Analyzing Team Performance Using Graph Theory Metrics

Lucas Cerqueira Portela
Department of Computer
Federal Center of Technological Education of Minas Gerais
Divinópolis, Brazil
lucas.portela@aluno.cefetmg.com

***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**

I. Introduction

Football, renowned for its **dynamic and multifaceted nature**, 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 **powerful analytical framework** to examine these complex relationships. 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 distribution, and tactical flexibility**.

This study aims to explore the **network characteristics** that differentiate **successful football teams** from **underperforming ones**. Utilizing **comprehensive player performance data** from *FBref* [1], 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.

II. Related Works

Numerous studies have explored the application of **network theory** in **football strategy analysis**. For instance, *López Peña and Touchette* [3] 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.* [4] 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.* [5] 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**.

III. Methodology

A. Data Collection

Player performance data was meticulously 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:

B. 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.

1) Player Information: This category includes fundamental details about each player, such as:

- **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.

2) Offensive Metrics: Offensive performance is assessed through metrics that measure a player's involvement in goal-scoring opportunities:

- **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 goal-scoring opportunities, such as key passes into the final third or through balls breaking defensive lines.

3) Defensive Metrics: Defensive metrics reflect a player's effectiveness in preventing the opposing team from advancing:

- **Defensive Third Tackles**: Tackles made in the team's defensive third, near their own goal.
- **Midfield Tackles**: Tackles performed in the midfield area, crucial for disrupting opposition play before they reach the defensive line.
- **Attacking Third Tackles**: Defensive actions executed in the attacking third, often as part of high-pressing strategies.

4) Possession Metrics: Controlling the Game: Possession metrics assess how well a player or team retains and controls the ball in different areas of the pitch:

- **Defensive Penalty Area Possession**: How often a player is in control of 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, a key indicator of offensive pressure.
- **Attacking Penalty Area Possession**: How frequently a player has possession inside the opponent's penalty box.

C. Data Preprocessing

To ensure **fair and consistent 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.

1) 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.

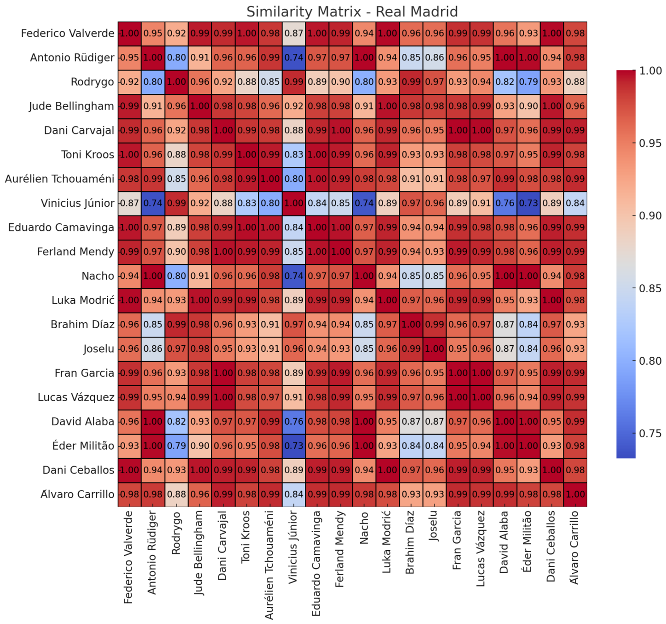

Fig. 1. Similarity Matrix of Real Madrid - 2023/24 Season

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**.

2) Mathematical Representation: Mathematically, 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** containing the performance metrics of players A and B.
- n is the **total number of attributes** used for comparison.

This approach was adapted from the methodology proposed by *Mendes et al.* [4], who applied graph theory to analyze player interactions and tactical decisions. By following this method, we constructed a similarity-based network model to evaluate team structures across multiple La Liga seasons.

3) 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 goalkeepers 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.

D. 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.

For enhanced visualization and interpretability 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** [2], a widely recognized open-source tool for network visualization and analysis. The **Fruchterman-Reingold** 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.

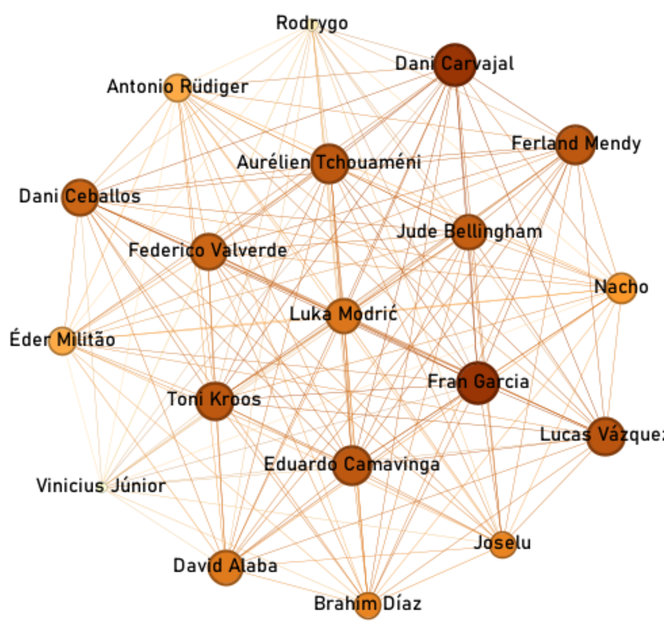

Fig. 2. Graph Representation of Real Madrid - 23/24

E. 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.

1) 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.

2) 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.

3) Closeness Centrality: Measuring Tactical Connectivity: **Closeness Centrality** assesses how quickly a player can reach all other teammates through passing sequences. A high score indicates that the player is **well-positioned to distribute the ball efficiently**, reducing reliance on long or risky passes.

Example: Players in *possession-based systems*, like *Luka Modrić* at Real Madrid, often score high in this metric due to their ability to quickly link up with multiple teammates.

4) Clustering Coefficient: Evaluating Local Team Cohesion: The **Clustering Coefficient** measures how closely connected a player is to his immediate teammates. High values suggest that a player is part of **a well-knit, collaborative unit**, which fosters smooth passing combinations and positional play.

Example: This is often high in *teams with a short-passing strategy*, such as *Barcelona's tiki-taka*, where players operate in tight triangles to maintain ball possession.

5) 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.

6) 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.

IV. Results

A. Key Differences Between Championship-Winning and Relegated Teams

1) Betweenness Centrality: Measuring Team Dependence: Our analysis indicates that **championship-winning teams** exhibit **lower average Betweenness Centrality**, 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 *López Peña and Touchette* [3], 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.

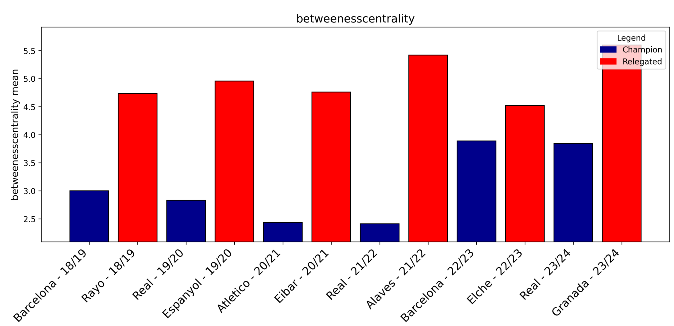

Fig. 3. Betweenness Centrality Mean for Winning and Relegated Teams

2) Authority Score: Identifying Key Pass Recipients: The **Authority Score** 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.

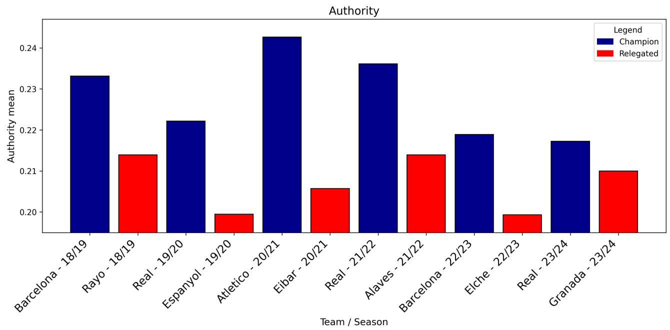

Fig. 4. Authority Score Mean for Winning and Relegated Teams

3) Hub Score: Evaluating Pass Distribution: Analysis of the **Hub Score** 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 Hub Scores**, suggesting **inefficiencies in pass distribution**. This often results in disconnected midfield structures and reduced attacking fluidity.

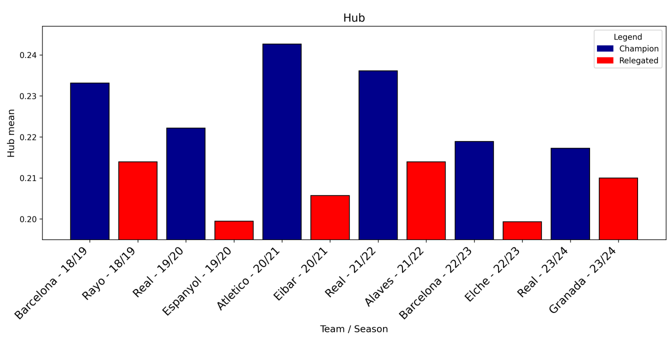

Fig. 5. Hub Score Mean for Winning and Relegated Teams

4) PageRank Score: Measuring Player Influence: The **PageRank Score** 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.

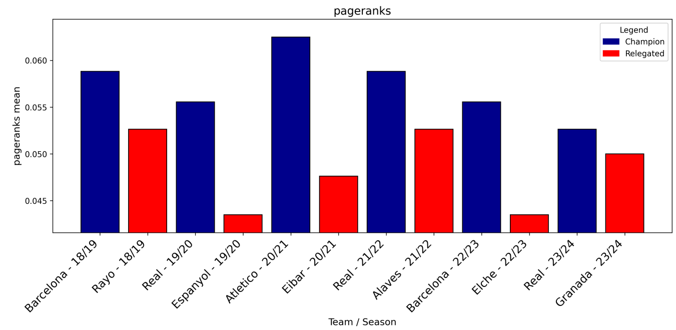

Fig. 6. PageRank Score Mean for Winning and Relegated Teams

5) Closeness Centrality: Evaluating Tactical Connectivity: **Closeness Centrality** measures how efficiently a player can reach other teammates within the passing network. A high Closeness Centrality score indicates that a player is **well-integrated into the team's structure** and can quickly distribute the ball across different areas of the field.

At the team level, **high average Closeness Centrality** suggests a playing style where **short passing distances** promote **fluid ball movement and tactical flexibility**. This characteristic is prevalent in **successful teams**, which maintain strong positional structures and ensure continuous ball circulation.

These results align with the findings of *López Peña and Touchette* [3], who demonstrated that *Spain's 2010 World Cup-winning team had exceptionally high Closeness Centrality values*. Their famous *"tiki-taka"* strategy relied on short, frequent passes, maximizing ball retention and minimizing opposition disruptions.

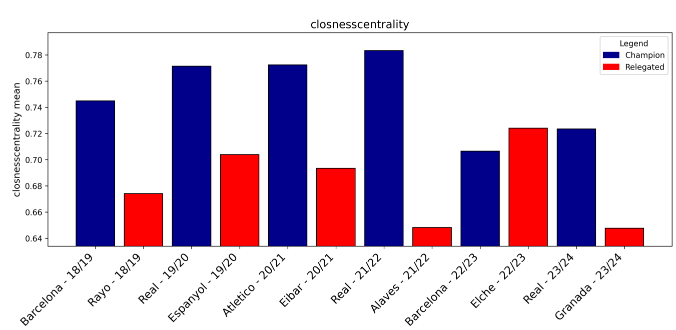

Fig. 7. Closeness Centrality Mean for Winning and Relegated Teams

B. Analyzing Graphs Differences

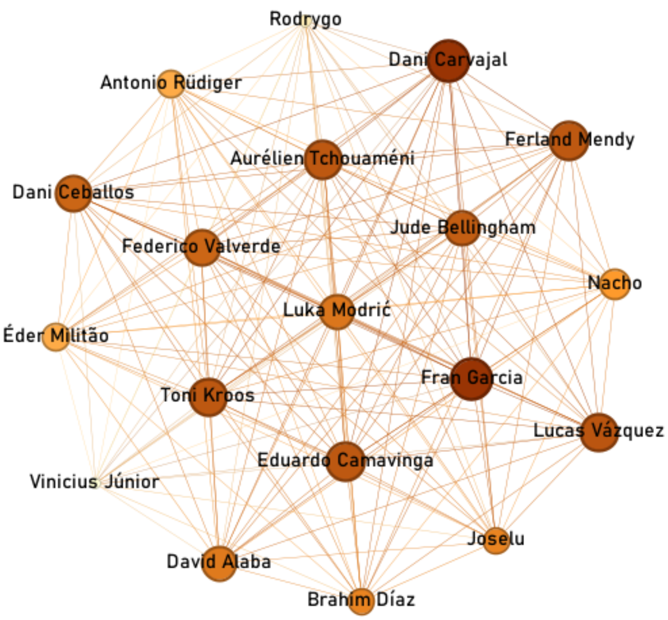

Fig. 8. Real Madrid 23/24 - Graph Representation

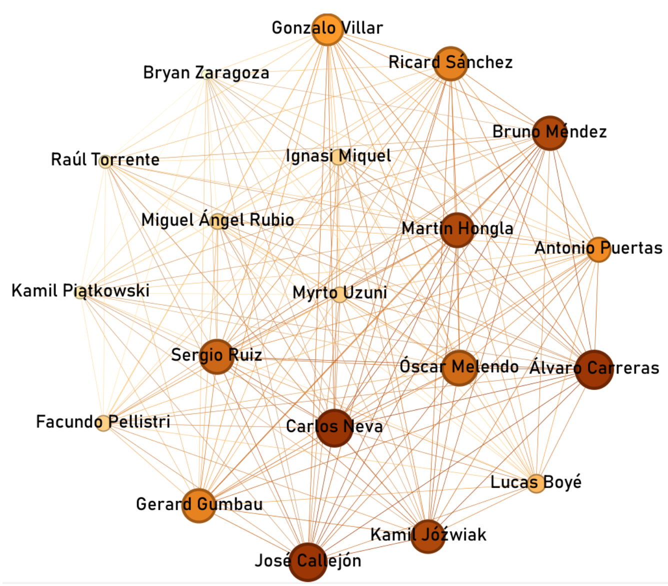

Fig. 9. Granada 23/24 - Graph Representation

To analyze the differences between the player similarity graphs of **Real Madrid (2023/24)** and **Granada (2023/24)**, we examine key structural characteristics such as *node connectivity, edge density, and centrality distribution*. These differences provide insight into how each team's composition and player interactions influence their overall performance.

1) Network Structure and Player Influence Distribution: The player similarity network of **Real Madrid** exhibits a well-balanced structure, where multiple players contribute significantly to the team's overall performance. The connectivity is evenly distributed, ensuring that no single player overwhelmingly dictates the network's dynamics. This structure aligns with the characteristics of successful teams, as identified by López Peña and Touchette [3], where a highly connected and balanced network fosters greater tactical adaptability.

Conversely, the **Granada** network reveals a more centralized structure, where a few key players (e.g., *Carlos Neva, José Callejón, and Sergio Ruiz*) have significantly higher influence within the network. This suggests that Granada relied heavily on specific individuals for transitions and overall gameplay, which increases tactical predictability and makes the team more vulnerable to defensive strategies.

2) Betweenness Centrality and Team Dependence: Betweenness centrality measures the extent to which certain players act as intermediaries in the flow of team dynamics. A comparative analysis reveals:

- **Real Madrid's network** exhibits a lower average betweenness centrality, meaning that the team's gameplay is well-distributed, reducing reliance on a single playmaker and ensuring multiple avenues for transitions and creativity.
- **Granada's network** demonstrates higher betweenness centrality concentrated in a few individuals, indicating a more hierarchical structure. As observed by López Peña and Touchette [3], teams that depend on isolated key players for transitions become more susceptible to defensive disruption when these players are effectively marked.

3) Clustering Coefficient and Team Cohesion: The clustering coefficient provides insights into the cohesion of player relationships within the network:

- **Real Madrid's high clustering coefficient** suggests a well-connected, cooperative structure where players exhibit high interaction density. This pattern is consistent with successful teams, such as Spain's 2010 World Cup squad, known for their cohesive and synchronized playing style [3].
- **Granada's network** presents a lower clustering coefficient, suggesting less integration among players, possibly leading to fragmented tactical execution and reduced gameplay fluidity.

4) Impact on Performance and Tactical Flexibility: These structural differences reveal how each team's network characteristics shape their in-game performance:

- **Real Madrid's well-distributed network** allows for greater tactical flexibility, making it difficult for opponents to isolate specific players and disrupt gameplay.
- **Granada's dependence on a few key players** reduces overall adaptability, making the team more predictable and vulnerable to defensive targeting strategies.

V. Conclusion

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 over-reliance 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

[1] FBref, "Football Statistics and History," Available at: *https://fbref.com*.

[2] M. Bastian, S. Heymann, and M. Jacomy, "Gephi: An Open Source Software for Exploring and Manipulating Networks," in *International AAAI Conference on Weblogs and Social Media*, 2009.

[3] J. López Peña and H. Touchette, "A network theory analysis of football strategies," arXiv preprint, arXiv:1206.6904, 2012. [Online]. Available: https://arxiv.org/pdf/1206.6904.

[4] F. W. O. Mendes, T. M. R. Dias, A. M. da Silva, A. L. M. Silva, and M. P. da 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, Nov. 2024. [Online]. Available: https://publicacoes.softaliza.com.br/cilamce/article/view/10166/7190

[5] F. Korte, D. Link, J. Groll, and M. Lames, "Play-by-Play Network Analysis in Football," in *Frontiers in Psychology*, vol. 10, p. 1738, July 2019. [Online]. Available: https://www.researchgate.net/publication/334545984_Play-by-Play_Network_Analysis_in_Football

[6] Y. Yajima and A. Inokuchi, "Refining Similarity Matrices to Cluster Attributed Networks Accurately," in *Proceedings of the IEEE International Conference on Data Mining Workshops (ICDMW)*, Sorrento, Italy, Nov. 2020, pp. 264-271. [Online]. Available: https://ieeexplore.ieee.org/document/9355957

[7] L. C. Portela, "Analyzing Team Performance Using Graph Theory Metrics", Federal Center of Technological Education of Minas Gerais, 2024. [Online]. Available: https://github.com/LucasPorteladev/AnalyzingTeamPerfomanceUsingGraph.git