XLVI CILAMCE
VITÓRIA - ES - BRAZIL - 2025
XLVI IBERO-LATIN AMERICAN CONGRESS ON COMPUTATIONAL METHODS IN ENGENEERING

Analyzing Team Performance Using Graph Theory Metrics

Lucas C. Portela[1], Felipe W. O. Mendes[2], Michel P. Silva[3]

Dept. of Computer Engineering, Federal Center of Technological Education of Minas Gerais
Divinopolis, Brazil
lucas.portela@aluno.cefetmg.com[1], felipewerneck@aluno.cefetmg.com[2], michel@cefetmg.com[3]

Abstract. In the realm of sports analytics, identifying the critical factors that distinguish successful football teams from those that underperform is of paramount importance. This paper conducts a comprehensive investigation into the similarities and differences between championship-winning teams and relegated teams in La Liga by harnessing the power of graph theory. We develop detailed player similarity networks, where individual nodes represent players and weighted edges denote the degree of performance similarity derived from an extensive set of statistical data. By employing a range of centrality and connectivity metrics, we uncover structural patterns that are intrinsically linked to team success. Our findings reveal that championship teams typically exhibit higher values in metrics such as degree, closeness centrality, clustering coefficient, and PageRank—indicators of robust team cohesion and distributed influence—while relegated teams often rely on a few key players with high betweenness centrality, a pattern that exposes vulnerabilities in their network structure. These insights contribute valuable perspectives to team performance analysis, recruitment strategies, and tactical planning in modern football.

Index Terms—Graph Theory, Football Analytics, Network Metrics, Team Performance

1 Introduction

Football, renowned for its dynamic and multifaceted na ture, is a sport where success depends not only on individual talent but also on the intricate interplay among players. Traditional performance evaluation methods have primarily focused on individual statistics, such as goals scored, assists, and defensive interventions. However, these conventional approaches often fail to capture the underlying team structures and collective dynamics that ultimately drive a team's success or failure. In recent years, network science has emerged as a pow erful analytical framework to examine these complex rela tionships. By modeling football teams as player similarity graphs—where nodes represent individual players and edges quantify performance similarity—we can uncover structural properties that distinguish successful teams from struggling ones. Graph theory, therefore, provides a novel perspective on team performance evaluation by enabling the identification of critical attributes such as connectivity, influence distribu tion, and tactical flexibility.

This study aims to explore the network characteristics that differentiate successful football teams from underperform ing ones. Utilizing comprehensive player performance data from FBREF [FBref], we construct similarity-based graphs and apply various centrality measures to identify key structural differences. Specifically, we evaluate metrics such as degree, closeness centrality, betweenness centrality, PageRank, clustering coefficient, and eccentricity to understand how these properties influence overall team performance.

The results of this study contribute to the field of football analytics by offering a data-driven approach to team evaluation. The insights derived from this analysis have practical applications in tactical planning, player recruitment, and strategic decision making. Our findings indicate that championship-winning teams tend to exhibit a highly connected structure with a balanced distribution of influence, whereas relegated teams often rely on a limited number of pivotal players, making them more vulnerable to strategic disruptions.

2 Related Works

Numerous studies have explored the application of network theory in football strategy analysis. For instance, López Peña and Touchette [2] demonstrated that network theory can effectively model football strategies by representing teams as weighted networks. Their research, which utilized passing data from the 2010 FIFA World Cup,

quantified player influence through centrality measures such as closeness, betweenness, and PageRank. The study highlighted how different playing styles—exemplified by Spain's highly interconnected tiki-taka approach—can be characterized by network properties such as clustering and edge connectivity.

Another recent study by Mendes et al.[3] applied graph theory to examine how coaching decisions impact team dynamics and player interactions. Their work emphasizes the influence of tactical strategies on passing structures and overall team cohesion, revealing significant correlations between metrics such as clustering coefficient and average path length with team performance. While their research focuses on coaching influences, our study concentrates on player similarity networks to identify structural patterns that differentiate successful teams from underperforming ones. Both approaches contribute significantly to the understanding of football performance through the lens of network analysis.

Additionally, Korte et al.[4] introduced an innovative methodology for football network analysis by employing play-by-play data to quantify player participation and influence throughout matches. Using data from the Bundesliga, they developed new metrics—such as flow betweenness—to identify intermediary players during different phases of the game. Their research highlights the importance of temporal network analysis, moving beyond aggregated match-level data to capture real-time interactions. Although their study primarily focuses on pinpointing key players during possessions, our analysis extends this perspective by considering team structures over an entire season. This broader scope enables us to uncover patterns that distinguish championship winning teams from relegated ones.

In summary, while previous research has leveraged network science to understand team performance, our approach offers a distinct methodological and analytical perspective, emphasizing season-long player similarity networks rather than isolated match-specific interactions.

3 Methodology

3.1 Data Collection

Player performance data was extracted from [FBref], encompassing multiple seasons in La Liga from 2018/2019 to 2023/2024. The choice of these seasons was driven by the availability of detailed performance statistics, as earlier seasons featured simpler metrics that could potentially limit the depth of the analysis. The dataset comprises both championship-winning teams and last-placed teams during the selected period. La Liga was chosen as the focus due to its diverse range of championship winners and relegated teams, in contrast to leagues such as the Premier League or Bundesliga, where a few teams often dominate. The collected statistics are organized into the following categories:

3.2 Categories of Player Performance Metrics

To analyze team structures effectively, we classify player performance data into four key categories: Player Information, Offensive Metrics, Defensive Metrics and Possession Metrics. Each category represents specific aspects of a player's role and contribution to team performance. Below, we provide a detailed explanation of these categories.

Player Information:

- Name: The player's full name.
- Position: The role they play on the field (e.g., defender, midfielder, forward).
- Age: The player's age, which can be relevant when assessing physical capabilities and career trajectory.
- Matches Played: The number of matches the player has participated in during a given season.

Offensive Metrics:

- Goals: The number of times the player has successfully scored.
- Assists: The number of passes that directly resulted in a goal.
- Completed Passes: The total number of successful passes a player has made.
- Attempted Passes: The number of total passes attempted, regardless of success.
- Dangerous Passes: Passes that lead to potential goalscoring opportunities, such as key passes into the final third or through balls breaking defensive lines.

Defensive Metrics

- Defensive Third Tackles: Tackles made in the team's defensive third, near their own goal.
- Midfield Tackles: Tackles performed in the midfield area.
- Attacking Third Tackles: Defensive actions executed in the attacking third.

Possession Metrics

- Defensive Penalty Area Possession: How often a player controls the ball inside their own penalty box.
- Defensive Third Possession: The percentage of ball control maintained in the team's defensive third
- Midfield Possession: The amount of time a player or team retains the ball in the midfield zone.
- Attacking Third Possession: The level of ball control in the final attacking third.
- Attacking Penalty Area Possession: How often a player has possession inside the opponent's penalty box.

3.3 Data Preprocessing

To ensure fair comparisons between players, all raw performance statistics were normalized based on the number of minutes played by each individual. This normalization process is essential to mitigate biases that may arise from differences in playing time, ensuring a more equitable basis for analysis. For instance, a player who has played 1,000 minutes in a season cannot be directly compared to another who has played 3,000 minutes using raw statistics alone. By normalizing data per 90 minutes, we ensure that performance metrics reflect a player's actual efficiency and contribution, rather than just their total playing time.

Constructing the Similarity Matrix

Before building the network graph, we computed a similarity matrix using the cosine similarity metric. This matrix quantifies the statistical resemblance between players, where higher similarity values indicate stronger correlations in playing styles and performance patterns.

The cosine similarity metric was chosen because it effectively captures relative performance patterns rather than absolute values. Unlike direct numerical comparisons, which may be influenced by differences in playing style, cosine similarity focuses on how players perform in relation to each other. This makes it particularly useful for analyzing teams where players have different roles and tactical responsibilities, the cosine similarity between two players A and B is defined as:

$$\cos(\theta) = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}} \tag{1}$$

where:

- A_i and B_i represent the feature vectors of players A and B
- n is the number of statistical attributes used for comparison.

This approach was adapted from the methodology proposed by Mendes et al. [3], who applied graph theory to analyze player interactions and tactical decisions.

Data Filtering and Selection Criteria

To enhance the reliability of our analysis, we applied specific filtering criteria to exclude data points that could introduce noise:

- Players with fewer than five matches played were excluded. This ensures that only players with a meaningful contribution to the team's structure are considered.
- Goalkeepers were omitted from the similarity analysis. Since they have very distinct playing styles and responsibilities compared to outfield players, including them in the similarity matrix would distort the results.

By implementing these filtering steps, we ensure that our dataset accurately reflects the true performance patterns of players who have played a significant role in their respective teams.

Graph Construction

Each team is represented as a graph where:

- Nodes correspond to individual players.
- Edges represent the cosine similarity between players based on their statistical performance.
- Edge weights quantify the strength of the similarity between the players.

For enhanced visualization of the network, a color and size encoding scheme was implemented:

- Node color is determined by the PageRank score, with a gradient transition from light yellow (indicating lower values) to deep red (indicating higher values).
- Node size is based on the Authority score, meaning that larger nodes correspond to players who have greater influence in receiving passes.
- Edge color reflects the Weight of the connection, with darker shades representing stronger player similarities.

The graph analysis was conducted using Gephi[5], a widely recognized open source tool for network visualization and analysis. The Fruchterman Reingold [6] layout was adopted for its capacity to evenly distribute nodes and reveal underlying patterns and clusters. Furthermore, a thresholding filter was applied to remove edges with a Similarity Value below 0.95, ensuring that only robust player connections were incorporated into the network.

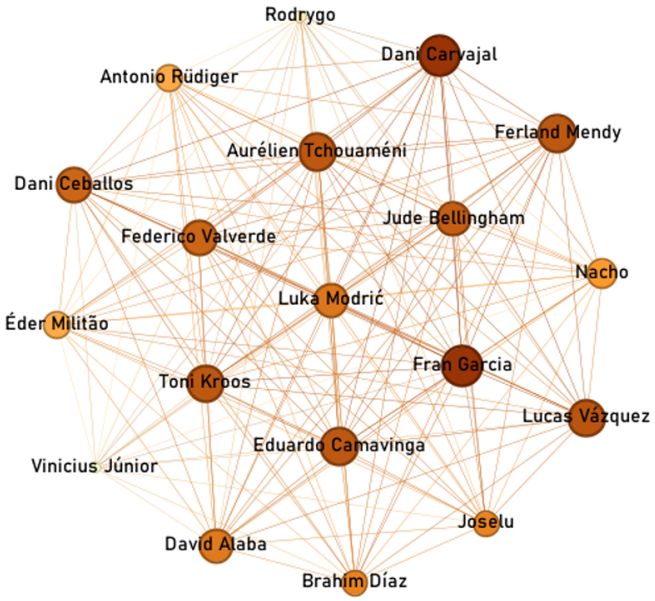


Figure 1. Visual representation of a championship-winning team's player similarity network. Node size reflects Authority Score, node color encodes PageRank value, and edge color indicates similarity strength (darker = stronger). Network layout generated using the Fruchterman-Reingold algorithm.

3.4 Metrics Used

To understand the structural differences between successful and underperforming football teams, we apply various graph based network metrics. These measures help quantify how players interact within the team's passing network, highlighting patterns that influence overall performance.

Authority Score - Measuring Key Pass Recipients:

The Authority Score indicates the significance of a player in receiving passes from key playmakers. A high Authority Score means that a player is trusted with the ball in crucial moments, often involved in offensive transitions or build-up play. Example: In teams like Manchester City, attacking midfielders such as Kevin De Bruyne typically have high Authority Scores because they frequently receive and distribute decisive passes in advanced areas.

Betweenness Centrality - Identifying Playmakers and Transition Points:

Betweenness Centrality highlights players who serve as critical connectors between different sections of the network. These players facilitate ball movement across the pitch, controlling how the team transitions from defense to attack. Example: A defensive midfielder like Sergio Busquets often has high Betweenness Centrality, as he acts as a bridge between defenders and attackers, dictating the flow of possession.

Hub Score - Identifying Key Distributors:

A player with a high Hub Score is a key distributor, channeling possession to other influential teammates. Unlike Authority Score (which measures reception of passes), Hub Score focuses on pass distribution. Example:

A fullback like Trent Alexander-Arnold may have a high Hub Score because he frequently delivers passes to key attackers, initiating offensive moves from wide positions.

PageRank Score - Evaluating Overall Influence:

Inspired by Google's PageRank algorithm, this metric evaluates a player's overall influence within the passing network, considering both direct and indirect connections. A high PageRank Score means the player is not only frequently involved but also interacts with other influential teammates. Example: In well-balanced teams, central midfielders such as Toni Kroos often have high PageRank Scores, as they continuously interact with key playmakers, defenders, and attackers, maintaining team structure

4 Results

4.1 Key Differences Between Championship-Winning and Relegated Teams

Betweenness Centrality - Measuring Team Dependence:

Our analysis indicates that championship-winning teams exhibit lower average Betweenness Centrality[2] , meaning that passing responsibilities are more evenly distributed among multiple players. This balanced structure reduces reliance on a single playmaker to transition the ball, making the team more adaptable and less predictable. In contrast, relegated teams tend to have higher Betweenness Centrality, meaning that only one or two key players dominate ball transitions. This structure increases tactical vulnerability, as opponents can focus on marking these pivotal players, disrupting the entire team's offensive flow. These findings align with the research by Lopez Pena and Touchette [2], who observed that weaker teams in the 2010 World Cup, such as Paraguay, relied heavily on a few players with high Betweenness Centrality. This over-reliance made them more susceptible to defensive pressure and strategic containment.

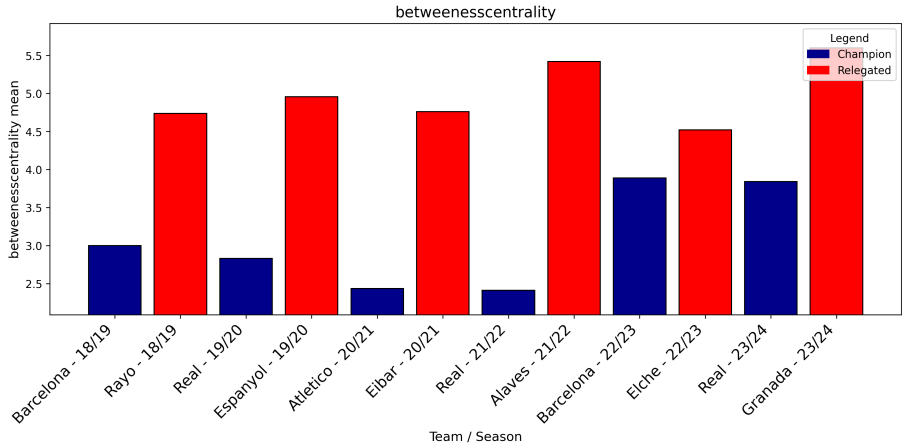


Figure 2. Comparison of average Betweenness Centrality between championship-winning and relegated teams. Higher values in relegated teams indicate reliance on a few key players for ball transitions.

Authority Score - Identifying Key Pass Recipients:

The Authority Score [3] reveals that championship-winning teams generally have a higher mean Authority Score than relegated teams. This suggests that successful teams possess multiple influential players capable of receiving key passes, reinforcing a structured and versatile attacking strategy. Conversely, relegated teams typically exhibit lower Authority Scores, indicating a heavy reliance on isolated players rather than a well-integrated passing system. This structural flaw reduces flexibility in build-up play and limits offensive options.

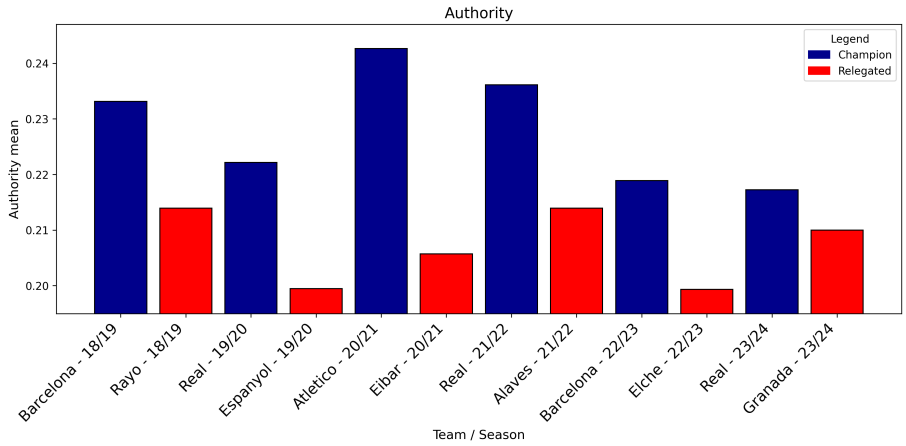


Figure 3. Distribution of Authority Scores across championship-winning and relegated teams. Successful teams exhibit higher average Authority, suggesting more players are involved as key pass recipients.

Hub Score - Evaluating Pass Distribution:

Analysis of the Hub Score [4] demonstrates that championship-winning teams tend to achieve higher values, signifying that a greater number of players actively contribute to ball distribution. This structure reflects a well-balanced and collaborative midfield, where multiple players participate in possession buildup. On the other hand, relegated teams show lower HubScores, suggesting inefficiencies in pass distribution. This often results in disconnected midfield structures and reduced attacking fluidity.

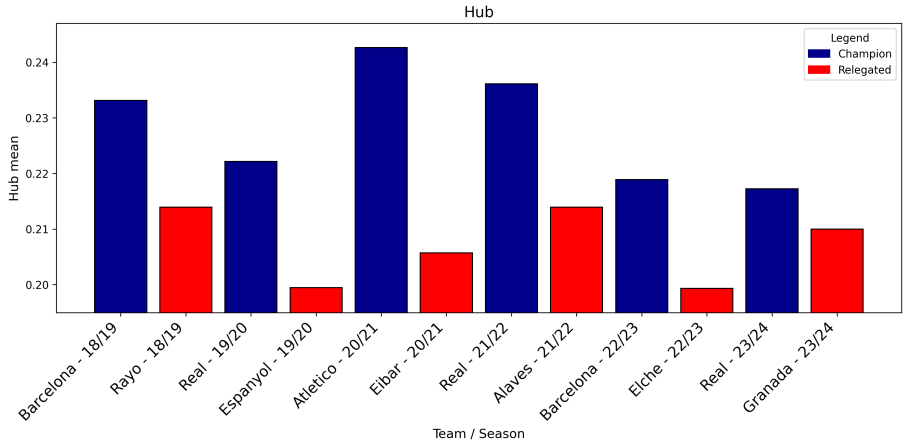


Figure 4. Hub Score values for championship-winning versus relegated teams. Higher scores among champions indicate a more distributed pass distribution structure and stronger midfield connectivity.

PageRank Score: Measuring Player Influence:

The PageRank Score [5] analysis further validates the trends observed in Authority Score. Championship-winning teams consistently show higher PageRank values, indicating that influential players are not only well-connected but also interact with other key teammates, thereby reinforcing a robust team network. In contrast, relegated teams exhibit lower average PageRank scores, which suggests that influence is concentrated in a few players rather than being widely distributed across the team. This can make the team predictable and easier to counter strategically.

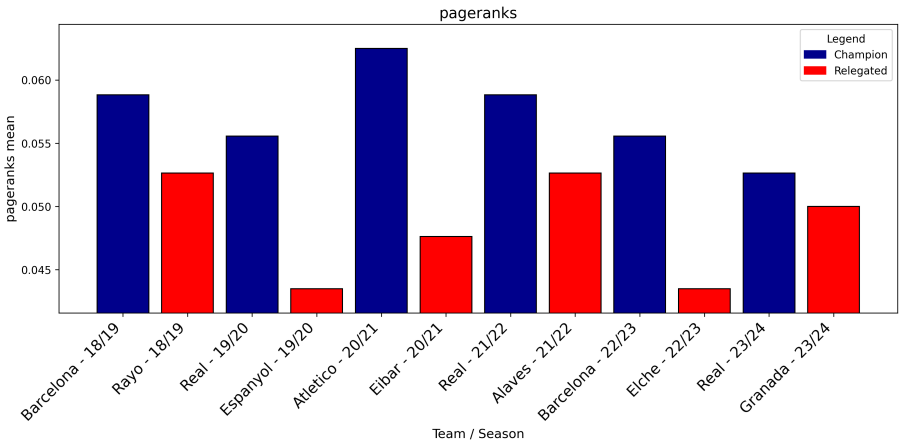


Figure 5. Evaluation of average PageRank Scores, highlighting player influence within team structures. Championship teams demonstrate higher and more balanced influence networks compared to relegated teams.

5 Conclusions

This study highlights the potential of graph theory in analyzing and distinguishing structural differences between championship-winning and relegated teams in La Liga. By constructing player similarity networks and applying centrality and connectivity metrics, we uncovered fundamental patterns that influence team success. Our findings indicate that successful teams exhibit higher levels of cohesion and distributed influence, as reflected in metrics such as degree centrality, closeness centrality, clustering coefficient, and PageRank. These attributes suggest a well-balanced structure where multiple players contribute effectively to team dynamics, reducing overreliance on a single playmaker and enhancing adaptability. Conversely, relegated teams tend to exhibit high betweenness centrality, signifying a dependence on a limited number of key players. This structural weakness increases tactical predictability, making such teams more susceptible to defensive disruptions. By providing a novel approach to team performance evaluation, this research contributes to football analytics, recruitment strategies, and tactical decision-making. Clubs can leverage these insights to design more resilient team structures, optimize player acquisitions, and implement strategies that enhance overall cohesion. Future studies could expand this analysis by incorporating temporal aspects of player interactions, examining different leagues, or integrating real-time match data to refine predictive models for team success. Ultimately, this work underscores the interdisciplinary potential of network science in sports analytics, offering a data driven framework for understanding and improving football team performance.

References

[FBref] FBref. Football Statistics and History. Available at: `https://fbref.com`.

[2] J. L. Peña and H. Touchette. A network theory analysis of football strategies. *arXiv preprint*, vol. arXiv:1206.6904, 2012.

[3] F. W. O. Mendes, T. M. R. Dias, da A. M. Silva, A. L. M. Silva, and da M. P. Silva. Technical decisions influences on dynamics and results in football: An analytical approach based on graph theory. In *Proceedings of the XLV Ibero-Latin-American Congress on Computational Methods in Engineering (CILAMCE-2024)*, Maceió, Alagoas, Brazil, 2024.

[4] F. Korte, D. Link, J. Groll, and M. Lames. Play-by-play network analysis in football. *Frontiers in Psychology*, vol. 10, pp. 1738, 2019.

[5] Gephi Consortium. Gephi: The Open Graph Viz Platform. Accessed: 2025-07-24, 2008.

[6] T. M. J. Fruchterman and E. M. Reingold. Graph drawing by force-directed placement. *Software: Practice and Experience*, vol. 21, n. 11, pp. 1129–1164, 1991.