

Figure 9 – Defender - Out(ball passing) pass and xT eigenvector centrality comparison.

As seen in Figure 10 and Figure 11, midfielder is a classification which describes a broader range of players in the context of their role in the passing network of the team. Examples at both extremes of eigenvector centrality difference can be seen particularly for receiving. Some midfielders tend to be important for receiving large numbers of non-threatening passes whilst others are important for receiving aggressive attacking passes. The results indicate it is difficult to be important for receiving many passes and receiving dangerous passes. As passers, the spread of eigenvector centrality difference is smaller, with those who are important for making a large number of passes also being important for contributing threat.

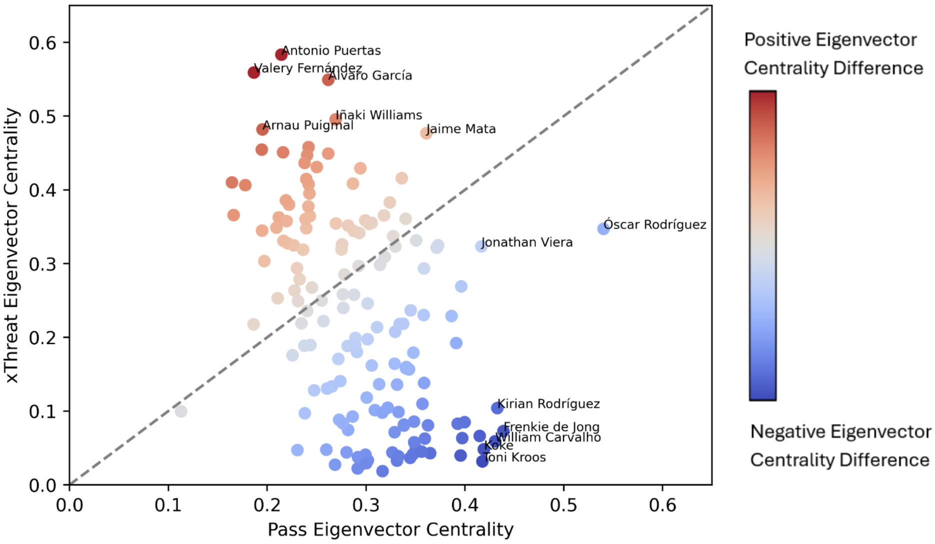

Figure 10 – Midfielder - In(ball receiving) pass and xT eigenvector centrality.

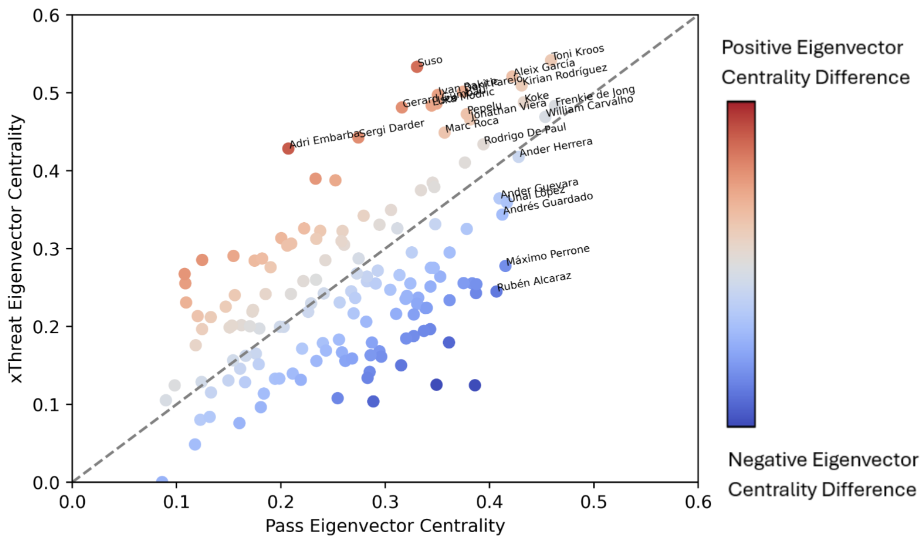

Figure 11 – Midfielder - Out(ball passing) pass and xT eigenvector centrality.

In Figure 12 and Figure 13, we can see most strikers have a strongly positive eigenvector centrality difference. This aligns with the concept that strikers aim to receive the ball in dangerous areas and attempt to score, acting as sinks for passes particularly into high threat areas. However, as passers, there is a large variation in striker importance to both passing and xThreat contribution. It can be concluded that some strikers are more involved in earlier phases of play and some are more focused only in the final attacking actions.

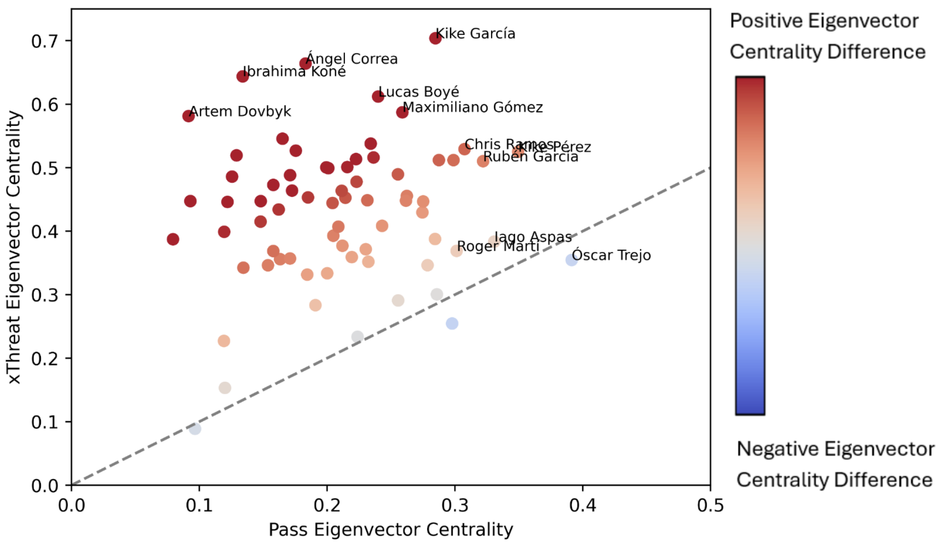

Figure 12 – Striker - In(ball receiving) pass and xT eigenvector centrality.

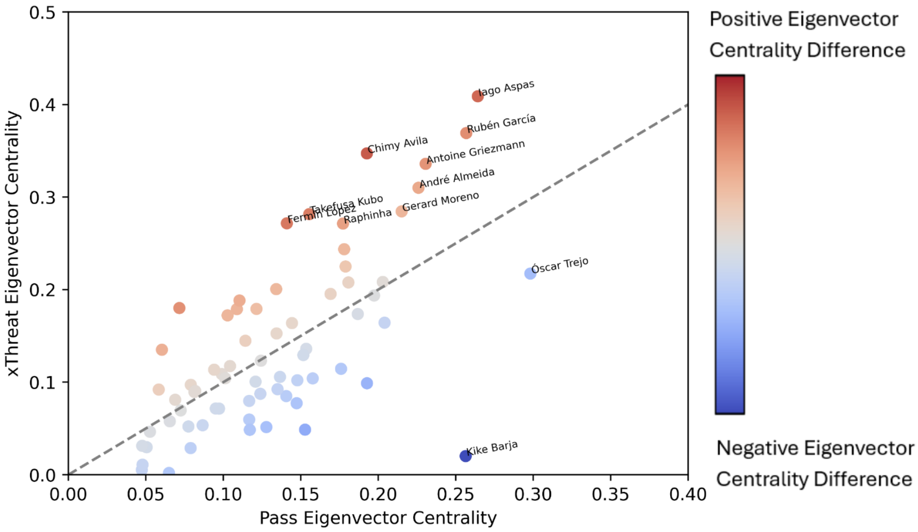

Figure 13 – Striker - Out(ball passing) pass and xT eigenvector centrality.

4 Conclusion

Training an expected threat model involves working back from goals to assign expected value to areas of the pitch. As goals are sparse, we require a large sample to achieve a reasonable model. The data from the 192 games analysed contained 510 goals in total. As a result the patterns in the xT grid for all the recorded matches and those of the individual teams discussed previously may arise from the inherent sparsity of goals, which can be considered as the noise in the goal, pass and dribble distributions, rather than indicative of the teams style of play. Considering a team's xT grid over a certain time period as a representation of their attacking tendencies may not be suitable due to the inherent dynamism of football. Teams change tactics and lineups over time, and in response to opponents and game situations. This intricacy cannot be captured by xT grids, and referring to passing networks themselves could provide a better insight into team attacking patterns.

The expected threat grid generated allows quantification of on-ball event value and provides additional context to a scoreline in a game which is unavailable through traditional metrics. However, using event data and considering only on-ball actions results in the loss of a large amount of context such as the location of other players on the pitch and game state. Improvements to expected threat which include this missing context have been developed such as Expected Possession Value[5] which uses tracking data to capture information on all player locations. This is more computationally expensive compared to xT and event data is cheaper and available for a wider pool of competitions. For applications such as scouting, particularly in less prominent competitions, xThreat is an important metric for quantifying individual contributions during the progress of the ball.

An xT weighted passing network provides more context as an attack-weighted network compared to the those which only consider pass angle to determine attacking intent. Passes with low or negative xT contribution can be attributed to retaining possession or fulfilling a tactical objective on a higher level not capture by xT. Therefore, it may be suggested that players with a high pass count EC and a low xT EC take advantage of pitch space and regulate the team formation in order to overcome the chaos that arises during attacking phases. This can be observed in Figure 10 and 11. It is worth noting that Toni Kroos, Frenkie de Jong, William Carvalho are together in the same cluster.

As an extension to the application of network science in football, it can be seen that considering the positive xT contributions as edge weights gives insight beyond the pass count and allows for a more nuanced understanding of the importance of a player within their team and their role within the network particularly in attacking phases, and expands upon the discrete classification that position labels provide. This can be seen in the expansion of midfielders into those who are important to receiving large volumes of safe passes and those who are important in receiving small volumes of dangerous passes. This presents uses in profiling players for recruitment and to understand how they may be deployed in games.

This project focused on the use of network measures for analysis of individual players, however, the nature of network science means that behaviour is captured within a complex system and the results can not be considered to be a global measure of an individual's performance or ability. To what extent, the findings are due to the team around a player, or the task assigned to them by a coach, is unknown. They could give an indication a player is able to perform a similar role in a similar setup or an indication a player was unable to fulfil the role asked of them. In this way, network measures may be applied as tools for player profiling and performance analysis.

5 Future Work

As mentioned above, the outputs from graph networks provide insights into a player the context of a team and match. In order to extract information on players outside of that context, we can consider the tracking data. Because tracking data permits to reconstruct the occasion where the player was in, then the context is extracted and the most important point is that the information is player-centered. Through this method, we could look to predict the performance of players in contexts beyond their current team. Based on this, it is possible for us to construct player embedding, which is considered as the interesting future work. We get the insight of eigenvector centrality that it is the detailed balanced state of a team for passing, as known as the team passing pattern. This is the strong team-level evidence to construct the cost function during the player node embedding.

In this work, the mean values of network measures for players over the games available was analysed. Future work could consider the variability in these measures which may be beneficial to understand consistency of teams or variability of player role across matches. The analysis of additional network measures could also provide further clarity into the profile of a player and be beneficial for recruitment and performance analysis.

References

[1] Sergey Brin and Lawrence Page. "The anatomy of a large-scale hypertextual Web search engine". In: *Computer Networks and ISDN Systems* 30.1 (1998). Proceedings of the Seventh International World Wide Web Conference, pp. 107–117. ISSN: 0169-7552. DOI: https://doi.org/10.1016/S0169-7552(98)00110-X. URL: https://www.sciencedirect.com/science/article/pii/S016975529800110X.

[2] Javier M Buldú et al. "Using network science to analyse football passing networks: Dynamics, space, time, and the multilayer nature of the game". In: *Frontiers in psychology* 9 (2018), p. 1900.

[3] JM Buldu et al. "Defining a historic football team: Using Network Science to analyze Guardiola's FC Barcelona". In: *Scientific reports* 9.1 (2019), p. 13602.

[4] Keith Davids et al. "How small-sided and conditioned games enhance acquisition of movement and decision-making skills". In: *Exercise and sport sciences reviews* 41.3 (2013), pp. 154–161.

[5] Javier Fernández, Luke Bornn, and Dan Cervone. "Decomposing the immeasurable sport: A deep learning expected possession value framework for soccer". In: *13th MIT Sloan Sports Analytics Conference*. 2019.

[6] Mark Newman. *Networks*. Oxford university press, 2018.

[7] Mark EJ Newman. "Analysis of weighted networks". In: *Physical review E* 70.5 (2004), p. 056131.

[8] Alex Rathke. "An examination of expected goals and shot efficiency in soccer". In: *Journal of Human Sport and Exercise* 12.2 (2017), pp. 514–529.

[9] Hiroki Sayama. *Introduction to the modeling and analysis of complex systems*. Open SUNY Textbooks, 2015.

[10] Karun Singh. *Introducing Expected Threat(xT)*. 2023. URL: https://karun.in/blog/expected-threat.html.

[11] Michael Stöckl, Denise Plück, and Martin Lames. "Modelling game sports as complex systems–application of recurrence analysis to golf and soccer". In: *Mathematical and Computer Modelling of Dynamical Systems* 23.4 (2017), pp. 399–415.

Appendix.

All of the source code is in the GitHub repository link.

Table 4 – Event Data(Left) and Pass(Right) Collection Data Structure

Attribute	**Data Structure**
event_code	Integer
event_type	String
team_id	String
origin_player	String
origin_pos_x	Float
origin_pos_y	Float
minute	Integer
second	Integer
period	Integer
time	Integer
match_id	String
team_name	String
player_name	String
outcome	Integer
pattern_of_play	Integer
destination_pos_x	Float
destination_pos_y	Float
extra_detail	Integer
xthreat	Float

Attribute	**Data Structure**
pass_id	String
match_id	String
team_id	String
minute	Integer
second	Integer
period	Integer
time	Integer
origin_player	String
destination_player	String
outcome	Integer
origin_pos_x	Float
origin_pos_y	Float
destination_pos_x	Float
destination_pos_y	Float
offside	Integer
possession	Integer
sequence	Integer
xthreat	Float