Chaos, Solitons and Fractals 178 (2024) 114355

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

journal homepage: www.elsevier.com/locate/chaos

A multilayer network framework for soccer analysis

Álvaro Novillo [a,b], Bingnan Gong [c], Johann H. Martínez [d], Ricardo Resta [e], Roberto López del Campo [e], Javier M. Buldú [a,b,*]

[a] Complex Systems Group, Universidad Rey Juan Carlos, Móstoles, Madrid, Spain
[b] Grupo Interdisciplinar de Sistemas Complejos (GISC), Spain
[c] Institute of Physical Education and Training, Capital University of Physical Education and Sports, China
[d] Interdisciplinary Mathematics Institute, Department Mathematical Analysis & Applied Mathematics, Universidad Complutense, Madrid, Spain
[e] Mediacoach, LaLiga, Madrid, Spain

ARTICLE INFO

Keywords:
Network science
Sports analytics
Multilayer networks
Eigenvector centrality
Soccer passing networks

ABSTRACT

In this paper, we define a novel methodology for analyzing soccer matches and teams using spatial multilayer networks. Departing from a segmentation of the pitch into $h \times v$ regions, we create 2-layer networks that capture the exchange of ball possessions between teams throughout a match. To assess the significance of each node, we employed eigenvector centrality measures within the constructed multilayer network. Furthermore, we introduce three additional metrics, namely the *leakage*, *recovery* and *switching factor*, which quantify the possession transitions between layers. Finally, we apply our methodology to analyze the performance of Spanish soccer teams over an entire season, using the aforementioned multilayer parameters, and discuss the relation with the playing style and ranking of soccer teams.

1. Introduction

Network Science has emerged as a powerful framework for understanding complex systems across various domains, including social networks, biological networks, and technological networks [1]. Recently, researchers have recognized the relevance and potential of Network Science in the realm of sports analysis, where the study of sports encompasses a myriad of interconnected components such as team dynamics, player interactions, and strategic decision-making [2,3]. These components can be effectively examined and understood using network-based approaches, shedding light on the underlying structures and dynamics that shape sporting events. In this way, network science has become an increasingly important tool for studying the dynamics of relational structures in sports teams during practice and competition [4].

The application of network science in sports has led to the emergence of new fields, such as *network physiology of exercise* (NPE), which focuses on understanding how physiological states and functions emerge and improve the efficacy of exercise in health and sport performance [5]. Social network theory and analysis have also been used to gain insights into advantageous techniques and insights that can be offered in sport management and organizational behavior [6,7]. Furthermore, network analysis has allowed for the identification of research hotspots in sports science and the exploration of the relationship between knowledge networks and scientific performance [8].

In the context of a soccer match, the dynamics and outcomes are intricately intertwined with the interplay between two teams, representing two distinct networks. Hence, to comprehensively understand team performance, it is crucial to analyze the passing network of a team in conjunction with the network of the opposing team [3,9]. This approach allows for drawing insights into how a team adapts its gameplay in response to the opponent and the topological structures that contribute to superior outcomes. Notably, studies in other domains focusing on network-of-networks have revealed that when networks establish connections, essential properties of the ensemble system undergo modifications [10,11]. However, a multilayer description of a football match, encompassing the interaction of two layers representing the internal passes of each team, remains absent. Such a framework would be pivotal in unraveling the evolution and adaptability of teams throughout a match, as these aspects cannot be comprehended without considering the opponent's responses. The competing nature of the two team-networks, driven by the pursuit of a shared resource and objectives that inherently entail interaction and competition with other networks, suggests that Network Science can offer novel perspectives to comprehend and forecast optimal strategies [12]. Previous research consistently demonstrates that ball possession cannot guarantee winning in the game [13,14]. However, the team with organized and continuous passing actions can be purposefully executing tactical

* Corresponding author at: Complex Systems Group, Universidad Rey Juan Carlos, Móstoles, Madrid, Spain.
E-mail address: javier.buldu@urjc.es (J.M. Buldú).

https://doi.org/10.1016/j.chaos.2023.114355
Received 7 November 2023; Accepted 3 December 2023
Available online 14 December 2023

objectives and setting the pace and rhythm of the game [15–17]. Furthermore, there has been an excessive focus on high-percentage ball possession or low-percentage ball possession, while neglecting the network spatial value of collective patterns of ball possession and ball transitions [18–20]. Considering the game's nature, team performance during a match is influenced by the interactions between the opposing sides. Traditional research has placed a high focus on static performance indicators while overlooking the mechanisms of mutual interaction within performance behaviors. In a soccer game, the attacking team aims to create attacking space and scoring opportunities through continuous team tactical actions, while the defending team seeks to constantly restrict the opponent's performance through off-the-ball tactical actions. In pursuit of this goal, the tactical relationship between the attacking and defending sides in a soccer match exhibits the physical phenomenon of coupled oscillators [21,22]. The opposing sides engage in self-organizing activities, interacting with varying intensity and across different temporal and spatial scales [23]. What is even more significant is the oversight of the tactical behavioral interaction during transition play in matches. Hence, the challenge for coaches is how to train players and teams to understand the complexity of the game and adapt their behaviors to the constantly changing and unpredictable complex and dynamic match environment. Incorporating a multilayer network perspective into football analysis holds promise for advancing our understanding of team dynamics and enhancing strategy development within the sport. With this target in mind, in this paper, we introduce a novel methodology to construct and analyze multilayer soccer networks. Departing from a spatial partition of the pitch into $h \times v$ regions, we construct the pitch passing networks, which capture the structure of the passing patterns elaborated by each team [23]. For a given match, two pitch passing networks are obtained (one for each team), which correspond to each of the two layers of a multilayer network. Next, we include the inter-layer links detecting the ball exchange between teams and finally analyze the structure of the multilayer network with different centrality metrics. Importantly, we introduce three new network parameters to evaluate the ability of retaining and recovering the ball by each team.

2. Materials and methods

2.1. Dataset description

The dataset used were provided by Opta [24–26] and encompassed comprehensive pass data of all teams participating in the Spanish national league (*"La Liga"*) during the 2018/2019 season. The dataset comprises a total of 380 matches, with each of the 20 teams of the competition playing 38 matches. The dataset encompasses vital information of each pass made during the competition, including (i) the player responsible for executing the pass, (ii) the player who received the pass, (iii) the positional coordinates representing the position of the sender and receiver players and (iv) the timestamp indicating the precise moment when the pass occurred.

2.2. Multilayer network construction

We use the same methodology of [23] to create the pitch-passing networks of each team. We divide the pitch into $N = h \times v$ patches; where N is the number of nodes (pitch areas), h is the number of horizontal subdivisions (X direction) of the pitch and v is the number of vertical subdivisions (Y direction). Without loss of generality, we set $h = 4$ and $v = 5$, thus leading to a division of the pitch into $N = 20$ regions. Each region will be a node of a network. A link from node i to node j is created when a pass is made from region i to j and we assign a weight that quantifies the total number of completed passes at each direction. In this way, we obtain weighted-directed networks with an adjacency matrix that is not symmetric. Note that each team has its own pitch network. Next, we assign the network of the home team to be the layer-1 of a 2-layer multilayer network, while the away team's pitch network is considered layer-2. It is crucial to label the nodes of the multilayer network adequately. The N regions of each pitch network will be projected into the $2 \times N$ nodes of a 2-layer network. In this multilayer network, the first N nodes correspond to the N regions of the pitch where the home team has the ball (home team pitch network). Nodes ranging from $N + 1$ to $2N$ correspond to regions of the pitch where the possession belongs to the away team (away team pitch network). By defining the node number in this way, node i and node $N + i$ correspond to the same region of the pitch but with possessions to the home (i) or away ($N + i$) teams. Finally, we create the inter-layer links by accounting for the actions where teams lose possession, and the ball goes to the other team. For example, a pass that is started from region i by the home team and is intercepted at region j by the away team will create an inter-layer link from node i to node $N + j$. In this way, intra-layer and inter-layer links constitute the elements of $\mathbf{G}$, a multilayer pitch passing network of a soccer match. Note that the supra-adjacency matrix of $\mathbf{G}$ has size $2N \times 2N$ and it is, by construction, non-symmetrical. See Figs. 1–2 for two different examples of multilayer pitch networks.

2.3. Parameter definition

Definition 1 (*Eigenvector Centrality*)**.** A measure of the importance of a node in a network. It is based on the concept of eigenvectors, which are a special set of vectors that do not change direction when multiplied by a matrix. In the context of networks, the eigenvector centrality of a node is proportional to the sum of the eigenvector centralities of its neighbors (that is, nodes with high eigenvector centralities are those which are connected to important nodes which are, in turn, connected to important nodes, and so on). The formal definition of eigenvector centrality is given by the equation:

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x} \tag{1}$$

where $\mathbf{A}$ is the adjacency matrix of the network, $\mathbf{x}$ is the eigenvector corresponding to the largest eigenvalue of $\mathbf{A}$, and λ is the corresponding eigenvalue. In the context of our analysis, $\mathbf{A}$ can be either the adjacency matrices of the pitch passing networks of the home and away teams ($\mathbf{A_1}$ and $\mathbf{A_2}$, respectively) or the supra-adjacency matrix of the match, $\mathbf{G}$.

Definition 2 (*Leakage*)**.** Percentage of possessions that goes to the rival team at jth zone of the pitch:

$$L_j^i = \frac{P_{out,j}^{*i}}{P_{in,j}^i} \tag{2}$$

where $P_{in,j}^i$ is the total in-degree of the area j of layer i, with $i \in [1, 2]$, which joins (i) the inner in-degree of the layer of each zone (that is the number of passes at zone j) plus (ii) the inter-layer in-degree $P_{in,j}^{*i}$ accounting the balls that are regained from the rival team. $P_{out,j}^{*i}$ is the inter-layer out-degree of zone j (the balls that are lost to the rival team).

Definition 3 (*Recovery*)**.** Percentage of balls that have been recovered from the rival at jth zone of the pitch:

$$R_j^i = \frac{P_{in,j}^{*i}}{P_{in,j}^i} \tag{3}$$

where $P_{in,j}^i$ is the total in-degree of the area j of layer i and $P_{in,j}^{*i}$ is the inter-layer in-degree of zone j (the balls that are recovered from the rival).

Definition 4 (*Switching Factor*)**.** The relative influence of the leakage and recovery of a team compared to the total number of passes:

$$\langle S \rangle^i = \frac{1}{2} \sum_j \left(\frac{P_{in,j}^{*i} + P_{out,j}^{*i}}{Passes_j^i} \right) \tag{4}$$

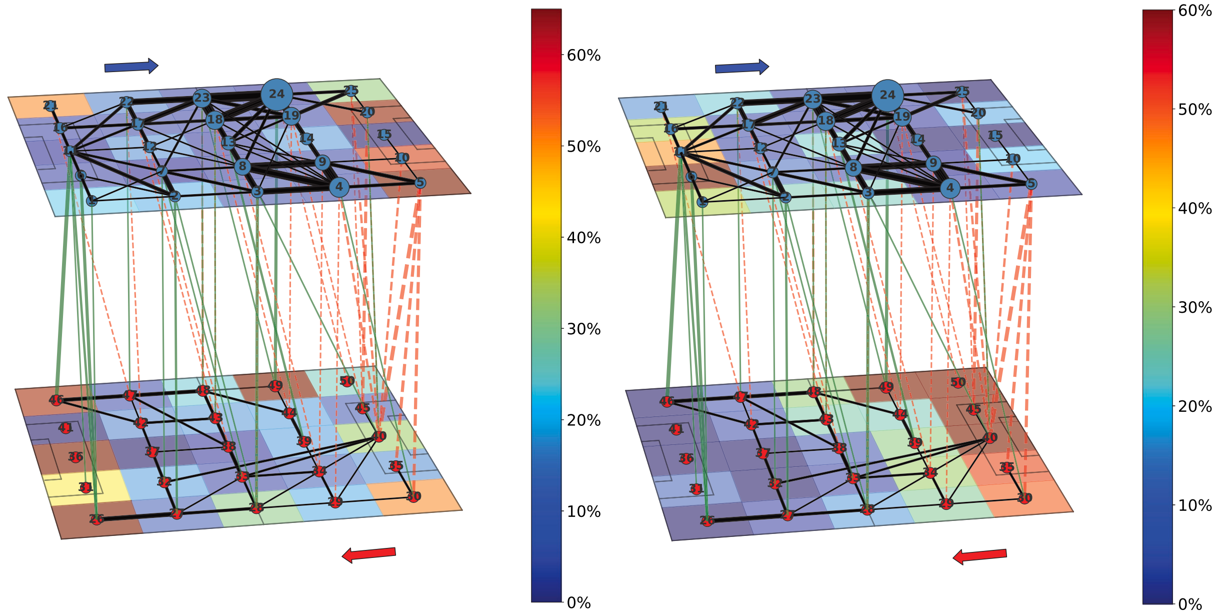

Fig. 1. Multilayer network of the match between Real Madrid (home team, top layer) and Atlético de Madrid (away team, bottom layer). Inter-layer links account for the ball losses between teams. Green inter-links indicate ball transfers from the away team (bottom) to the home team (top), while red inter-links represent the reverse direction. Node size corresponds to their eigenvector centrality, and link width is proportional to the ball transfer between nodes, which can be either passes or changes of possession. The heatmap on the left represents the *leakage parameter* of each zone of the pitch. On the right, the heatmap accounts for the *recovery parameter*. Arrows indicate the attacking direction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

where $Passes^i_j$ stands for the in-degree of pitch zone j, and $i \in [1,2]$ indicates what team is considered, with $i = 2$ for the home team and $i = 2$ for the away team.

2.4. Statistical analysis

The z-score was employed to quantify the deviation of the parameters observed for a specific team relative to the mean values observed across all teams. In statistics, the z-score (also known as a standard score) is a measure that describes a value's relationship to the mean of a group of values. It represents the number of standard deviations a given data point is from the mean of the dataset. Z-scores are used to compare data points from different datasets or to identify outliers within a single dataset. The z-score is calculated using the following formula:

$$z = \frac{x - \mu}{\sigma}$$

Where z is the z-score, x is the value of the data point, μ is the mean of the dataset and σ is the standard deviation of the dataset.

3. Results

First, we obtained the networks G_i for each of the i matches of the competition, with $1 \leq i \leq 380$, and computed the eigenvector centrality of each node (region of the pitch) of the network and its corresponding leakage, recovery and switching parameters. For example, Fig. 1 shows the multilayer network corresponding to the match between Real Madrid and Atlético de Madrid. Layer 1 (top) corresponds to Real Madrid (home team), and layer 2 (bottom) to Atlético (away team). The width of the intra- and inter-layer links is proportional to the number of passes between regions. Inter-links going from the bottom to the top layer are plotted in green, while red inter-links go from the top to the bottom layer. To ease the visualization of the network, only links with a weight higher than one are plotted (i.e., with two or more passes). The size of the nodes is proportional to their eigenvector centrality (i.e., importance) in the multilayer network. Finally, the color of the regions is proportional to either the leakage (Fig. 1, left) or the recovery (Fig. 1, right) parameters, according to the color coding shown in the color bars of Fig. 1.

We can observe how the top layer (Real Madrid) has more intra-layer connections than the bottom one. Furthermore, the passing network of Real Madrid is densely connected at regions close to the center but over the opponent's part of the pitch. All these signals reveal the dominance of Real Madrid during that match, who accumulated 66% of possession. We can also observe how the majority of balls lost by Real Madrid were close to Atlético's box, as indicated by the red links going to the bottom layer. Reversely, Atlético's losses of possession were mainly distributed at regions of the pitch close to Real Madrid's box (green inter-layer links).

The eigenvector centrality reinforces the idea of Real Madrid's dominance. Node 24 is the one accumulating more centrality (importance), but furthermore, we can observe how eigenvector centrality is basically distributed over the nodes of the top layer. Concerning the leakage parameter (colors over the pitch on the left plot), regions of the pitch closer to the opponent's goal are the ones accumulating more leakage. While in general, the central regions of the pitch are the ones with lower values. It is worth noting how in the case of Real Madrid, the leakage at the center of the opponent's goal is not that high. The reason is the low number of passes that arrive into this region since when the ball enters, it is controlled by Atlético in most cases. This fact is captured by the recovery parameter (right plot), where we can see how Atlético de Madrid has the highest value at the center of its box.

3.1. Multilayer networks and node centrality in soccer

Our first aim was to understand what regions of the pitch play a crucial role in the multilayer passing networks obtained from soccer datasets. With this objective in mind, we constructed the G_i networks of the 380 matches of the season and averaged them, considering all teams. The result is shown in Fig. 2, where the top layer corresponds to the home teams of the 380 matches, and the bottom layer contains the away teams. We computed the eigenvector centrality of each region of the pitch for the home and away teams and, second, the corresponding leakage and recovery parameters (left and right plots, respectively).

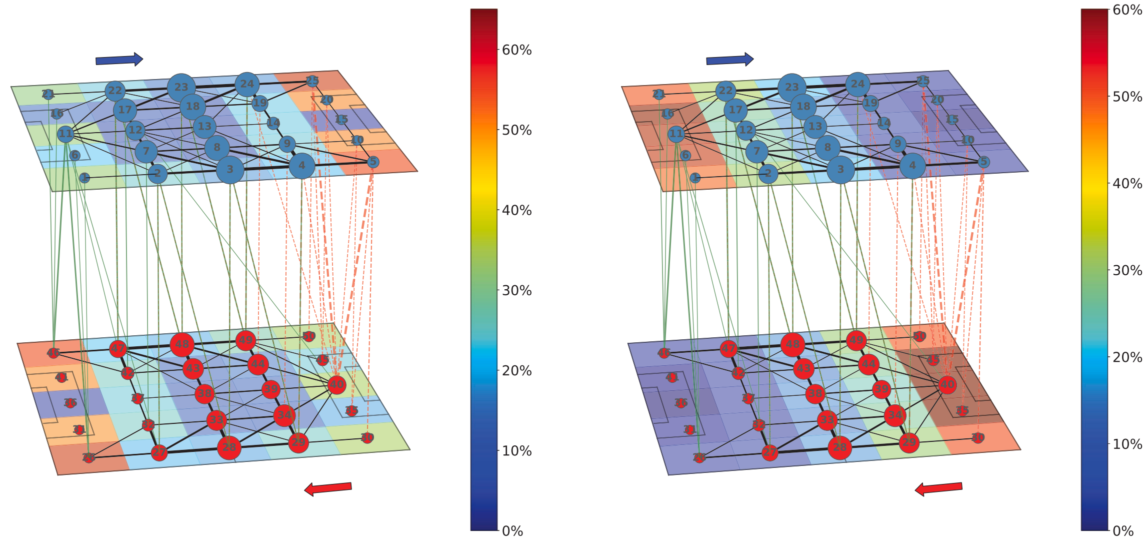

Fig. 2. Average multilayer networks encompassing all teams in the league. Dashed links are created when the possession of the ball changes from one team to another. Green links indicate ball transfers from the away team (bottom layer) to the away team (top layer), while red links represent the reverse direction. Node size corresponds to eigenvector centrality. The heatmap of the figure on the left represents the *leakage* parameter of each pitch zone. On the right, the heatmap accounts for the *recovery* parameter. Arrows indicate the attacking direction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

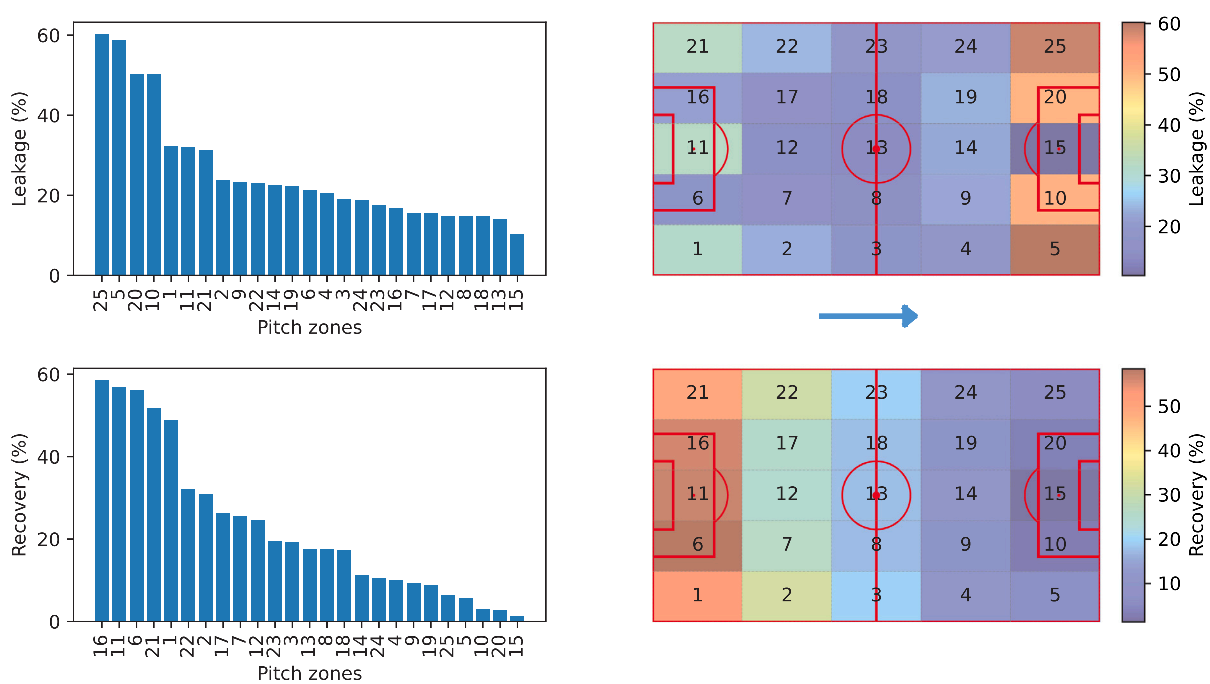

Fig. 3. Average leakage (top plots) and recovery (bottom plots) parameters for all teams in the league. In the left plots, bars are sorted from the pitch zone with the highest leakage/recovery to the lowest one. The right plot contains the color map of the leakage (top) and recovery (bottom) values.

Fig. 2 shows the nodes with a size that is proportional to their importance (eigenvector centrality) in the multilayer network. We can observe that regions at the pitch's center accumulate more importance, while regions at the corners and the lateral of the boxes are those with the lowest importance. This result is somehow expected since the number of passes in the lateral regions is lower. Interestingly, the region where the goalkeeper stays the majority of the time also has a high eigenvector centrality. This is due to the passes made by goalkeepers, which normally begin in front of their goals. Another interesting result of Fig. 2 is that the importance of the pitch regions is highly symmetric with regard to the layers, indicating that the effect of playing at home or away is not relevant in the distribution of eigenvector centrality.

The analysis of the leakage (left plot) indicates that the center of the pitch is the place where the percentage of possession loss is the lowest, as indicated by the bluish colors. The boxes on both sides are the regions where the leakage shows intermediate values, while the corners of the pitch are the places with the highest leakage. Note that the leakage considers where the ball is lost but not where it is recovered; for example, an uncompleted pass sent from the corner to the box is lost at the corner by one team and recovered at the box by the other team. The multilayer network containing the recovery parameter (Fig. 2, right plot) shows how the closer to the box, the higher the recovery parameter. This is expected, since the defending players' density increases as we get closer to the goal.

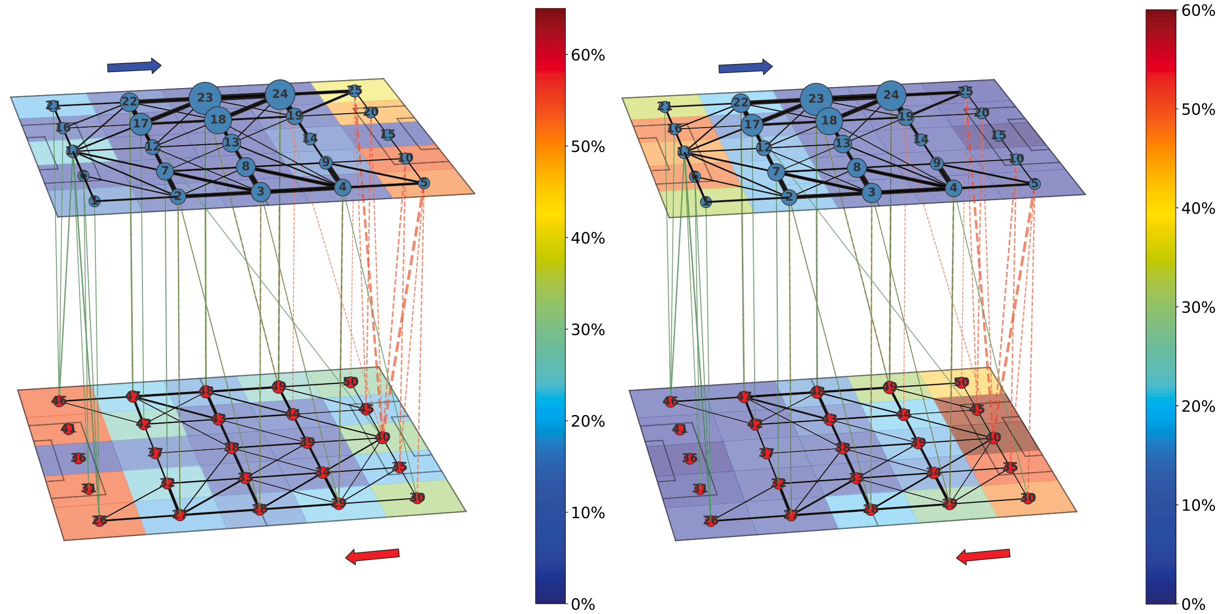

Fig. 4. Real Madrid multilayer network (throughout the season). The top-layer corresponds to Real Madrid pitch passing network, while the bottom layer is the average of its rivals. Dashed links indicate ball losses between teams. Green links reflect the average ball transfers from the rival teams to Real Madrid, while red links represent the reverse direction. Node size is proportional to the eigenvector centrality, and edge width corresponds to link weight (number of ball transfers). The heatmap on the left is proportional to the *leakage* parameter of each zone of the pitch. On the right, the heatmap represents the *recovery* parameter. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

This behavior can be better observed in Fig. 3, where we plot the leakage and the recovery parameters of the home teams per region together with a 2-dimensional color map (top layer). A similar plot (not shown here) is obtained for the away teams. It is worth noting that region 15, which is at the center of the opponent's box, has a low leakage parameter compared to the sides of the box. The reason is that it is very difficult to complete passes inside this region. When a pass sent from surrounding regions into region 15 is lost, the leakage is assigned to the region from where the pass was sent and not to node 15. However, as explained before, the recovery parameter accounts for the passes recovered at each specific region, and we can see (Fig. 3, bottom plots) that region 11 (the center of the box) is, together with the other two regions of the box (16 and 11, respectively), the one accumulating higher values (as one may expect).

3.2. From team to team

Using this methodology, we can analyze the multilayer passing networks of specific soccer teams to unveil the properties of their paying patterns. For example, Fig. 4 contains the average multilayer networks of Real Madrid vs its rivals, computed for all matches played during the season. We can observe how Real Madrid accumulates much more centrality than its rivals, as indicated by the larger size of its nodes (proportional to the eigenvector centrality). This is due to the fact that Real Madrid normally has larger possessions than its rivals. Furthermore, nodes on the left side of the pitch accumulate higher eigenvector centralities, indicating a preference for attacking through this side.

Concerning the distribution of leakage, we can see how Real Madrid's layer has lower values than its rivals, with the corners being the regions with the highest leakage. It is also worth noting that the recovery parameter has the highest values at its own box, accumulating low to moderate values at the rest of the pitch (the lower, the closer to the opponent's goal).

Fig. 5 shows the same results for the average multilayer network of FC Barcelona, calculated along the whole season. In general terms, results are qualitatively similar, despite FC Barcelona accumulates higher centrality at the nodes placed at the center of the pitch and, furthermore, it does not show a preference in playing at one of the lateral lanes.

However, not all teams have similar multilayer networks. For example, Fig. 6 shows the networks obtained for Getafe CF, a team characterized by a very direct pattern of play, with short possessions and a very intense pressure after losing the ball. As a consequence, the multilayer networks show a high density of inter-layer links, due to the high number of possession exchanges. The size of the nodes, proportional to their eigenvector centrality, shows how the rivals of Getafe CF have longer possessions, as indicated by the small size of Getafe's nodes. It is worth noting how different is the distribution of the leakage and recovery parameters of Getafe compared to those of Real Madrid and FC Barcelona. Concerning the leakage (left network), we can observe the high values it accumulates over all regions of the pitch, indicating that Getafe loses the ball quite fast (and frequently), probably due to its direct play. However, the recovery parameter (right network) is much higher, overcoming both Real Madrid and Barcelona and recollecting the ability of Getafe to recover the ball.

The comparison with the performance of other teams is better unveiled by the z-score, which quantifies the deviation of the leakage/recovery parameters of each region of the Getafe's pitch compared with the average of all teams of the league. Values of z-score higher than zero indicate positive deviations while negative values are obtained when the parameter is below the average. In this way, Fig. 7 shows the regions of the pitch where Getafe is particularly strongest at its leakage and recovery parameters. The yellowish and reddish colors indicate that Getafe has a positive z-score both for the leakage and recovery parameter, in some case close to two times the standard deviation of both parameters. Appendix, containing the z-scores of Real Madrid and FC Barcelona, shows that each team has its particular distribution of z-scores.

3.3. Comparison between teams

Next, we compared the average values of all teams in the competition. Fig. 8 shows a box plot of the eigenvector centrality accumulated

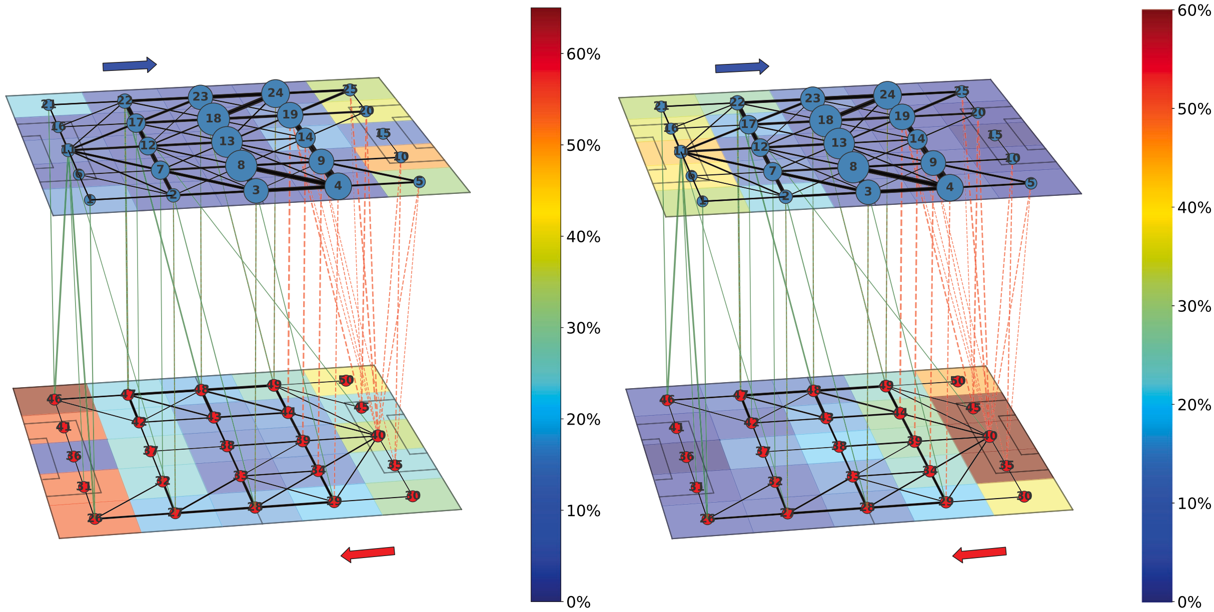

Fig. 5. FC Barcelona multilayer network. The top-layer corresponds to FC Barcelona pitch passing network, while the bottom layer is the average of its rivals. Inter-layer dashed links indicate possession losses (red) and recoveries (green). Node size corresponds to eigenvector centrality, and edge width is proportional to the number of passes (intra-layer links) or ball recoveries (inter-layer links). The heatmap on the left is proportional to the *leakage* parameter of each zone of the pitch. On the right, the heatmap represents the *recovery* parameter. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

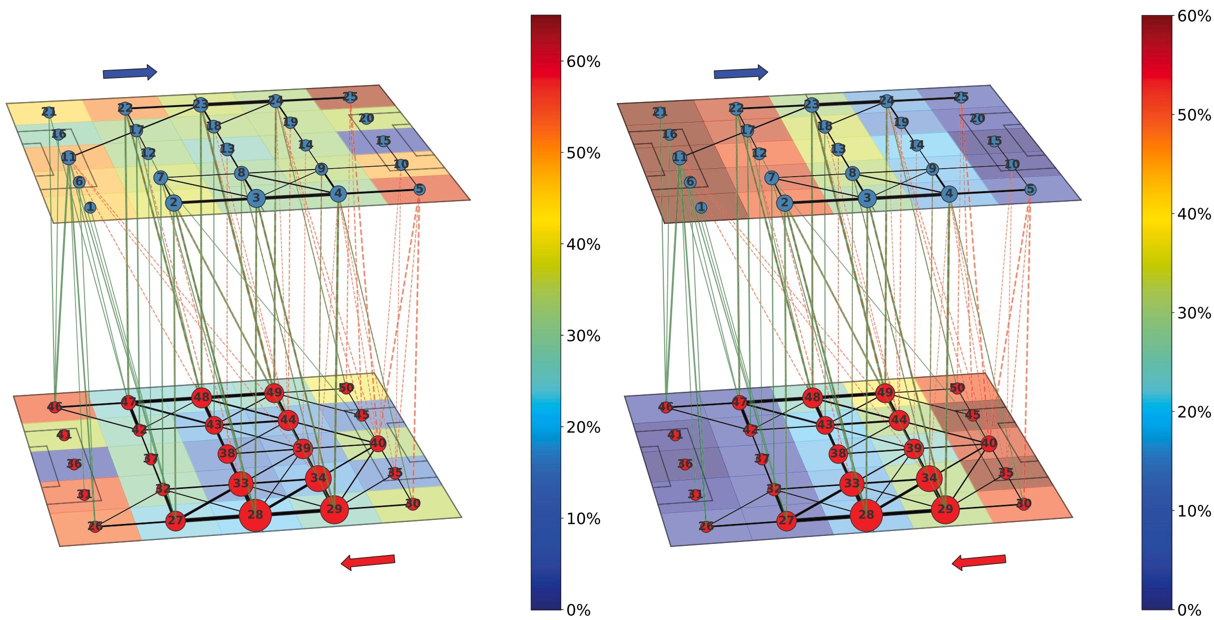

Fig. 6. Getafe CF multilayer network. The top-layer corresponds to Getafe pitch passing network, while the bottom layer is the average of its rivals. Inter-layer dashed links indicate possession losses (red) and recoveries (green). Node size corresponds to eigenvector centrality, and edge width is proportional to the number of passes (intra-layer links) or ball recoveries (inter-layer links). The heatmap on the left is proportional to the *leakage* parameter of each zone of the pitch. On the right, the heatmap represents the *recovery* parameter. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

by all nodes of each team in the multilayer network of each match, with teams ordered by their median and the boxes indicating the 25–75 percentage of the values. The accumulated eigenvector centrality is a classical way of measuring what layer is dominating over the other [12], in this case, with regard to the passing network. F.C. Barcelona, Real Betis and Real Madrid are, respectively, the teams with the highest eigenvector centrality due to a way of playing based on retaining the possession of the ball. At the bottom of the distribution, Real Valladolid, Getafe CF and Deportivo Alavés are the teams with the lowest eigenvector centralities (see Table 1 for a summary of the average values of each team).

Fig. 9 contains a similar box plot of the switching factor of each team. In this case, Getafe CF is leading the ranking due to the high alternance of possession during its matches thanks to a playing style that combines high pressure on the opponent combined with a very direct play. On the contrary, FC Barcelona and Real Madrid are the teams with the lowest switching factor, both of them characterized by long possessions.

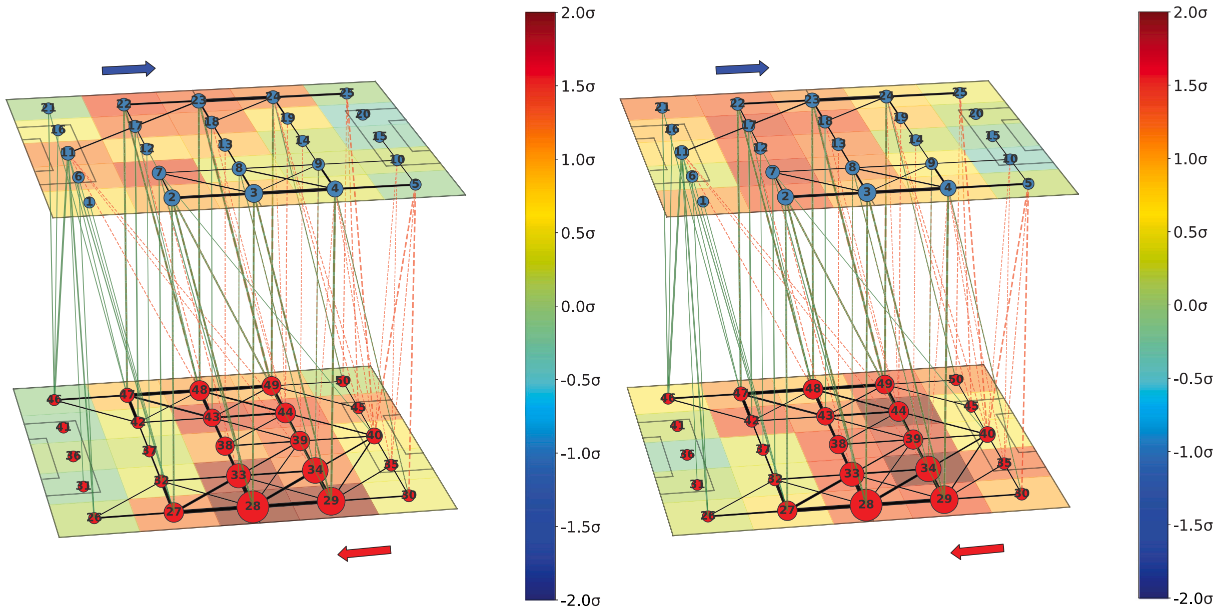

Fig. 7. Z-score of Getafe CF multilayer network. Nodes and links are obtained in the same way as Fig. 6 (see caption). Colors indicate the z-score of each region of Getafe's pitch compared with the rest of *LaLiga*'s teams. Yellowish and reddish colors indicate that Getafe has a positive z-score both for the leakage and recovery parameter (see colorbar on the left of the networks), revealing that its values are higher than the average. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

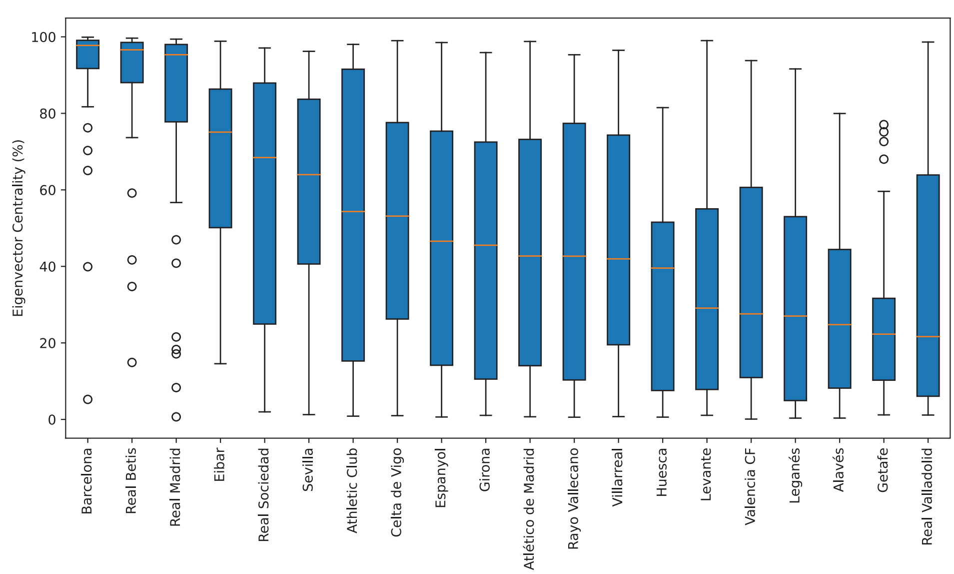

Fig. 8. Box plot of the accumulated eigenvector centralities of each team across all matches.

Finally, in Fig. 10, we have plotted the percentage of accumulated eigenvector centrality as a function of the number of points of each team at the end of the season. We can observe how Barcelona has the highest percentage of eigenvector centrality and, at the same time, the highest number of points. Despite being a positive correlation, the accumulated eigenvector centrality is not a good indicator of the final position in the ranking. This is due to the fact that there are teams, such as Atlético de Madrid, which ranked second, whose style of play is based on short possessions, leading to a "weak" passing network and, therefore, accumulating low values of eigenvector centrality.

As a summary, Table 1 contains the average values of the accumulated eigenvector centrality, the leakage and recovery parameters and the switching factor of all teams, ordered by their corresponding eigenvector centrality. It is worth noting how the accumulated eigenvector centrality negatively correlates with the rest of the parameters. However, the relation between them is not trivial, and we can see teams like SD Eibar or Athletic Club departing from this correlation.

4. Discussion

The multilayer network framework in the analysis of soccer matches introduces a parallel perspective for the exploration of team tactical behaviors, offering a novel lens through which to comprehend team dynamics by considering pitch spatial distribution and ball transitions. From the average multilayer networks encompassing all teams in the league, we observed that the central regions of the field hold greater significance, with a substantial number of passing performance behaviors associated with these areas [27]. This additional evidence reinforces the importance of the midfield zone during the offensive phase [28]. Conversely, regions situated at the corners and the sides

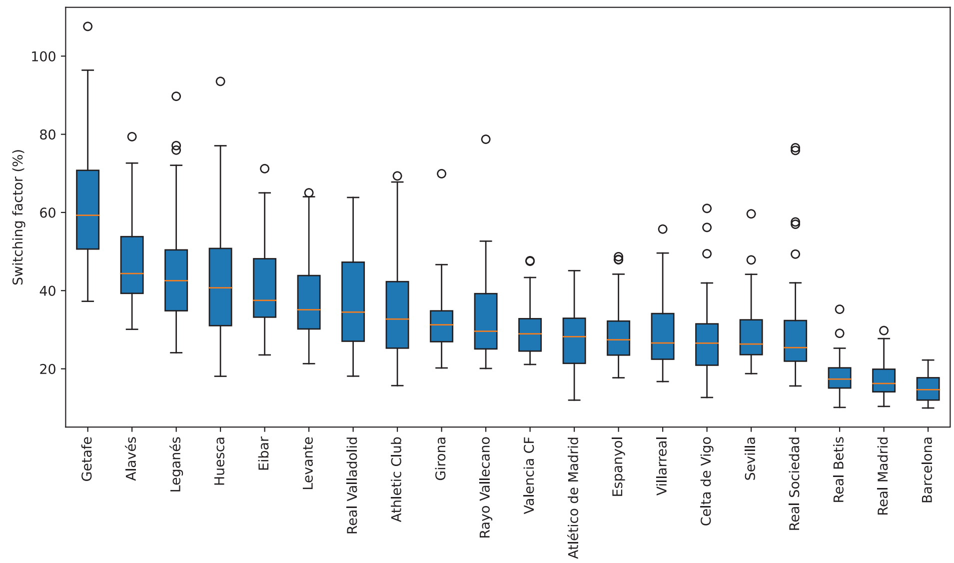

Fig. 9. Box plot of the switching factor of each team across all matches.

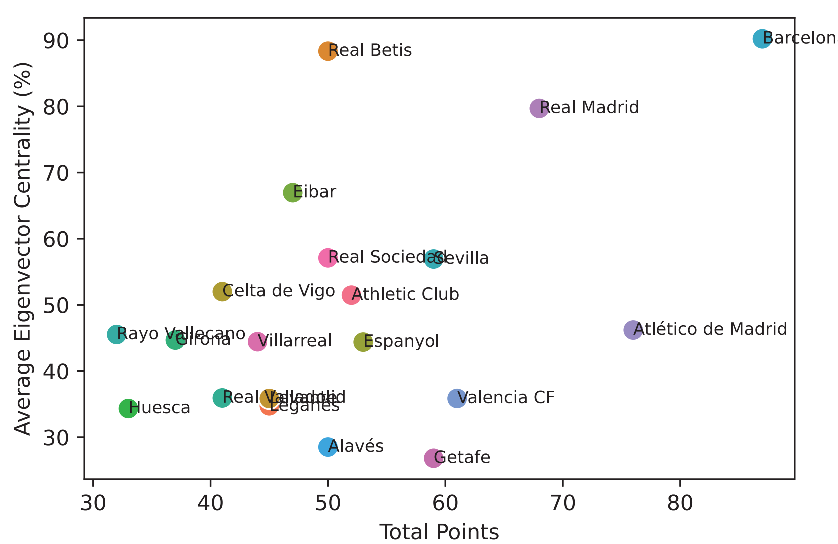

Fig. 10. Scatter plot of the average eigenvector centrality vs the total points accumulated at the end of the season.

of the penalty boxes exhibit the lowest importance. This indicates that these areas are closer to the opponent's defensive third, resulting in fewer opportunities for sustained passing sequences.

Examining the situation from a different vantage point, an analysis of the leakage indicator reveals that within the central areas of the pitch, all teams exhibit a propensity for high-quality and continuous passing organization, resulting in reduced instances of possession loss. On the flip side, as the ball approaches the corner kick areas, these regions manifest the highest leakage rates (Node 25, 5), albeit without a corresponding increase in possession recovery. This observation implies that, excluding the factor of returning the ball in the opposite direction in the corner kick area, securing precise passing in this area demonstrates a considerable challenge.

We have seen that, taking into consideration the team's penalty area regions, due to the typically dense presence of defenders and the numerical advantage in the penalty area, they tend to be a region with a higher success rate in regaining possession [29]. At the same time, for attacking players, the penalty area is a challenging place to engage in sustained passing actions because entering the penalty area often leads directly to a shooting opportunity [30,31].

The multilayer passing networks also capture the distinct playing styles and passing patterns of *LaLiga* teams. Real Madrid comprises a higher proportion of high-level players, many of whom exhibit exceptional individual ball control skills, particularly in terms of their capacity to create space through individual skill. Consequently, Real Madrid often commands a superior ball possession rate compared to their adversaries in matches. Moreover, an analysis of Real Madrid's multilayer networks reveals a conspicuous pattern of increased passing activities in the left midfield region, characterized by nodes 23, 24, and 18. Additionally, they demonstrate a proficient utilization of the lateral spaces, frequently exhibiting a proclivity for attacking down the left flank, thus establishing a discernible offensive advantage. On the other hand, FC Barcelona has consistently adhered to a possession style, particularly evident in their high passing linkages in the midfield and the opponent's half. These passing connections extend across most areas in the middle and attacking thirds of the pitch, with each region generally displaying a strong centrality. Barcelona boasts the highest percentage of eigenvector centrality in the league, concurrently achieving the highest position at the final ranking. This illustrates FC Barcelona's adeptness in controlling the speed and rhythm of the game effectively. Concerning the switching factor, FC Barcelona and Real

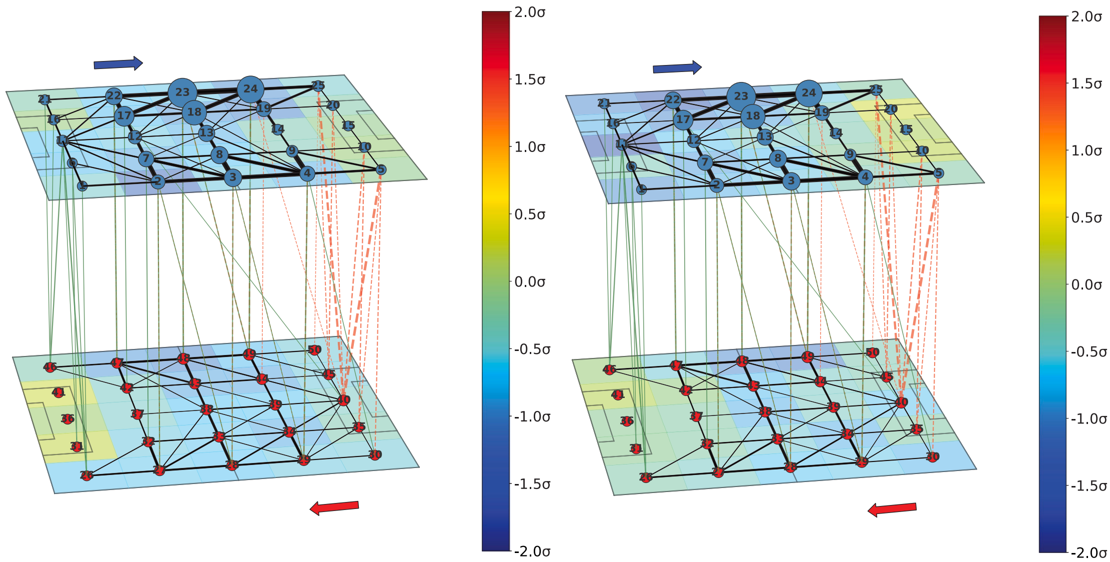

Fig. 11. Real Madrid multilayer network (throughout the season). The top-layer corresponds to Real Madrid's pitch passing network, while the bottom layer is the average of its rivals. Dashed links indicate ball losses between teams. Green links reflect the average ball transfers from the rival teams to Real Madrid, while red links represent the reverse direction. Node size is proportional to the eigenvector centrality, and edge width corresponds to link weight (number of ball transfers). The heatmap on the left is proportional to the z-score of the *leakage* parameter of each zone of the pitch. On the right, the heatmap represents the z-score of the *recovery* parameter. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Comparison of the average values of the eigenvector centrality, leakage parameter, recovery parameter and switching factor for the teams in LaLiga during the season 2018/2019. Teams are ordered by a decreasing raking of their eigenvector centralities.

Team	Eigenvector centrality (%)	Leakage (%)	Recovery (%)	Switching factor (%)
Barcelona	90.2%	16.3%	14.7%	14.5%
Real Betis	88.3%	19.5%	15.9%	17.4%
Real Madrid	79.7%	18.6%	16.3%	16.7%
Eibar	67.0%	33.8%	31.1%	39.7%
Real Sociedad	57.1%	25.4%	20.8%	27.6%
Sevilla	56.9%	26.1%	21.9%	27.9%
Celta de Vigo	52.0%	24.6%	21.4%	25.6%
Athletic Club	51.5%	28.9%	26.1%	32.5%
Atlético de Madrid	46.2%	25.6%	23.5%	26.4%
Rayo Vallecano	45.5%	27.1%	23.4%	31.3%
Girona	44.7%	28.9%	22.6%	30.9%
Espanyol	44.4%	24.8%	21.3%	28.0%
Villarreal	44.4%	25.4%	23.1%	27.8%
Real Valladolid	35.9%	30.0%	24.7%	33.8%
Valencia CF	35.9%	26.4%	24.0%	28.7%
Levante	35.9%	30.6%	26.0%	35.6%
Leganés	34.7%	32.9%	29.1%	40.8%
Huesca	34.3%	32.8%	28.4%	39.6%
Alavés	28.5%	34.8%	30.5%	45.6%
Getafe	26.8%	38.0%	37.8%	59.3%

Madrid register the lowest values during the league, signifying that both teams are known for their extended periods of ball possession and a higher frequency of passing actions.

Besides, some teams employ different tactical styles. For instance, Atlético de Madrid and Getafe leans toward a "direct play" approach with lower eigenvector centrality, aiming to advance quickly with fewer touches into the opponent's half or attacking third. Long passes are a prominent feature of this style [32]. However, such long passes or any rapid long-distance offensive progression often entail a lower pass success rate [13,33]. This lower ball possession rate can lead to more frequent transitions between attack and defense [34,35].

Additionally, it is interesting to note that Getafe simultaneously exhibits higher positive leakage and recovery parameters, which are evident across all regions of the pitch. This can be attributed to their playing style, which involves applying high pressure on the ball and employing a highly direct approach. As a result, the team excels in defensive actions during transition play, displaying both effectiveness and intensity in this aspect of the game.

From the perspective of intra-layer connections, this refers to the passing patterns between the different pitch areas during the attacking phase, which are primarily determined by a central aggregation of the collective passing actions [3,23]. The pattern of connectivity can aid in disclosing how the team coordinates and cooperates in different areas within the same layer. Elite soccer teams have different playing styles with small tactical groups specializing in different offensive purposes or game strategy. For instance, some teams tend to initiate their attacks from the flanks, while others excel in finding opportunities for central penetration in possession play. Different network spatial connectivity characteristics reflect the dominance and trends in a team's offensive tactics [36].

From the perspective of inter-layer connections, this focuses on the process of possession exchanges. The leakage, recovery and switching factors include the spatial features of the offensive actions transforming into a defensive action, as well as the spatial features of the defensive actions transforming into offensive actions. In the realm of modern football, the competition for dominance over temporal and spatial dimensions during the phases of transition has exhibited a marked intensification. The spatial distribution and frequency of transitions between teams reveal critical playing style of a game strategy [37,38].

In view of all, the multilayer network framework method, as a novel approach in sports performance analysis, represents a significant advancement building upon the foundation of passing networks [3]. It introduces spatial 2-layer networks to delve into the intricate dynamics of offensive and defensive transitions between competing teams. Moreover, it extends and innovates by proposing three additional metrics leakage, recovery, and switching factor aimed at quantifying the extent of possession transitions between layers. This approach offers a novel perspective and research dimensions for studying the complex, interactive, and dynamic collective behaviors in football matches. However, it is also worth noting that when applying the multilayer network framework to explain team's final position in the league ranking, one must carefully consider the consequences of the different teams' playing styles. It is not a matter of maximizing/minimizing any of the multilayer network parameters, such as the leakage, recovery of

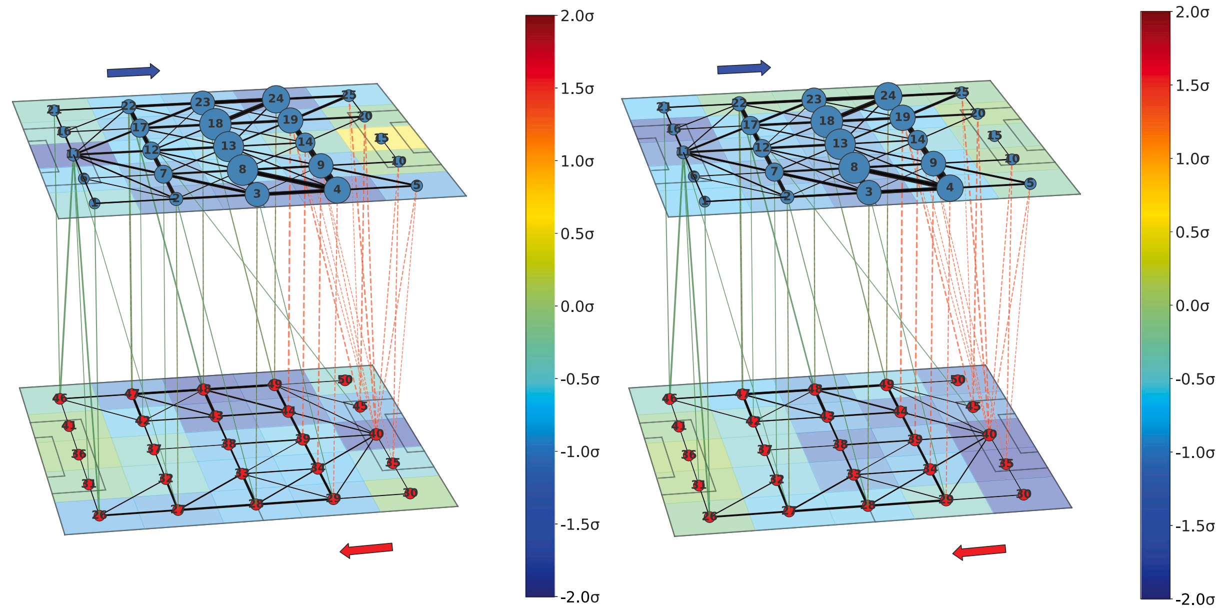

Fig. 12. FC Barcelona multilayer network (throughout the season). The top-layer corresponds to Barcelona's pitch passing network, while the bottom layer is the average of its rivals. Dashed links indicate ball losses between teams. Green links reflect the average ball transfers from the rival teams to Real Madrid, while red links represent the reverse direction. Node size is proportional to the eigenvector centrality, and edge width corresponds to link weight (number of ball transfers). The heatmap on the left is proportional to the z-score of the *leakage* parameter of each zone of the pitch. On the right, the heatmap represents the z-score of the *recovery* parameter. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

switching factor, but to understand how opponents perform in these parameters and adapt to have an marginal advantage to increase winning probability.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data contained in this paper may be accessible under request for appropriate studies.

Acknowledgments

Authors thank M.A. Gómez-Ruano and Jacobo Aguirre for fruitful discussions.

Funding

JMB is supported by Ministerio de Ciencia e Innovación (project PID2020-113737GB-I00). This research was conducted under the Sport Sciences Network (2022): 25/UPB/22 SPAA. Sports Performance Analysis Association.

Appendix

Z-score of the average networks of Real Madrid (Fig. 11) and F.C. Barcelona (Fig. 12) in terms of the leakage (left networks) and recovery (right networks) parameters.

References

[1] Newman M. Networks. Oxford University Press; 2018.
[2] Gudmundsson J, Horton M. Spatio-temporal analysis of team sports. ACM Comput Surv 2017;50(2):1–34.
[3] Buldú JM, Busquets J, Martínez JH, Herrera-Diestra JL, Echegoyen I, Galeano J, et al. Using network science to analyse football passing networks: Dynamics, space, time, and the multilayer nature of the game. Front Psychol 2018;9(1900).
[4] Ribeiro Silva, Davids Araújo, Ramos Lopes, et al. A multilevel hypernetworks approach to capture properties of team synergies at higher complexity levels. Eur J Sport Sci 2020;10(20):1318–28.
[5] Balagué Hristovski, Almarcha Garcia-Retortillo, Ivanov. Network physiology of exercise: Vision and perspectives. Front Physiol 2020;(11).
[6] Liu XF, Liu YL, Lu XH, Wang QX, Wang TX. The anatomy of the global football player transfer network: Club functionalities versus network properties. PLoS One 2016;11(6):e0156504.
[7] Clemente GP, Cornaro A. Community detection in attributed networks for global transfer market. Ann Oper Res 2023;325(1):57–83.
[8] Ma Wang, Wang Li, Fung Zhang, et al. The hotspots of sports science and the effects of knowledge network on scientific performance based on bibliometrics and social network analysis. Complexity 2021;(2021):1–12.
[9] Narizuka T, Yamamoto K, Yamazaki Y. Statistical properties of position-dependent ball-passing networks in football games. Physica A 2014;412:157–68.
[10] Kivelä M, Arenas A, Barthelemy M, Gleeson JP, Moreno Y, Porter MA. Multilayer networks. J Cplx Netw 2014;2:203–71.
[11] Boccaletti S, Bianconi G, Criado R, Del Genio CI, Gómez-Gardenes J, Romance M, et al. The structure and dynamics of multilayer networks. Phys Rep 2014;544:1–122.
[12] Aguirre J, Papo D, Buldú JM. Successful strategies for competing networks. Nat Phys 2013;9:230.
[13] Gong B, Cui Y, Zhang S, et al. Impact of technical and physical key performance indicators on ball possession in the Chinese Super League. Int J Perform Anal Sport 2021;21(6):909–21.
[14] Bradley PS, Lago-Peñas C, Rey E, et al. The effect of high and low percentage ball possession on physical and technical profiles in English FA premier league soccer matches. J Sports Sci 2013;31(12):1261–70.
[15] Gomez M-A, Reus M, Parmar N, et al. Exploring Elite Soccer Teams' performances during different match-status periods of close matches' comebacks. Chaos Solitons Fractals 2020;132(1095):66.
[16] K. Göral. Passing success percentages and ball possession rates of successful teams in 2014 FIFA World Cup. Int J Sport Cult 2015;3(1):86–95, 2015.
[17] Lorenzo-Martinez M, Kalén A, Rey E, et al. Do elite soccer players cover less distance when their team spent more time in possession of the ball? Sci Med Football 2021;5(4):310–6, (2021).

[18] Modric T, Versic S, Chmura P, et al. The effect of ball possession on physical and technical performance in UEFA Champions League matches. J Sports Sci 2022;40(24):2750–9.
[19] García-Calvo T, Ponce-Bordón J, Leo F, et al. How does ball possession affect the physical demands in Spanish LaLiga? A multilevel approach. Res Q Exercise Sport 2022;1–9.
[20] Bradley PS, Lago-Peñas C, Rey E, et al. The influence of situational variables on ball possession in the English Premier League. J Sports Sci 2014;32(20):1867–73.
[21] Marcelino R, Sampaio J, Amichay G, et al. Collective movement analysis reveals coordination tactics of team players in football matches. Chaos Solitons Fractals 2020;2020(1098):138: 31.
[22] T. McGarry. Sport competition as a dynamical self-organizing system: coupled oscillator dynamics of players and teams underscores game rhythm behaviours of different sports. In: Routledge handbook of sports performance analysis. Routledge; 2013, p. 52–63.
[23] Herrera-Diestra J, Echegoyen I, Martínez J, et al. Pitch networks reveal organizational and spatial patterns of Guardiola's FC Barcelona. Chaos Solitons Fractals 2020;138(1099):34.
[24] Mediacoach. Available online: https://www.mediacoach.es. [Accessed on 6 October 2023].
[25] Opta. Available online: http://www.optasports.com. [Accessed on 6 October 2023].
[26] Description of Opta events. 2023, Available online: https://www.statsperform.com/opta-event-definitions. [Accessed on 6 October 2023].
[27] Konefał M, Chmura P, Zając T, et al. A new approach to the analysis of pitch-positions in professional soccer. J Hum Kinetics 2019;66(143).
[28] Clemente FM, Martins FM, Couceiro MS, et al. Developing a tactical metric to estimate the defensive area of soccer teams: The defensive play area. Proc Inst Mech Eng P 2016;230(2):124–32.
[29] Ruiz-Ruiz C, Fradua L, Fernández-García Á, et al. Analysis of entries into the penalty area as a performance indicator in soccer. Eur J Sport Sci 2013;13(3):241–8.
[30] Herold M, Goes F, Nopp S, et al. Machine learning in men's professional football: Current applications and future directions for improving attacking play. Int J Sports Sci Coach 2019;14(6):798–817.
[31] Wang S, Y. Qin. Analysis of shooting and goal scoring patterns in the 2019 France Women's World Cup. J Phys Educ Sport 2020;20(6):3080–9, (2020).
[32] Yi Q, Gómez MA, Wang L, et al. Technical and physical match performance of teams in the 2018 FIFA World Cup: Effects of two different playing styles. J Sports Sci 2019;37(22):2569–77.
[33] Merlin M, Cunha SA, Moura FA, et al. Exploring the determinants of success in different clusters of ball possession sequences in soccer. Res Sports Med 2020;28(3):339–50.
[34] Gollan S, Ferrar K, K. Norton. Characterising game styles in the English Premier League using the moments of play framework. Int J Perform Anal Sport 2018;18(6):998–1009.
[35] Olthof SB, Frencken WG, Lemmink KA. When something is at stake: Differences in soccer performance in 11 vs. 11 during official matches and training games. J Strength Conditioning Res 2019;33(1):167.
[36] Gong BN, Zhou CJ, Gomez MA, et al. Identifiability of Chinese football teams: A complex networks approach. Chaos Solitons Fractals 2023;(166).
[37] Barreira D, Garganta J, Guimarães P, et al. Ball recovery patterns as a performance indicator in elite soccer. Proc Inst Mech Eng P 2014;228(1):61–72.
[38] Forcher L, Forcher L, Altmann S, et al. The keys of pressing to gain the ball–Characteristics of defensive pressure in elite soccer using tracking data. Sci Med Football 2022;1–9.