Chaos, Solitons and Fractals 138 (2020) 109934

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

Nonlinear Science, and Nonequilibrium and Complex Phenomena

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

Pitch networks reveal organizational and spatial patterns of Guardiola's F.C. Barcelona

J.L. Herrera-Diestra [a,b,*], I. Echegoyen [d,e,f], J.H. Martínez [c,e], D. Garrido [d,e,f], J. Busquets [g], F.Seirul. lo [h], J.M. Buldú [d,e,f,i]

[a] UNESP-IFT, ICTP-SAIFR, Sao Paulo, Brazil
[b] CeSiMo, Facultad de Ingeniería, Universidad de Los Andes, Mérida 5101, Venezuela
[c] Depto. Ingeniería Biomédica. Universidad de los Andes, Bogotá, Colombia
[d] Complex Systems Group, Universidad Rey Juan Carlos, Madrid, Spain
[e] Grupo Interdisciplinar de Sistemas Complejos (GISC), Spain
[f] Laboratory of Biological Networks, Center for Biomedical Technology, Universidad Politécnica de Madrid, Madrid, Spain
[g] E.S.A.D.E. Business School, Barcelona, Spain
[h] Departamento de Metodología, F.C. Barcelona, Barcelona, Spain
[i] Institute of Unmanned System and Center for OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an 710072, China

ARTICLE INFO

Article history:
Received 26 March 2020
Revised 14 May 2020
Accepted 22 May 2020

Keywords:
Football
Network science
Pitch networks
Team performance

ABSTRACT

We investigated the particular organization of Guardiola's F.C. Barcelona during season 2009/2010, using datasets from the Spanish National League *La Liga*. Specifically, we constructed the corresponding pitch networks, obtained from all passes successfully performed by a team during a football match. Pitch networks are composed of nodes consisting of particular subdivisions of the field, which are connected through links whose weight $\omega_{i,j}$ corresponds to the number of passes made from region i to region j. We performed a multi-scale analysis focused on evaluating the properties of pitch networks at different scales, from a partition of the pitch into 2×2 to 10×10 areas. For each scale, we calculated a diversity of network parameters of F.C. Barcelona and its opponents during the whole season. Next, we compared the properties of F.C. Barcelona pitch networks with the networks of its rivals. Our results show how, depending on the spatial scale, there are statistically significant differences between F.C. Barcelona and the rest of the teams of the Spanish league. These differences are particularly significant at the clustering coefficient, the network average shortest-path, and the number of nodes occupied by a team for partitions with a high number of subdivisions.

1. Introduction

Football is probably the sport that has benefited the most from the application of Network Science [1,2] to team sports, where the coordination between a group of players plays a crucial role. As long as there is a diversity of interacting systems and a way of determining a particular connection or interaction between them, it is possible to construct a network based on experimental observations and analyze its structure to understand the processes occurring in the network. Basketball [3], rugby [4], or baseball [5] are examples of how it is possible to translate an activity related to a particular sport into a network, obtaining a different point of view of specific problems, such as the understanding of team performance, player evaluation or outcome prediction. However, the way of constructing networks relies both on (i) the system itself and (ii) the availability of datasets. For example, in [3], the authors evaluated ball movements during a basketball match. In this case, the nodes of the network consisted of different game actions (inbounds, rebounds, shots, fouls, turnovers, ...). Next, the ball flow through all these actions allowed to create weighted-directed networks of actions occurring during the match. Finally, the role played by each action (node) was determined in terms of its importance in the whole network.

We can find many other ways of obtaining sports networks if we turn our attention to football. For example, in [6] authors investigated the structure of the transfer network, where nodes consisted of football clubs and connections between them were created when a player was transferred from one club to another. Using a dataset that contained close to 500.000 transfers along

* Corresponding author.
E-mail address: diestra@gmail.com (J.L. Herrera-Diestra).

https://doi.org/10.1016/j.chaos.2020.109934

26 years, it was possible to obtain a weighted-directed network, whose in-degree (number of incoming players), out-degree (number of outgoing players), and weight distribution were analyzed to detect the role of clubs in the transfer market. A complementary approach was followed by Maimone and Tasseri to assess the predictability of football using network science [7]. In this case, the authors analyzed the results of 11 major European leagues for 26 years. Again, football clubs where the nodes of the network, however, links consisted of arrows departing from the loser team of a match to the winner, disregarding ties. The computation of the eigenvector centrality [8], a network-based metric that quantifies the importance of a node in a network, showed that the inequality of teams has dramatically increased during the last years, increasing predictability of football results.

Other approaches introduced player-oriented networks, such as linking players who have worked at the same club at the same time, leading to a player co-working network [9].

However, the most extended way of transforming football datasets into a network is the construction of passing networks [10]. This approach, introduced in the late seventies by Gould and Gatrell [11], considers the players of a team to be linked through the passes they make during the match. In this way, it is possible to obtain weighted-directed networks, where nodes are players of a team, and links are arrows whose weight accounts for the number of passes between the players they are connecting. Passing networks have been used to quantify the role of players in a team [12,13], the organization of teams both at spatial and temporal scales [14] and, ultimately, team performance [15,16].

However, there is an alternative way of obtaining passing networks that puts the focus on the space instead of the player: *pitch passing networks*. In this kind of networks, nodes consist of different areas of the pitch that pass the ball between them. For example, in [17], authors evaluated the performance of teams at the Italian National League and the teams playing at FIFA World Cup 2014 by constructing pitch networks of 10×10 regions (i.e., 100 nodes). Based on the parameters of the resulting networks, authors obtained an indicator of team performance that correlated well with the outcomes of matches both at the Italian National League and the 2014 World Cup. Note that this kind of approach disregards the importance of a given player in a team, putting the attention to the collective behavior in terms of the spatial distribution of its passes.

There are also hybrid approaches where nodes are the combination of a player and its position on the pitch [13]. Using this methodology, a player can be present at different nodes of the network as long as he/she passes from different regions of the pitch. Narizuka et al. [18] followed this latter approach combining the 11 players of a team with 18 different pitch positions, leading to networks of 18×11 nodes. Furthermore, authors also considered a sliding window of 15 minutes to capture the changes in the organization of the pitch-player networks and how they were related to the phases of the game. They showed the existence of the small-world property [19], i.e., a network with a combination of a high clustering coefficient and a low average path between nodes. Finally, the authors designed a Markov chain model that was able to reproduce the statistical properties of the real player-pitch network, namely the clustering coefficient and the average shortest path.

However, the Achilles' heel of the partitions of the pitch considered in the literature is that they are quite subjective. The dimensions of the partitions are based on the previous experience of the researchers, which results on partitions of different sizes such as (3×3) 9 divisions in [13,20], 18 (6×3) divisions in [18] or 100 (10×10) in [17].

In this paper, we analyzed the organization of the pitch networks of F.C. Barcelona (FCB) when it was coached by Pep Guardiola, a football team considered as one of the most significant references in football history [21,22]. Pep's FCB was a team characterized by a particular way of playing where having the ball was the main priority. The advantage over the rival team was based on (i) the construction of a dense network of passes between players with the aim of generating disequilibria in the opponent and (ii) a high pressure after loosing the ball (in order to recover it as soon as possible) [23]. This historic Catalan team has been already analyzed under the scope of Network Science showing statistically significant differences in a series of network parameters. Buldú et al. [24], constructed player networks using the information (space and time) of all passes made by FCB during the 2009/2010 season of the National Spanish League (which we will refer to from now on as *La Liga)* and, next, analyzed the structure of these passing networks during a match, focusing on the fluctuations of their parameters and those of its rivals. Authors found differences in a series of network parameters that were related to a higher/lower probability of scoring/receiving a goal [24].

Here, we have gone one step beyond and analyzed the structure of the pitch-networks of FCB and its rivals during the 2009/2010 season of *La Liga*. The purpose was to identify what differences existed between Guardiola's team and the rest, trying to extract conclusions about the spatial organization of FCB passing networks and how they were related to Guardiola's way of understanding football. The crucial point about our analysis is that we did not focus on a single spatial scale of pitch-networks. On the contrary, we investigated how network parameters changed when the spatial scale at which pitch-networks are constructed was reduced, leading to subdivisions of the pitch with a smaller area and, as a consequence, increasing the number of nodes in the network. We observed how classical parameters such as the clustering coefficient or the shortest-path length [19] had statistical significant differences between FCB and the rest of the teams of *La Liga*, no matter what spatial scale was considered. However, other network parameters such as the largest eigenvalue of the adjacency matrix [1], the algebraic connectivity [25] or the percentage of occupied nodes drastically depended on the spatial scale, both for FCB and its rivals.

2. Methodology

Pitch networks

Datasets, provided by Opta [26], consisted of all passes completed during the 38 matches played by FCB during the season 2009/2010 of (*"La Liga"*). From each completed pass, we collected (i) the player who passed the ball, (ii) the player who received the ball, (iii) the position (x and y coordinates) of the sender/receiver players and (iv) the time at which the pass was made (see Table 1 for an illustrative example). It is worth noting that the dimensions of the different fields of *La Liga* are not the same. For this reason, we normalized all pitches to be bounded between the interval [0,100] both for the x and y coordinates. In this way, spatial units were called "field units" (f.u.) and, as a rule of thumb, one

Table 1
Example of the dataset structure. Time, in minutes and seconds, corresponds to the moment at which each pass was made. Player 1 and player 2 are, respectively, the sender and receiver of the pass, while $x_{1,2}$ and $y_{1,2}$ are the coordinates of both players, in field units (bounded, at both axis, between 0 and 100).

Time	Team	Player 1	x_1	y_1	Player 2	x_2	y_2
...	...	...	...	...	...	...	...
17:31	FCB	Piqué	13.30	51.30	Busquets	32.03	48.80
17:34	FCB	Busquets	35.00	50.10	Xavi	48.10	64.90
17:39	FCB	Xavi	51.15	60.20	Iniesta	66.40	55.50
...	...	...	...	...	...	...	...

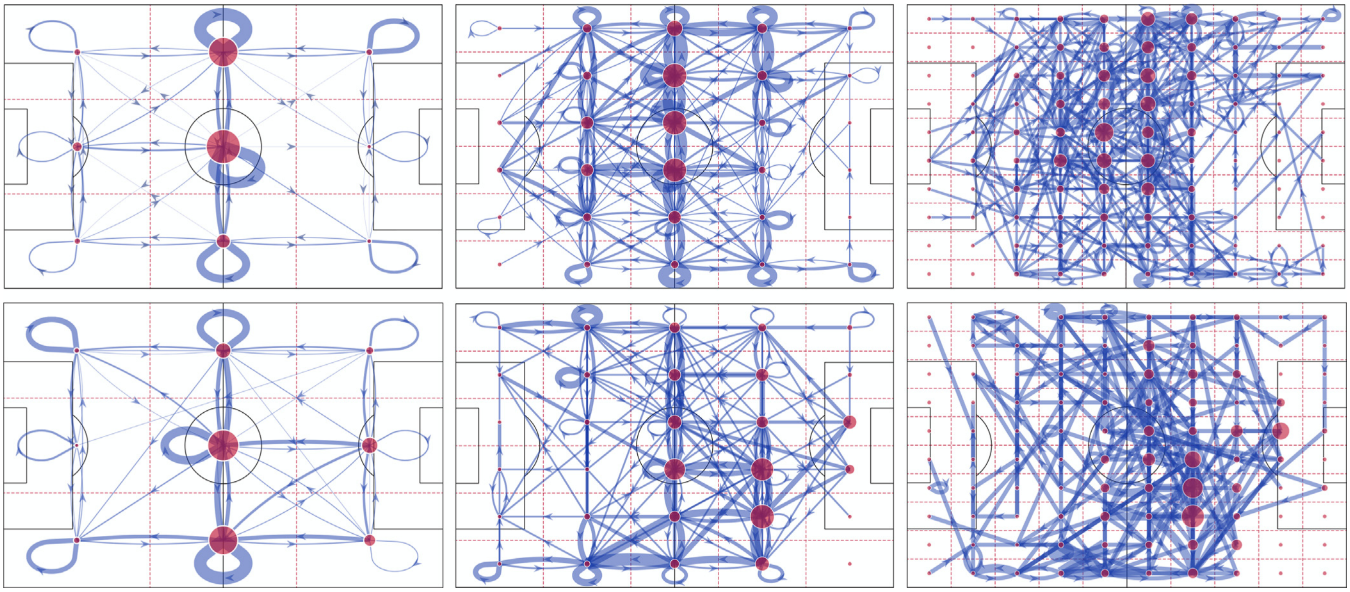

Fig. 1. Example of pitch networks at different scales for the match between FCB (home) and Real Madrid (away) during the season 2009/2010. Upper panel, plots of FCB's pitch networks for three different sizes: (left) $N = 9$ (3×3), (middle) $N = 30$ (5×6) and (right) $N = 100$ (10×10). Bottom panel correspond to the pitch networks of Real Madrid.

field unit in the x direction, i.e., the direction towards the opponent's goal, corresponded to approximately 1.05 meters, while one field unit in the y direction (parallel to the opponent's goal) was around 0.68 meters. The use of field units allows a more intuitive interpretation of the spatial coordinates where, for example, the position [50,50] is the center of the pitch and coordinates [100,50] correspond to the center of the opponent's goal.

Using these datasets, we constructed *pitch networks* by partitioning 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). Note that a network with $N = 6$ can have two different partitions: 3×2 and 2×3. A link from a node i to a node j is created when a pass is made from region i to j and has a weight that quantifies the total number of completed passes. In this way, we obtained weighted-directed networks with an adjacency matrix that was not symmetric. We built a total of $M = 81$ networks per match and team, resulting from the ordered combination of h and v, both variables taking values from 2 to 10. The smallest networks had a size of $N = 4$ (2×2), while the largest networks were of size $N = 100$ (10×10). As an example, Fig. 1 has pitch networks at three different scales (left: $3 \times 3 = 9$; middle: $5 \times 6 = 30$ and right: $10 \times 10 = 100$) for the match between FCB (home team) and Real Madrid (away team) played during the 2009/2010 season. Pitch partitions are highlighted by red dotted lines. Nodes were placed at the center of each pitch subdivision, and they are proportional to their importance in the passing network. Precisely, importance is measured using the eigenvector centrality [8], a measure of node importance that takes into account the number of passes received and, in turn, the importance of the regions from where these passes were made.

As we can see, splitting the pitch into divisions of different sizes leads to different networks. As a consequence, a series of natural questions arise: How are the properties of these pitch networks and how are they related? Are there any scales related between them? Furthermore, are there fundamental differences between teams depending on the specific scale? In what follows, we addressed these questions; however, let us first introduce the network parameters that we calculated and explain what information we can extract from each of them.

Definition of network parameters

Weighted adjacency matrix

The weighted adjacency matrix **W** contains the weights of the links that go from any node i to any node j. Its elements $w_{i,j}$ are proportional to the number of passes made from region i to region j. Specifically, we first obtained L, which is the total number of completed passes of a team during a match. Next, we counted all passes $(p_{i,j})$ made between any pair of regions i and j and finally we divided it by the total number of passes in order to obtain $w_{i,j} = p_{i,j}/L$. Using this normalization, if we define the network strength S as the sum of the elements of the adjacency matrix **W**, it will be 1 in all cases, no matter the number of passes made by a team. In this way, if we find differences between two teams in the network parameters, they could not be attributed to the number of passes, but to the differences in the organization of passes between the regions of the pitch.

Clustering coefficient

When links between networks do not have weights, the *local clustering coefficient* of a node i is commonly obtained as the percentage of the nodes directly connected to i that, in turn, are connected between them. In other words, it measures the probability of finding triangles around a given node. This measure can be averaged along the N nodes of the network to obtain the *average clustering coefficient* [1]. However, in our particular case, pitch networks are weighted and directed. For this reason, we used the weighted-directed version of the clustering coefficient of a node i, using the formula [27]:

$$C_w(i) = \frac{\frac{1}{2}[(\mathbf{W}+\mathbf{W}^T)(\mathbf{A}+\mathbf{A}^T)^2]_{ii}}{s_i^{tot}(d_i^{tot}-1)-2s_i^{\leftrightarrow}}, \qquad (1)$$

where $\mathbf{W}^T$ is the transpose of the weighted adjacency matrix **W; A** is the *unweighted* adjacency matrix, which is the binary version of **W** (i.e, $a_{i,j} = 1$ if $w_{i,j} > 0$ and $a_{i,j} = 0$ otherwise); s_i^{tot} is the total strength of node i ($s_i^{tot} = s_i^{in} + s_i^{out} = \sum_{j \neq i}(a_{ji}w_{ji} + a_{ij}w_{ij})$) and $s_i^{\leftrightarrow}$ accounts for the bidirectional links between the node i and its adjacent nodes ($s_i^{\leftrightarrow} = \sum_{j \neq i} a_{ij}a_{ji}(w_{ij} + w_{ji})/2$). Finally, the weighted clustering coefficient of the whole network is obtained by averaging $C_w(i)$ over all nodes, i.e., $C = \frac{1}{N}\sum_{i=1}^{N} C_w(i)$.

Shortest-path length

In pitch networks, the *shortest-path length SP* is the minimum number of nodes (areas of the pitch) that must be crossed by the ball to go from one sector of the pitch to any other. Since pitch networks are weighted (i.e., the number of passes involving a pitch division is different), we have to take into account the different weights of the links, considering that, the higher the number of passes between two areas (nodes), the higher the weight of the link and the shorter the "topological distance" between two nodes. Therefore, the topological length l_{ij} of the link between two pitch areas i and j is defined as the inverse of the link weight, $l_{ij} = 1/w_{ij}$. Besides, the shortest-path length between a pair of nodes may not be a direct link, since there could exist a shorter path by combining two (or more) alternative links. Therefore, we computed the minimal shortest-path p_{ij} between all pairs of nodes using the Dijkstra's algorithm [28]. Next, we defined the average shortest-path SP of each pitch network as:

$$SP = \frac{1}{N(N-1)} \sum_{i,j\ i \neq j} p_{ij} \tag{2}$$

where N is the total number of nodes of the pitch network. Note that, the lower the value of SP, the better connected pitch areas are, in terms of the number of steps required by the ball to reach any given area.

Largest eigenvalue of the adjacency matrix

The *largest eigenvalue* λ_1 of the weighted adjacency matrix $\mathbf{W}$ of a pitch network is a measure of its strength [29] in terms of robustness. The largest eigenvalue of $\mathbf{W}$ is bounded by the average strength of pitch areas $\langle s \rangle$, as $\lambda_1 \geq \langle s \rangle$, in the way $s_{\max} \geq \lambda_1 \geq \max(\langle s \rangle, \sqrt{s_{\max}})$ [30], where $s_{\max}$ is the highest strength of all areas (i.e., $max(s(i) = \sum_{j=1}^{N} w_{j,i})$). As a rule of thumb, networks with higher number of non-zero links (passes) have a higher λ_1 and networks with the important nodes connected between them also have a higher λ_1 than networks where the hubs (i.e., highly transited areas) are not directly connected between them.

Algebraic connectivity

The *algebraic connectivity* $\tilde{\lambda}_2$ corresponds to the second smallest eigenvalue of the Laplacian matrix $\tilde{L}$, which is defined as $\tilde{L} = \mathbf{S} - \mathbf{W}$, with $\mathbf{W}$ being the weighted adjacency matrix and $\mathbf{S}$ is a diagonal matrix whose i-element is the sum of the outgoing weights of area i. Algebraic connectivity is closely related to both structural and dynamical properties of networks [1,30,31]. On the one hand, the algebraic connectivity is an indicator of the modular structure of a network [32]: The lower the $\tilde{\lambda}_2$, the more independent groups inside the network, with the limit value of $\tilde{\lambda}_2 = 0$ indicating the existence of, at least, two completely disconnected groups. On the other hand, $1/\tilde{\lambda}_2$ is proportional to the time required to reach equilibrium in a linear diffusion process [33]. Additionally, the time t_{sync} to reach synchronization of an ensemble of phase oscillators that are linearly and diffusively coupled is also proportional to $1/\tilde{\lambda}_2$ [34].

Pitch coverage

The *coverage* parameter measures the percentage of areas of the pitch that are actually used by a team making passes. In this context, a node is "used" when it has at least one link with $w_{i,j} > 0$ and the coverage parameter is computed as the percentage:

$$coverage = 100 \times \frac{\text{number of nodes with at least one link}}{\text{number of possible nodes for a given scale}} \tag{3}$$

Note that, when all areas have a pass inside them, the coverage is 100%. However, when the pitch is divided into smaller areas, the probability of finding an empty region (i.e., a region without any pass) increases, leading to a decrease of the coverage parameter. Also note that the higher the number of passes made by a team, the higher the probability of covering more areas.

Pitch occupation

Since the coverage is highly dependent on the total number of passes, we can normalize its value in order to quantify the efforts of a team to distribute a limited number of passes through more regions of the pitch. With this aim, we defined the occupation parameter of each area as:

$$o^i_{in} = \frac{p^i_{in}}{L} - \frac{1}{N}, \tag{4}$$

where i is each of the subdivisions of the pitch (i.e., each node of the network), p^i_{in} is the number of incoming passes of a region i, L is the total number of passes made by a team, and N is the number of divisions of the pitch. Note that $\frac{p^i_{in}}{L}$ accounts for the probability of finding a pass entering a region i computed from the actual distribution of passes, while $\frac{1}{N}$ is the probability that a pass is entering to region i at random. In this way, o^i_{in} is indicating if the actual occupation of a region i is higher ($o^i_{in} > 0$) or lower ($o^i_{in} < 0$) than what would be expected if passes were randomly distributed.

Finally, the *occupation parameter* of the pitch is calculated as $O_{in} = \sqrt{\sum_{i=1}^{N} \left(o^i_{in}\right)^2}$. On one hand, when the pitch is used homogeneously by a team, i.e. all divisions receive the same number of passes, we obtain $O_{in} \sim 0$; on the other hand, when the use of the pitch is highly heterogeneous, O_{in} increases until reaching the extreme value of $O_{in} \sim 1$ for a large enough N.

Statistical analysis

To compare each network metric between groups (as seen in Figs. 2–7), we computed a non parametric test (Wilcoxon ran-ksum test) at each partition of the field. Each partition defined a network from which we obtained the parameter under study for each match, leading to 38 values per group. Assuming data is normally distributed is too risky with such small sample sizes, and it is safer turning to non-parametric approaches. The Wilcoxon ranksum test does not assume any underlying distribution, and compares medians instead of means, yielding a probability associated to the difference. Given the number of comparisons, we corrected all p-values following the common false discovery rate procedure developed by Benjamin and Yekutieli ([35]), setting $\alpha = 0.01$. Following this procedure ensured that all statistical comparisons met the most strict criteria, at the cost of not detecting subtler differences between groups. As we have shown in the Results section, this methodology does not prevent us from finding differences in various parameters for different partition schemes.

3. Results

We computed a series of network and pitch attributes in order to characterize the differences between FCB and its rivals for the 38 matches played during the 2009/2010 season of the Spanish national league. In Fig. 2 we plot the weighted clustering coefficient C of the pitch networks as a function of the different partitions considered for FCB (blue) and all its rivals (red). Labels of the horizontal axis indicate the type of partition, with the total number of nodes of the partition increasing to the right. Results are plotted using a box and whiskers representation that contains the median and the interquartile range IQR (i.e., the box), whose borders are the (lowest) 25th percentile, and the (highest) the 75th percentile. The whiskers are calculated as the first percentile minus 1.5 times the IQR (low) and the third percentile plus 1.5 times the IQR. Outliers are those values that fall outside these limits.

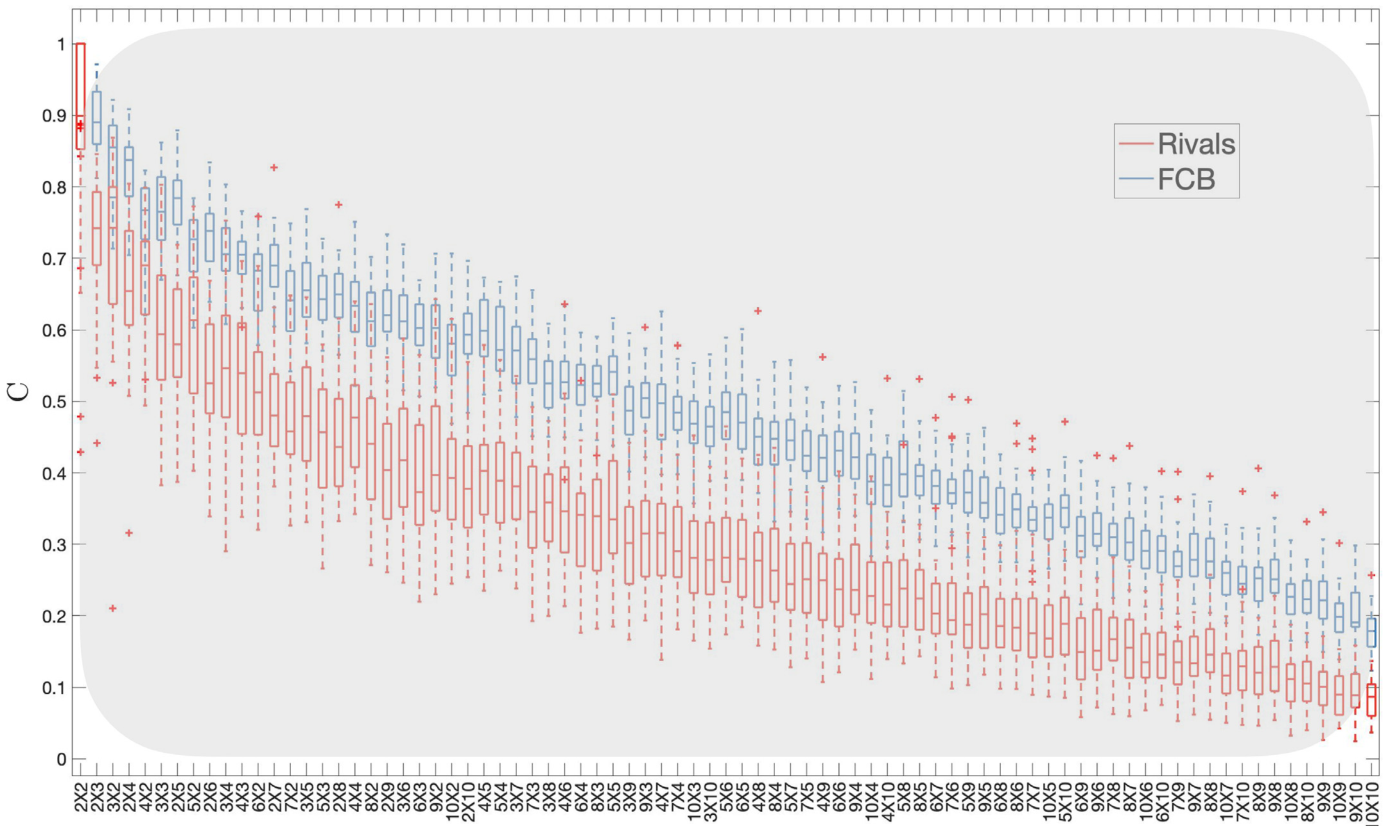

Fig. 2. Clustering coefficient (C) of pitch networks vs the size of the partition for FCB (blue) and its rivals during the season 2009/2010 (red). Measures highlighted with a grey background indicate that differences between FCB and its rivals are statistically significant. In this particular case, all scales (except the first pitch division) show statistically significant differences. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Finally, those values having statistically significant differences are highlighted with a grey shadowed background.

We can observe how, for all considered partitions, except the first-one (2×2), the clustering coefficient shows significant statistical differences between FCB and its rivals. Interestingly, FCB has a larger C in all cases, indicating that the number of triangles of the pitch passing networks is higher. Triangles are indicators of loops between three regions of the pitch and are related to a circulation of the ball between them.

In Fig. 3, we show the corresponding average shortest-path length *SP* of the aforementioned pitch networks. As in Fig. 2 there are significant statistical differences between FCB and its rivals at the different scales.

Interestingly, while the rivals increase their *SP* with the partition, FCB maintains its value for all partitions. This behavior of *SP* shows how tightly connected FCB remains, regardless of the scale of partition, i.e., the ball arrives from any region of the field to any other in a lower number of “topological” steps.

However, not all network metrics showed clear differences. In Fig. 4, we plot λ_1 as a function of the partition size. We can observe how, no matter the team, the value of λ_1 decreases monotonically as the number of nodes increases. This fact is a consequence of increasing the number of nodes (regions) while maintaining the number of links (passes). For a given number of passes, the lower the number of regions, the more connected they will be. Interestingly, we can observe how statistically significant differences only appear for certain partitions, which mainly have intermediate sizes. In these cases, it is FCB who has the higher value of λ_1, indicating that its passing networks are more robust than those of its rivals, since λ_1 is an indicator of the robustness of a network in terms of keeping its connectedness when links are removed. The algebraic connectivity $\tilde{\lambda}_2$ has a behaviour similar to λ_1. As shown in Fig. 5, only certain partitions show statistically significant differences. Furthermore, for pitch divisions implying a high number of nodes, the value $\tilde{\lambda}_2$ goes to zero due to a fragmentation of the networks (since $\tilde{\lambda}_2$=0). Again, in those cases where there are statistically significant differences, the value of $\tilde{\lambda}_2$ is higher in FCB, indicating that its passing networks are less prone to be splited into two disjoint sets than the networks of its rivals.

Now, let us take a closer look at the distribution of passes along the pitch. In Fig. 6 we show the *coverage* of the pitch networks as they are split into a larger number of areas. For small partitions, both FCB and its rivals cover the whole pitch; i.e., all nodes of the corresponding networks have at least one link (i.e., pass). As a consequence, there are no statistical differences between them, since the pitch coverage only measures the percentage of connected nodes in the network. However, as the number of partitions increases, the FCB’s rivals begin to have some areas of the pitch disconnected from the rest of the network, while FCB is still able to cover the whole pitch. Finally, for a sufficiently large number of partitions of the pitch, FCB fails in completing a pass from all areas of the pitch, and the coverage decreases below 100%. However, it is always higher than the average coverage of its rivals.

The higher values of the coverage reported in FCB may be simply explained by the higher amount of passes made by FCB during the season (see [24] for details). Therefore, we tried to quantify whether the coverage of the field is just related to the “quantity” of passes or it also includes their “quality”. With this aim, we calculated the occupation parameter O_{in}, which, in a few words, consists of a normalized version of the coverage, where the impact of the number of passes has been neutralized. The closer to zero O_{in} is, the more homogeneous the occupation of the pitch. Fig. 7 shows

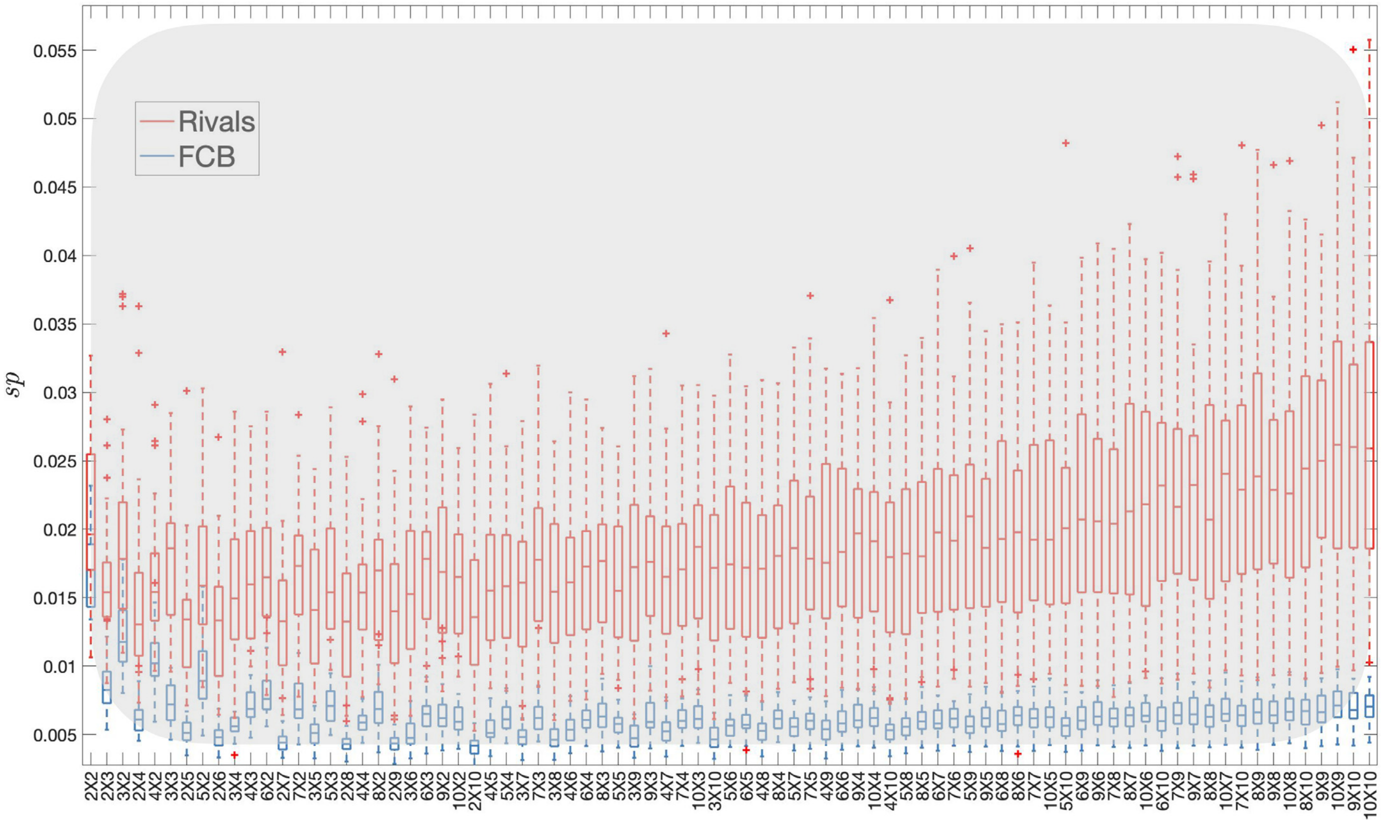

Fig. 3. Average shortest path length (*SP*) of pitch networks vs the size of the partition for FCB (blue) and its rivals during the season 2009/2010 (red). Measures highlighted with a grey background indicate that differences between FCB and its rivals are statistically significant. As in the case of the clustering coefficient, all scales (except the first pitch division) show statistically significant differences. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

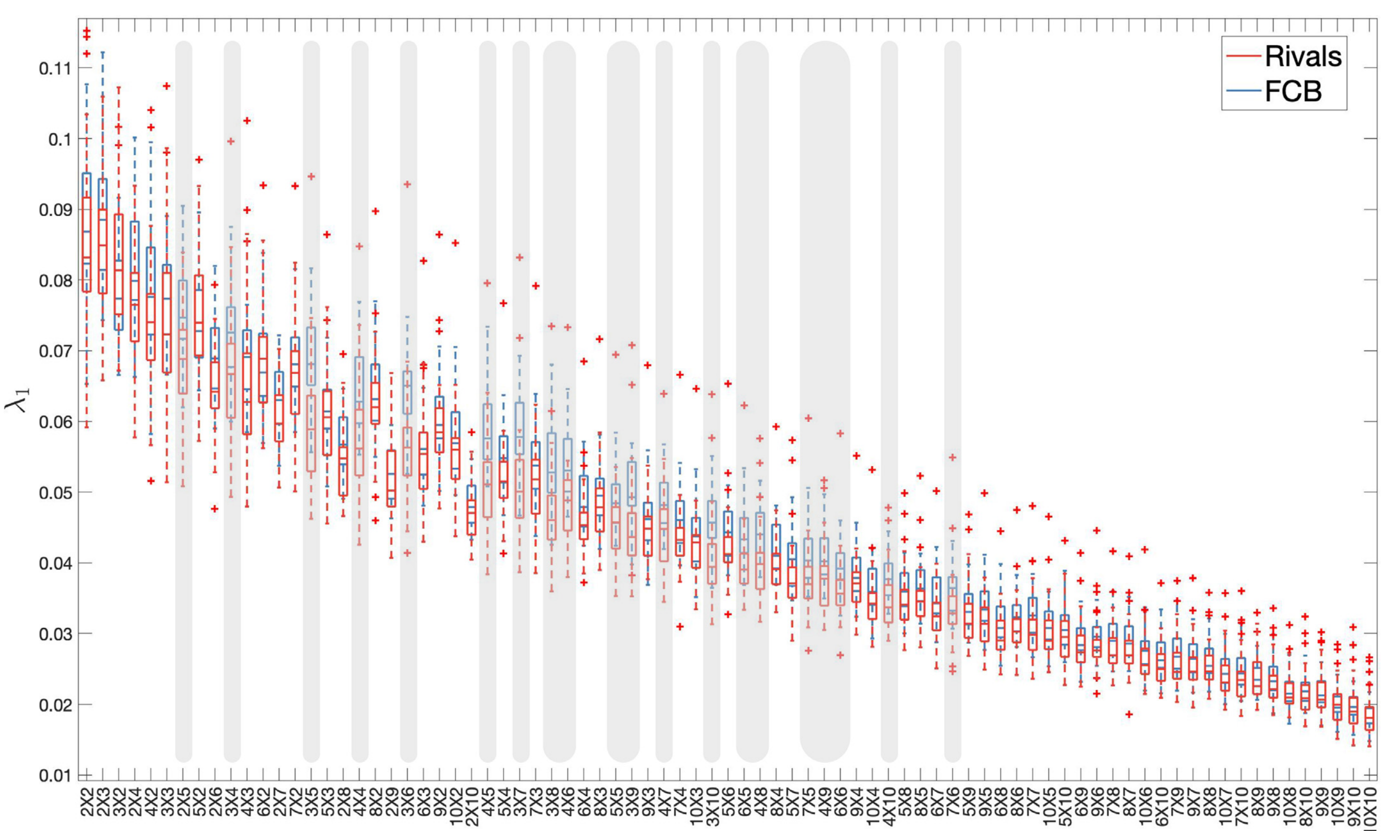

Fig. 4. Largest eigenvalue (λ_1) of the adjacency matrix of pitch networks vs the size of the partition for FCB (blue) and its rivals during the season 2009/2010 (red). Measures highlighted with a grey background indicate that differences between FCB and its rivals are statistically significant. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

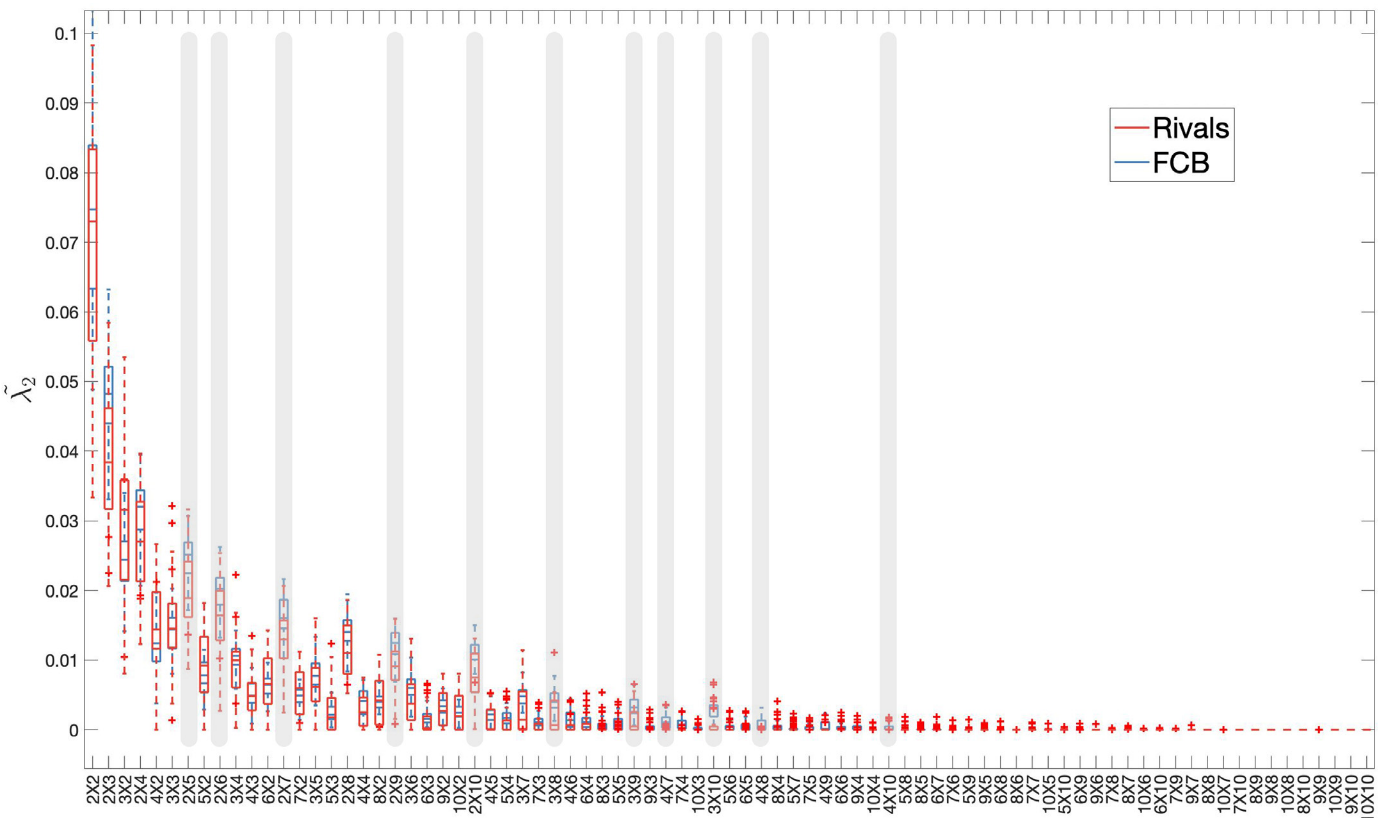

Fig. 5. Algebraic connectivity ($\tilde{\lambda}_2$) of pitch networks vs the size of the partition for FCB (blue) and its rivals during the season 2009/2010 (red). Measures highlighted with a grey background indicate that differences between FCB and its rivals are statistically significant. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

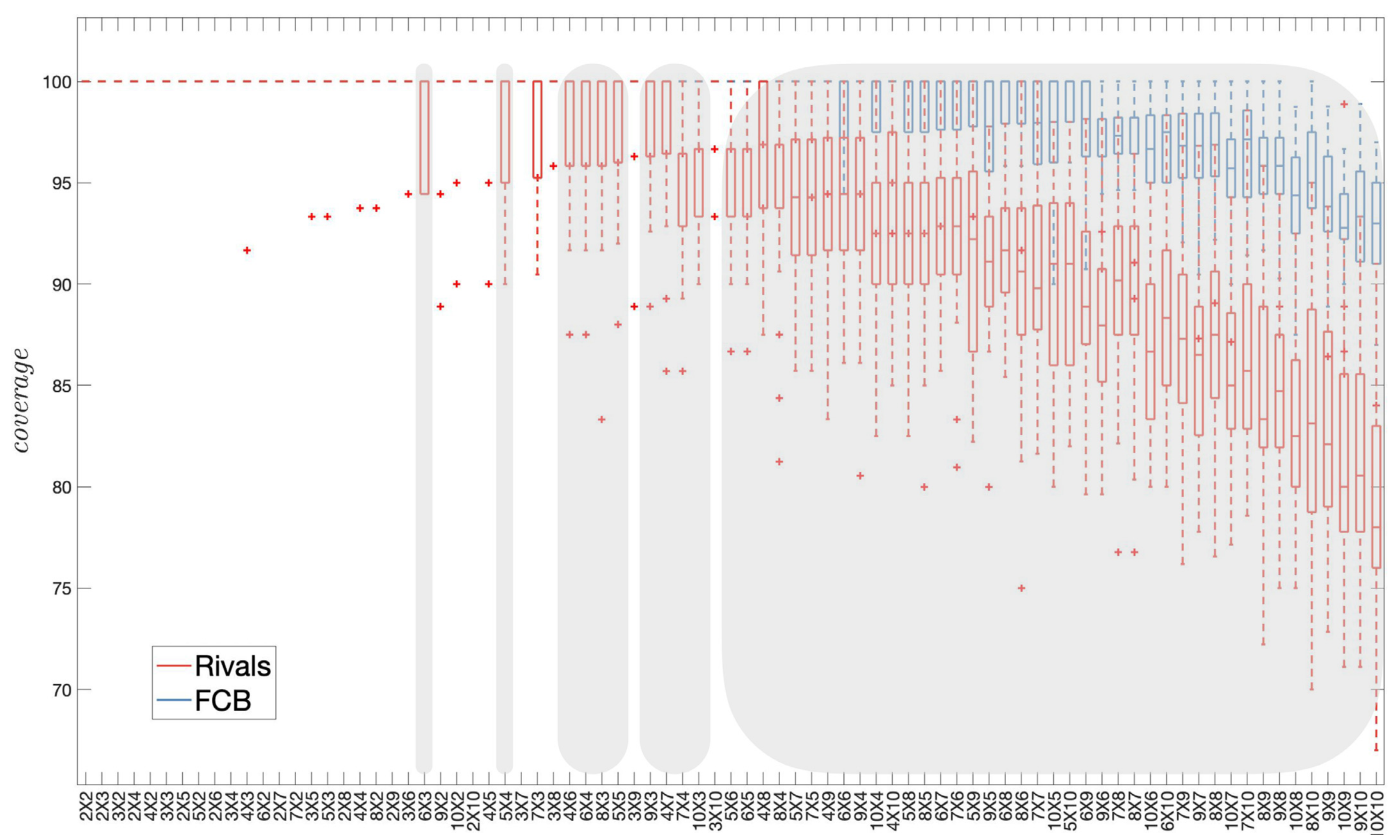

Fig. 6. Coverage of pitch networks vs the size of the partition for FCB (blue) and its rivals during the season 2009/2010 (red). Measures highlighted with a grey background indicate that differences between FCB and its rivals are statistically significant. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

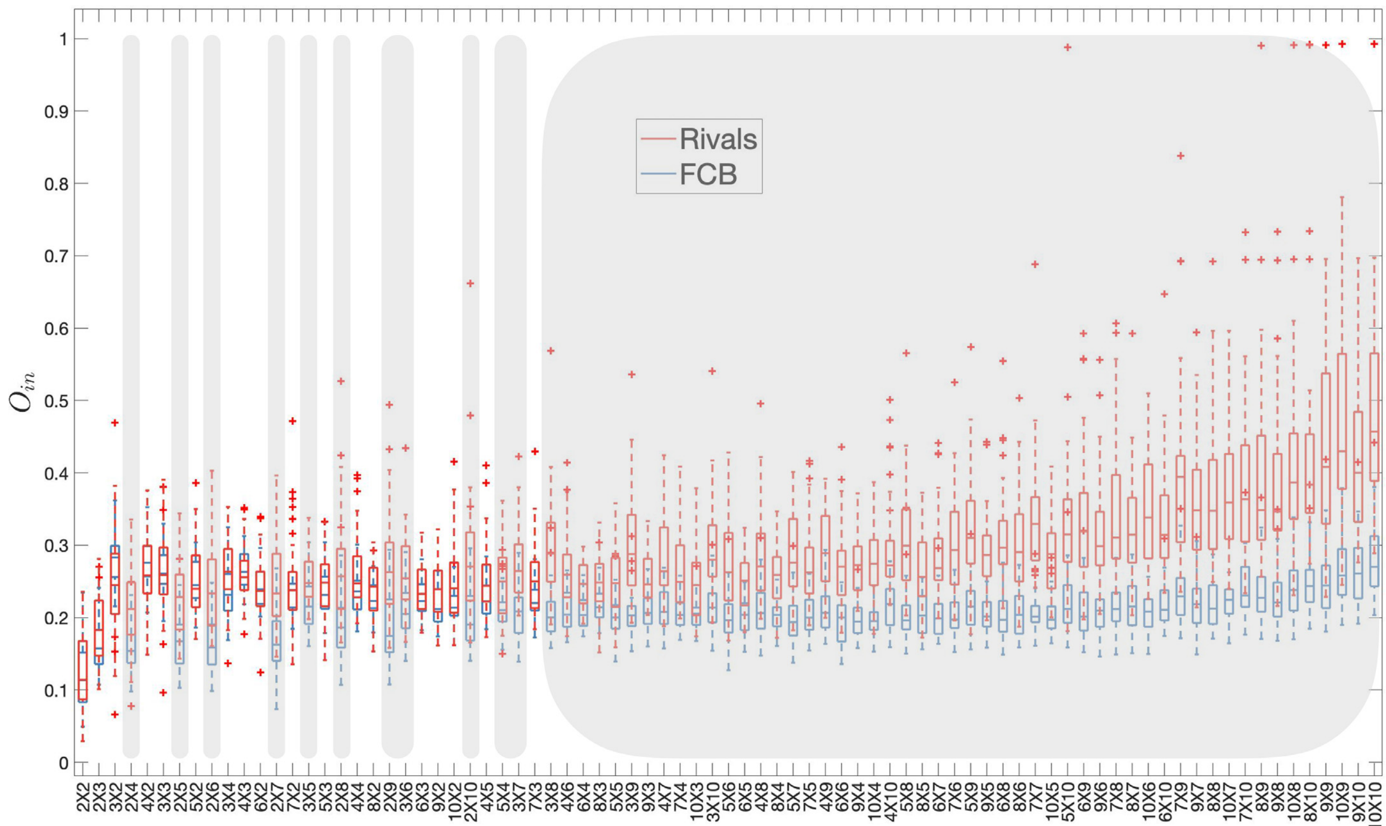

Fig. 7. Occupation parameter $\langle O_{in} \rangle$ of pitch networks vs the size of the partition for FCB (blue) and its rivals during the season 2009/2010 (red). Measures highlighted with a grey background indicate that differences between FCB and its rivals are statistically significant. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

how for partitions with intermediate to high number of divisions, O_{in} is smaller for FCB, indicating that it occupies the pitch more homogeneously than its rivals. In turn, the high values of O_{in} of the rivals show that they are prone to confine themselves into given zones of the pitch. Since O_{in} does not take into account the total number of passes, we can conclude that the style of playing developed by FCB leads to occupy more space in the pitch.

4. Conclusions

Football datasets can be translated into a diversity of different networks whose nodes can be, for example, teams [7], players [24], areas of the pitch [17] or actions carried out during a match [36], while links between nodes can be inferred from the result of a match [7], players exchanged by teams [6], the number of passes between players [12] or regions of the pitch [17]. Here, we made use of Network Science to analyze the structure of F.C. Barcelona (FCB) during the season 2009/2010, when Pep Guardiola coached it. From the diversity of approaches, we decided to focus on pitch passing networks, paying special attention to the importance of the spatial scale when dividing the pitch into subdivisions. Using this methodology, we reported statistically significant differences between the network parameters of Guardiola's team and the rest of its rivals in the Spanish national league. Two of the most studied network parameters, the clustering coefficient and the average shortest path length, showed that FCB was better organized than its rivals. The reason is twofold; on one hand, FCB created a higher number of triangles between areas of the pitch, which is related to the high robustness of networks at local scales. The reason is that, in a triangle, when one of the possible connections is cut, there is always an alternative path to reach any of the three nodes belonging to the triangle. Furthermore, networks with a high clustering coefficient have been demonstrated to enhance the transmission of information at global scales [37], a fact that, translated to pitch passing networks, would indicate that the ball better reaches all areas of the pitch. On the other hand, the lower values of the shortest path length indicate that the number of subdivisions that the ball crossed to go from an area of the pitch to any other was lower for FCB when compared to its rivals. This fact is good news for Guardiola's team since it demonstrates that the team, as a whole, was better connecting the field through the network of passes.

We also analyzed the spatial distribution of passes at different scales. Two parameters, the coverage and the occupation of the field, showed that FCB better used space than its rivals. Importantly, FCB made more passes from different locations or, in other words, it played over a larger field than its rivals. Both parameters indicate that FCB made more passes from different locations or, in other words, it played over a larger field than its rivals. However, this merit is not just a matter of making more passes since the occupation parameter is normalized to be independent of the total number of completed passes. Furthermore, the larger the number of subdivisions of the field, the higher the difference between FCB and its rivals.

It is worth noting that Guardiola's FCB had been previously analyzed using network science. In [24], it was shown that player-passing networks of FCB had much better indicators from those of its rivals in terms of robustness and connectivity between players. However, the use of pitch-passing networks puts the role of space at the forefront. In this way, we were able to (i) relate the differences between network parameters with the spatial scales and

(ii) we could detect the higher occupation of the pitch made by FCB. Both observations would not be made using player-passing networks.

Finally, two of the network parameters, the largest eigenvalue of the adjacency matrix and the algebraic connectivity, did not show as many differences between FCB and its rivals as the rest of the parameters. Only some particular divisions of the pitch showed statistically significant differences. However, when these differences existed, they always were in favor of FCB, showing a more united and robust network of passes in the Catalan team. Given all, we believe that the use of pitch passing networks can be a useful way of analyzing team organization in football and team sports in general. For example, in sports like basketball or hockey, where passing networks could be constructed, our methodology could be applied directly as long as the position of all passes had been recorded. If it is so, it would be possible to identify the scales and areas of the field where teams are more different from each other and to use this information to identify the particular spatial features of teams.

Author statement

All authors contributes equally in all stages of the research.

Declaration of Competing Interest

The authors declare that they do not have any financial or non-financial conflict of interests.

Acknowledgements

JHM thanks Colciencias Call N. 811 and Dept. Biomedical Sciences Universidad de los Andes for the support. JMB is funded by MINECO, Spain (FIS2017-84151-P). JLHD was supported by the São Paulo Research Foundation (FAPESP) under Grants no. 2016/01343-7 and no. 2017/05770-0.

References

[1] Newman MEJ. Networks: an introduction. New York: Oxford University Press; 2010.
[2] Barabási AL. Network science. Cambridge university press; 2016.
[3] Fewell JH, Armbruster D, Ingraham J, Petersen A, Waters JS. Basketball teams as strategic networks. PLoS ONE 2012;7(11):e47445.
[4] Cintia P, Coscia M, Pappalardo L. The haka network: evaluating rugby team performance with dynamic graph analysis. In: In Proceedings of the 2016 IEEE/ACM international conference on advances in social networks analysis and mining. IEEE Press.; 2016. p. 1095–102.
[5] Saavedra S, Powers S, McCotter T, Porter MA, Mucha PJ. Mutually-antagonistic interactions in baseball networks. Physica A 2010;389(5).
[6] Li MX, Zhou WX, Stanley HE. Network analysis of the worldwide footballer transfer market. EPL (Europhys Lett) 2019;125(1):18005.
[7] Maimone V.M., Yasseri T.. Football is becoming boring; network analysis of 88 thousands matches in 11 major leagues. 2019. ArXiv:1908.08991. unknown->arXiv preprint;
[8] Bonacich P. Factoring and weighting approaches to status scores and clique identification. JMathSociol 1972;2(1):113–20.
[9] Onody RN, de Castro PA. Complex network study of Brazilian soccer players. Phys Rev E 2004;70(3):37103.
[10] 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.
[11] Gould P, Gatrell A. A structural analysis of a game: the liverpool vs manchester united cup final of 1977. Soc Netw 1979;2:253–73.
[12] Duch J, Waitzman JS. Amaral LAN. quantifying the performance of individual players in a team activity. PLoS ONE 2010(5):e10937.
[13] Cotta C, Mora AM, Merelo JJ, Merelo-Molina C. A network analysis of the 2010 FIFA world cup champion team play. J Syst Sci Complex 2013;26:21.
[14] Martínez JH, Garrido D, Herrera-Diestra JL, Busquets J, Sevilla-Escoboza R, Buldú JM. Spatial and temporal entropies in the spanish football league: a network science perspective. Entropy 2020;22(2):172.
[15] Bialkowski A, Lucey P, Peter G, Carr K, Yue Y, Matthews I. Win at home and draw away: automatic formation analysis highlighting the differences in home and away team behaviors. In: Proceedings of the 8th Annual MIT Sloan Sports Analytics Conference (Boston, MA: MIT); 2014. p. 1–7.
[16] Duarte R, Araújo D, Folgado H, Esteves P, Marques P, Davids K. Capturing complex, non-linear team behaviours during competitive football performance. J Syst Sci Complex 2012;26:62–72. doi:10.1007/s11424-013-2290-3.
[17] Cintia P, Rinzivillo S, Pappalardo L. A network-based approach to evaluate the performance of football teams. In Machine Learning and Data Mining for Sports Analytics Workshop, Porto, Portugal; 2015.
[18] Narizuka T, Yamamoto K. Physica A 2014;412:157–68.
[19] Watts DJ, Strogatz SH. Collective dynamics of small-world networks. Nature 1998;393:440–2.
[20] Diquigiovanni J, Scarpa B. Analysis of association football playing styles: an innovative method to cluster networks. Stat Model 2019;19(1):28–54.
[21] Violan MA. Pep guardiola: the philosophy that changed the game. Meyer & Meyer Verlag; 2014.
[22] Perarnau M. Pep guardiola: the evolution. Birlinn Ltd.; 2016.
[23] Balague G. Pep guardiola: another way of winning: the biography. Orion 2012.
[24] Buldú JM, Busquets J, Echegoyen I, Seirullo F. Defining a historic football team: using network science to analyze guardiola's FC Barcelona. Sci Rep 2019;9(1):1–14.
[25] Fiedler M. Algebraic connectivity of graphs. Czechoslovak Math J 1973;23:298–305.
[26] https://www.optasports.com/sports/football.
[27] Clemente GP, Grassi R. Directed clustering in weighted networks: a new perspective. Chaos Solitons Fractals 2018;107:26–38.
[28] Dijkstra EW. A note on two problems in connection with graphs. Numer Math 1959;1:269–71.
[29] Aguirre J, Papo D, Buldú JM. Successful strategies for competing networks. Nat Phys 2013;9.
[30] Van Mieghem P. Graph spectra for complex networks. Cambridge, UK: Cambridge University Press; 2011.
[31] Masuda N, Porter MA, Lambiotte R. Random walks and diffusion on networks. Phys Rep 2017;716–717:1–58.
[32] Fortunato S, Hric D. Community detection in networks: a user guide. Phys Rep 2016;659:1–44.
[33] Gómez S, Díaz-Guilera A, Gómez-Gardeñes J, Pérez-Vicente CJ, Moreno Y, Arenas A. Diffusion dynamics on multiplex networks. Phys Rev Lett 2013;110:028701.
[34] Almendral JA, Díaz-Guilera A. Dynamical and spectral properties of complex networks. New J Phys 2007;9:187.
[35] Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. Ann Stat 2001;29:1165–88.
[36] Ramos J, Lopes RJ, Marques P, Araújo D. Hypernetworks reveal compound variables that capture cooperative and competitive interactions in a soccer match. Front Psychol 2017;8:1379.
[37] Centola D. The spread of behavior in an online social network experiment. Science 2010;329(5996):1194–7.