Technical Decisions Influences on Dynamics and Results in Football: An Analytical Approach Based-On Graph Theory

Felipe Werneck de Oliveira Mendes, Thiago Magela Rodrigues Dias, Alisson Marques da Silva, André Luis Maravilha Silva, Michel Pires da Silva

Department of Computer Science, Federal Center of Technological Education of Minas Gerais
Rua Álvares de Azevedo 400 - Bela Vista - Divinópolis - MG - Brasil - CEP 35503-822
felipwerneck@gmail.com, {magela, alisson, maravilha, michel}@cefetmg.br

Abstract. Soccer players are meticulously and continuously observed and evaluated during games based on their roles, positions, and characteristics. In a scenario of great tactical complexity, it becomes essential to identify variations in athletes' performance, a dynamic that directly impacts the quality of the team during each competition. In this context, relevant challenges emerge and require meticulous analysis of the teams and their behavior during the game. Among these challenges, evaluating the player's performance is an important aspect that is deeply influenced by the strategic approach adopted by the coach. In this paper, we explore the issue of individual player performance, given the different philosophies and strategies that coaches use during each match. In this context, we introduced an analysis approach based on graph theory, aiming to evaluate the relationships between players, the quality of assistance in the match, and the coaches' strategies to develop a model capable of identifying the different impacts of strategic composition in individual gameplay. The results show that training strategies adopted can considerably influence gameplay quality and, consequently, the player's performance during a match.

Keywords: Gameplay Analysis; Tactical Formation; Performance Evaluation

1 Introduction

Football, one of the most popular sports globally, places player performance as one of the fundamental pillars in developing successful strategies for each confrontation. In this context, the tactical base can consider, as part of its strategy, an attacker with a high goal rate, a midfielder with a great presence on the field, and a defender skilled in tackling. Regardless of the tactical principles addressed, it is important to recognize the dynamics and complexity of the player's actions on the field, which are directly influenced by the various decisions made.

The complexity of player behavioral analysis drives elite teams to invest in computational tools to improve their strategies. In this context, the continuous production of significant volumes of data in real-time stands out, enabling the implementation of a diversity of decision-making techniques [1]. This data offers a wide spectrum of possibilities for exploration, ranging from predictions of match results to suggestions for tactical repositioning and decisions about substitutions and reconfiguration on the field. In [2], the data collected represents both passes on the field and the interrelationships between players and identifies the strategy adopted by the team during the match.

Among the widely adopted collection mechanisms, those that track players' movements and the ball during a football match stand out. Data obtained through these mechanisms are applied to generate various *insights* about the performance and positioning of the team [3]. Although these methods effectively deal with data and employ advanced computational strategies, the dynamism of decisions often needs to consider the coach's characteristics as part of the information explored, which continues to be a significant challenge in the concepts used.

This study explores how a team's strategies and technical decisions can affect its performance on the field. To that end, we take an innovative approach, investigating graph theory using complex networks to analyze the impact of strategic decisions from the perspective of the technical team and its decision-making model. The study gathers data from individual players and the team to observe how the established strategies and positioning decisions influence the players' behaviors and the team's overall performance on the field, offering a fresh perspective on team performance analysis.

To illustrate the relevant impacts that the coaching team and its strategies have on the players' individual and collective performance, the remainder of this work is structured as follows: In Section 2, a comprehensive contextualization details the challenges surrounding this issue; In Section 3, related work is discussed and emerging computationally based strategies are outlined; In Section 4, the proposed solution to map the relationships above is presented in detail; In Section 5, the results obtained are presented and discussed; Last but not least, in Section 6, there are conclusions and suggestions for future work.

2 Background

The evolution of computational models for performance analysis has proven to be an indispensable tool in several contexts, especially when dealing with significant volumes of data, much of which is generated in real time. In disciplines related to sport, such data predominantly originates and is observed from the actions of athletes during the practice of their activities. Due to these actions' dynamic and complex nature, analysis using conventional tools represents a significant challenge, leading different areas of sports science to seek performance solutions through computing.

Analyzing a highly dynamic, multiple-relationship environment where different conditions or approaches impact the observed domain space outlines relevant and significant challenges, especially in computational solutions. The growing tendency to use emerging computational solutions is evident in this scenario, many of which are based on graph theory and complex networks. In the sporting context, models of complex systems built on dynamic and non-linear mathematical foundations highlight the profound challenges in proposing effective analysis processes, often conditioned on observing behavioral relationships in real-time [4].

According to Miranda et al. (2021) [5], applying emerging computing concepts to the analysis of sports data offers the advantage of creating effective statistical mechanisms. These mechanisms can make relevant predictions that can, for example, influence the outcome of a match, identify deficiencies in team composition, and make behavioral predictions, among other possibilities. Research such as that of Yan et al. (2019) [6] and Talha et al. (2018) [7] use graph theory and complex networks to examine the behavior of individuals when they are part of a team, observing their influence and the impact they have on the analysis network. Studies like these bring a new perspective of evaluation to the universe of sports sciences, allowing a deep understanding of the dynamics behind different success stories in the world of sports while at the same time making it possible to amplify the team's potential through strategies that better encompass the characteristics evidenced by the modeling of relationship networks between athletes.

In this line of evidence, this work aims to investigate how actions external to network relationships can impact the influence of athletes, substantially altering the behavior of established interactions and their relevance. To this end, we explored introducing these conditions into the network based on the decisions and deliberations carried out by the technical team of a football team. This approach enables a more comprehensive and realistic analysis of the sporting scenario, taking into account not only the interactions between athletes but also the external context and the strategies adopted by the technical team as impact factors on the team's results and performance.

2.1 Related Works

Earlier research has explored the application of network theory to the analysis of football strategies. The studies reviewed highlight a range of possible approaches, from analyzing passing networks in football matches to predicting outcomes based on sentiments expressed on social media.

Efforts introduced by Miranda et al. (2021) [5] address the application of graph theory and sentiment analysis in predicting the results of football matches. The research focuses on the analysis of the social network Twitter, highlighting the 2019/2020 Premier League season as the study context. During a specific game week, 3,000 tweets were evaluated. The study proposes a machine learning model to automate pre-match data collection and classify expressed sentiments, aiming to calculate a team's probability of victory. The results obtained reveal that, when considering the centrality of the network, it is possible to discern the intensity of support from fans and, therefore, predict results with an accuracy of 54%, disregarding random game statistics but using the knowledge expressed in fans' interactions.

In investigations conducted by Andrienko et al. (2017) [3], a computational approach is addressed to detect and quantify pressure relationships during football matches. The research proposes a parameterized formula to numerically express the pressure exerted by players, considering the relative position of defenders concerning the evaluation target, which can be the ball or an opposing player. The analysis of the German Bundesliga shows that the active pressing technique represents a fundamental element for the success of elite clubs such as FC Barcelona, Borussia Dortmund, and Atlético de Madrid. This study contributes significantly to data analysis in the football context, providing new perspectives on collective behavior and the game strategies teams adopt.

Another relevant study in the analysis of players is presented by Pena et al. (2012) [2]. In this study, the authors investigate the exchange of passes during the 2010 World Cup matches. Representations of the tactical strategies adopted by football teams were obtained through the construction of weighted and directed networks. The visualizations of these networks made it possible to identify different game patterns, high-activity zones, and possible vulnerabilities in the team structure. The study also calculates centrality measures for individual players, highlighting their relative importance, popularity, and the impact of their absence on team dynamics. The findings offer valuable insights for coaches and sports journalists, enriching their understanding of team performance and the strategies adopted.

Although sports analysis studies have progressed significantly, there needs to be more literature regarding computational discussion that specifically addresses the coaching staff's deliberate perspectives and the impacts of these deliberations on playing conditions and individual players' performance on the field. This study aims to fill this gap, thoroughly analyzing how technical decisions influence game conditions through multiple interactions on the field and how these conditions are reflected in the match's final result.

3 Impact of Technical Decisions on Gameplay and Results in Football Matches

As mentioned, decisions made by the technical team impact the quality of players' performance on the field, in line with work related to neuroscience and neuro-social synchronism, as evidenced by professor Miguel Nicolelis in studies such as [8]. In this context, we seek to analyze such evidence by investigating the reduction or amplification of interactions between key players based on their performances under different technical decision approaches. To this end, we use graph mapping strategies associated with analysis and path models, which allow us to evaluate whether these performance variations can be identified exclusively through computational strategies without the need for neuro-social data.

To investigate the impact of technical decisions on gameplay and players' performance, we use an exploration flow that integrates data analysis processes based on information about players' relationships and their broad individual aspects on the field. To this end, we used as a source of observation the statistical and historical information presented on the FBREF website [9].

Based on the data available on the website mentioned above, we established an analysis of relationships using graph theory and breadth-first search algorithms as a strategy. To this end, we normalized the data to ensure uniformity between players. We then applied a filtering criterion, selecting only players who played more than ten games in the season. In this process, we excluded goalkeepers due to their unique characteristics that do not influence the deliberate study in this work. We finish by building a database with each player's name and respective metrics.

After extracting the input data, normalization ensures consistency across multiple value variations. This standardization process aligns the metric values under a uniform evaluative scale. To achieve this, we employ a statistical dispersion strategy [10, 11]. Specifically, we use the StandardScaler function from the sklearn API [12] to facilitate this normalization.

With the data normalized, the next step aimed to define a similarity matrix. We use the performance metrics extracted for each observed player to do this. To calculate this matrix, we sought, among the available metrics, to identify the covariance factors between players, basing the calculations on the Pearson correlation, as detailed in the equation 1.

$$R_{ij} = \frac{C_{ij}}{\sqrt{C_{ii} \times C_{jj}}} \tag{1}$$

Where R represents the correlation coefficient matrix, and C represents the covariance matrix. To normalize the values, each matrix element is divided by the square root of the products of the corresponding diagonal elements in the covariance matrix. This procedure ensures that the values are in the range of -1 to 1, which is the valid range for Pearson correlation coefficients.

After extracting the covariance matrix according to equation 1, we build a model representative of the relationships and their relevance. We used the principles of graph theory, where players were represented as nodes (or vertices) and their interactions on the field as edges. The value of these interactions is represented by the values established in the covariance matrix. During this process, the relationships between two players were considered based on a pre-established *threshold*. In the context of this study, which aims to highlight the impacts of technical decisions, we chose to keep any positive value of the matrix as a valid connection relationship.

Finally, we adopted a variant of the breadth-first search (BFS) algorithm to analyze the connections and distances of a specific player concerning the rest of the team. This allowed us to evaluate his performance guidelines on the field, considering his interactions with other players. At this stage, our objective was to identify players who, despite playing for different teams, had similar characteristics and demonstrated the ability to influence the game's dynamics. Figure 1 details an overview of the workflow adopted to carry out the process proposal of this work.

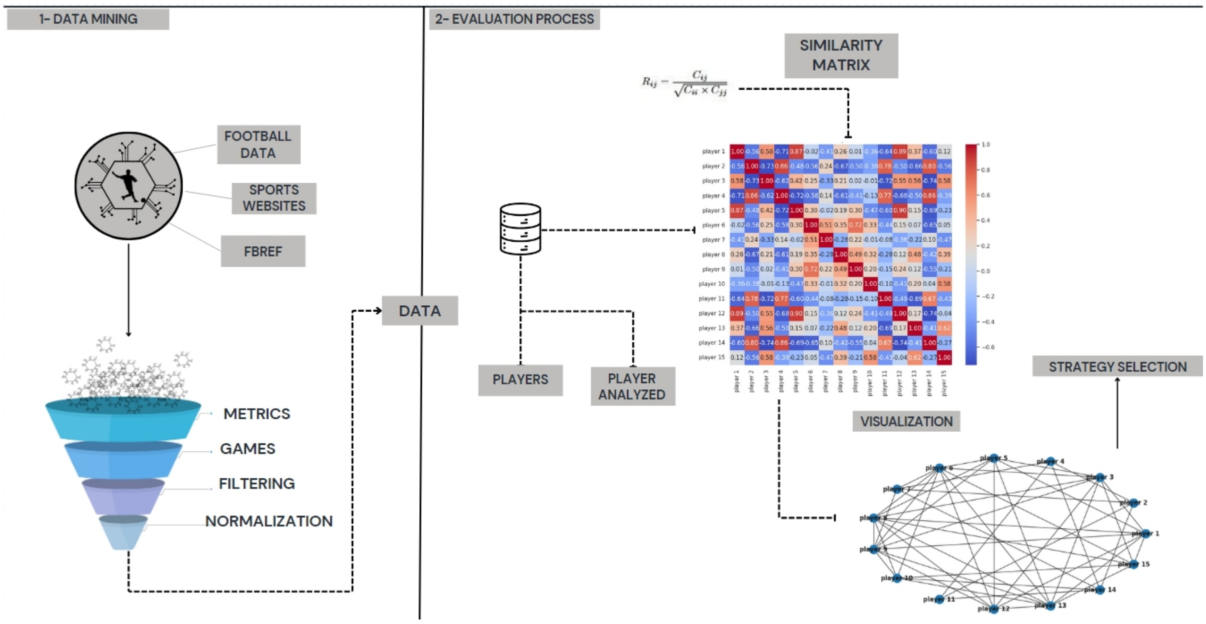

Figure 1. Multi-stage process to carry out the analysis of the coach's influence on the performance and dynamics of the football presented by the players. 1- Data mining. 2- Evaluation process.

As Figure 1 demonstrates, the proposed workflow is outlined by two fundamental stages: data mining and evaluation. In this context, the results of this study are interpreted based on interactions between players, emphasizing the importance of data analysis as much as conclusive evaluation. In the modeling stage, we use the functionalities available in the NetworkX API [13] to normalize the relationships obtained as an auxiliary resource. Understanding the potential of these stages and the observed relationships is discussed in the next section, contextualizing the results obtained.

4 Results

Once the information retrieval process and data mining used to compose the exploration domain presented in this work are presented, this section discusses the results obtained from the analysis methodology. Presentations are carried out following the multi-state process 1, employing such concepts to elucidate the main factors of influence exerted by the technical team on the player and his behavior on the field.

To argue about behavior and gameplay, we adopted as a starting point the relationship between different metrics presented by interaction networks and player performance data during a pre-determined period. Among the possible metrics available in interaction network models, we consider the agglomeration coefficient and the average path length between the study's target player and the other athletes on the field to be the most appropriate.

Regarding the analysis of the defined metrics, the agglomeration coefficient was used to observe the density of relationships, examining how greater density impacts the amplification of the interactions of the players under study within the group of athletes on the field. In turn, the average path length was used to evaluate the degree of interaction between the target player and other team members based on the assumption that a greater average length reflects an increase in movement variations and game strategies, which, in practical terms, implies an increase in the options for moves made.

To compose the discussions and report the efficiency of the methodology used in this study for the behavioral analysis of players from the perspective of decisions and technical strategies, two players who played in different teams throughout their careers and the analysis period were considered. The objective of this choice was to analyze the performance and behavior of these players under the guidance of technical teams with different strategic methodologies. Of the players available in the observed study domain, we chose to examine the performance and behavior of Erling Haaland and Robert Lewandowski. This decision is based on the high quality of these players and the independence of their technical characteristics concerning the team in which they play.

4.1 Erling Haaland

Erling Haaland, the Norwegian player, has notable spells in his history at Borussia Dortmund and Manchester City, playing for these clubs in the 2022 and 2023 seasons, respectively. At Borussia Dortmund, Haaland was under the guidance of the technical team led by Marco Rose, while at Manchester City, he was part of the team led by Pep Guardiola. Both teams have different tactical strategies, as illustrated in Figure 2.

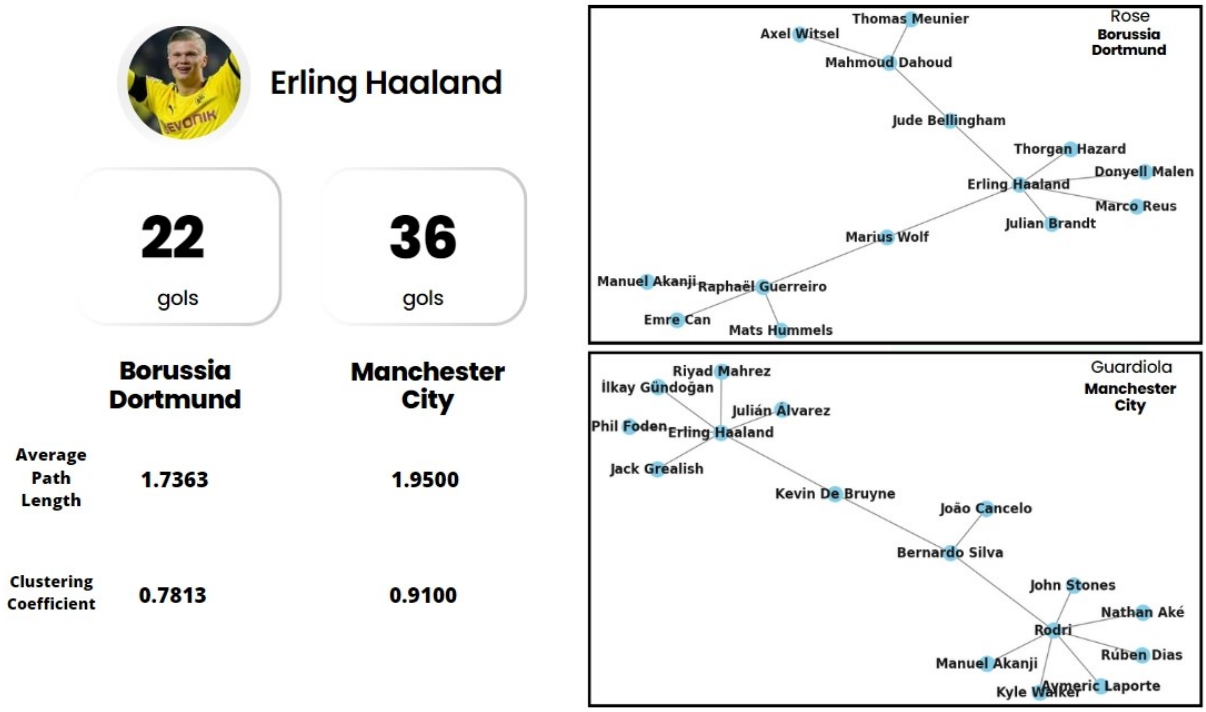

Figure 2. Relationship between the number of goals Erling Haaland scored for the Borussia Dortmund and Manchester City teams and the relationships on the field through the strategic and technical decisions used by each team.

Considering the data presented in Figure 2 and analyzing Erling Haaland's performance in the two teams studied, it is observed that his performance at Manchester City was significantly more effective. This increase in effectiveness is evidenced by the number of goals achieved during his time with the team.

When examining the interaction networks in both teams, it is noted that their agglomeration coefficient at Manchester City is higher than that at Borussia Dortmund. This may indicate that their performance on the field is associated with a greater need for cohesion between players, resulting in a more collaborative interaction network. This factor can amplify the creation of scoring opportunities as players establish better connections, facilitating the exchange of passes more efficiently and quickly. For Haaland, this cohesive network means he receives more support and better assists from his teammates, enhancing his chances of finishing and success. This underlines the crucial role of a cohesive network and tactical flexibility in player performance.

Manchester City's higher average path length could also be interpreted as a testament to their greater tactical flexibility. This implies that the team adopts a wide variety of movements on the field and interrelationship strategies, offering a plethora of play options. This flexibility allows Haaland to position himself better and explore the spaces between the opponent's lines, taking advantage of opportunities that arise throughout the matches due to the team's tactics.

4.2 Robert Lewandowski

Robert Lewandowski, the Polish player, made a significant move in 2023 to join the Barcelona team, where he adapted to the game dynamics under the influential management of Xavi. Following his successful stint at Bayern Munich under Hansi Flick, this transition showcased the impact of different coaching styles on a player's performance, as shown in Figure of 3.

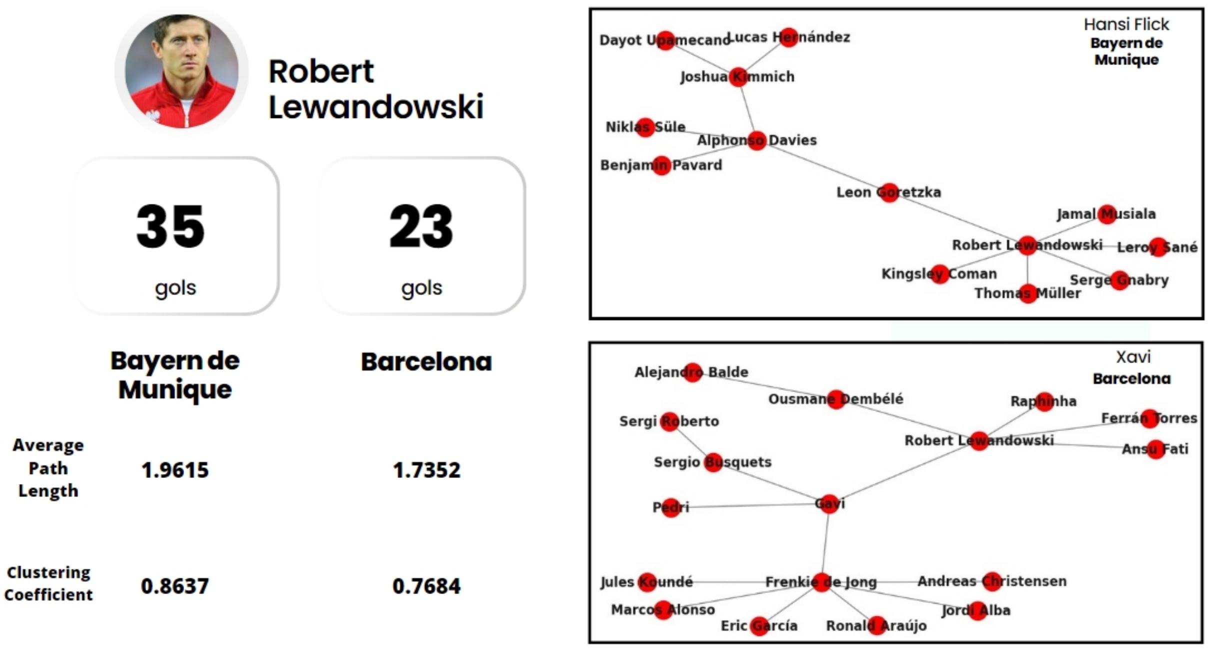

Figure 3. Relationship between the number of goals Robert Lewandowski scored for the Bayern Munich and Barcelona teams and the relationships on the field through the strategic and technical decisions used by each team.

When analyzing Robert Lewandowski's performance, it is observed that he had a superior performance while playing at Bayern Munich, where he scored 35 goals throughout the season. The relationship metrics confirm the previously mentioned characterizations, which include specific characterizations.

In such a context, Bayern Munich's higher crowding coefficient, which measures [specific metric], may indicate greater team cohesion, facilitating collaboration and efficiency on the field. Furthermore, Bayern Munich's higher average path length, which measures [specific metric], suggests greater tactical flexibility, allowing various movements and strategies. This contributes to the development of Lewandowski's characteristics, directly reflecting on his performance on the field and his results.

4.3 Brandstorm and Overview

Considering the clustering coefficient and average path length metrics as player performance indicators on the field reveals implicit signatures that reflect individual strategic needs. This insight significantly contributes to improved strategic decision-making during matches. Furthermore, certain players exhibit a greater dependence on interpersonal relationships in their performance characteristics.

Integrating these characterizations with technical strategies and team composition can enhance individual and team performance on the field. This integration can improve results, expanding the possibilities for exploring spaces and playing opportunities during matches.

Observations depicted in Figures 2 and 3 show that, despite different interaction networks, game dynamics are partly influenced by the collaborative model defined on the field. This influences player behavior and interactions, a concept deeply explored by Professor Miguel Nicolelis, who details various synchronization models from a cerebral perspective. These interaction networks, as addressed by Nicolelis, are significant influencing factors. The dynamics developed by the technical team induce different behaviors, making aligning these dynamics with the observed characterizations in this study relevant.

5 Conclusions

The study introduced in this paper is of significant importance to the field of sports science and analytics. It evaluates the behavior of various players on the field and examines their performance and the dynamics shaped by the technical team. It underscores the connections between individual attributes and team synchronization during gameplay. Analysis of interaction network metrics reveals a direct correlation between on-field behavior and player correlations and interaction patterns. Moreover, specific strategic compositions amplify individual traits, enhancing conditions for success.

While it has been established that technical dynamics influence player relationships and that field decisions correlate with team synchrony and success metrics, this study paves the way for further exploration. Future steps involve expanding the evaluation to encompass a broader range of metrics from diverse player perspectives and assessing the consistency of the identified characterizations. These efforts aim to construct signatures derived from interaction networks that reflect the distinctive traits of different teams across varying success levels. Such signatures will serve as inputs for different artificial intelligence strategies, enhancing decision-making support based on individual and collective team behaviors.

References

[1] T. A. Herberger and C. Litke. The impact of big data and sports analytics on professional football: A systematic literature review. In T. A. Herberger and J. J. Dötsch, eds, *Digitalization, Digital Transformation and Sustainability in the Global Economy*, pp. 147–171, Cham. Springer International Publishing, 2021.

[2] J. L. Peña and H. Touchette. A network theory analysis of football strategies, 2012.

[3] G. Andrienko, N. Andrienko, G. Budziak, J. Dykes, G. Fuchs, von T. Landesberger, and H. Weber. Visual analysis of pressure in football. *Data Mining and Knowledge Discovery*, vol. 31, n. 6, pp. 1793–1839, 2017.

[4] N. Balague, C. Torrents, R. Hristovski, K. Davids, and D. Araújo. Overview of complex systems in sport. *Journal of Systems Science and Complexity*, vol. 26, n. 1, pp. 4–13, 2013.

[5] C. Miranda-Peña, H. Ceballos, L. Hervert-Escobar, and M. Gonzalez-Mendoza. Previsão de resultados de futebol por meio de análise de sentimento: uma abordagem da teoria dos grafos. In M. Paszynski, D. Kranzlmüller, V. Krzhizhanovskaya, J. Dongarra, and P. Sloot, eds, *Ciência Computacional – ICCS 2021*, volume 12747 of *Notas de aula em Ciência da Computação*. Springer, Cham, 2021.

[6] G. Yan, N. Watanabe, S. Shapiro, M. Naraine, and K. Hull. Unfolding the twitter scene of the 2017 uefa champions league final: social media networks and power dynamics. *European Sport Management Quarterly*, vol. 19, n. 4, pp. 419–436, 2019.

[7] A. Talha, M. Simsek, and I. Belenli. The wisdom of the silent crowd: predicting the match results of world cup 2018 through twitter. *International Journal of Computer Applications*, vol. 182, n. 27, pp. 40–45, 2018.

[8] P.-H. Seng, S. Rajangam, G. Lehew, M. A. Lebedev, and M. A. Nicolelis. Interbrain cortical synchronization encodes multiple aspects of social interactions in monkey pairs. *Scientific Reports*, vol. 8, pp. 4699, 2018.

[9] FBref. Estatísticas de futebol e história. `https://fbref.com/pt/`. Acesso em: 19 maio 2024, 2024.

[10] M. Rothschild and J. Stiglitz. Increasing risk i: A definition. *Journal of Economic Theory*, vol. 2, n. 3, pp. 225–243, 1970.

[11] D. A. McQuarrie. *Statistical Mechanics*. Harper & Row, New York, NY, 1st edition, 1976.

[12] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, and others. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.

[13] A. Hagberg, P. Swart, and D. S Chult. *NetworkX: High-productivity software for complex networks*. NetworkX Developers, 2021.