ANALYSIS OF SCORED AND CONCEDED GOALS BY A FOOTBALL TEAM THROUGHOUT A SEASON: A NETWORK ANALYSIS

Filipe Manuel Clemente[1,2], Fernando Manuel Lourenço Martins[2,3], and Rui Sousa Mendes[3]

[1]*Instituto Politécnico de Viana do Castelo, Escola Superior de Desporto e Lazer, Melgaço, Portugal*
[2]*Instituto de Telecomunicações (Institute of Telecommunications), Delegação da Covilhã, Portugal*
[3]*Polytechnic Institute of Coimbra, ESEC, Department of Education, RoboCorp, ASSERT, Portugal*

Original scientific paper
UDC: 796.332:796.012.574.45

Abstract:

Network analysis can provide a new set of important information about players and teams' behaviour throughout a match. Despite their pertinence, football analysis using network metrics is limited. Therefore, the aim of this study was to analyse the goals scored and goals conceded by a single team during a full season using network methods. Thirty-six games were analysed and from these the data about players' connectivity in actions preceding 46 goals scored and 15 goals conceded were collected. The results showed that the most used players during the plays that resulted in the goals scored were forward players in the forward regions, mainly in the penalty area. The lateral defenders and midfield players were the players that mostly initiated the attacking plays that resulted in the goals scored. In sum, it was possible to conclude that network metrics can provide a new visualization and understanding of team-members' behaviour, as well as characterize some patterns of play giving sport's analysts information complementary to the traditional notational analysis.

Key words: *match analysis, football, network, collective behaviour, goals scored, goals conceded*

Introduction

Match analysis has been growing in the last few years supported by new technological advances and new multidisciplinary approaches mainly using methods of social sciences and mathematics (Duarte, et al., 2012). Such contributions are extremely important in improving the quality of match analysis, as well as in better understanding of the patterns of play of each team that justify the final outcomes such as passes, shots or goals scored (Barreira, Garganta, Guimarães, Machado, & Anguera, 2014). In fact, the traditional notational analysis, based on individual indicators, have now been complemented by a more complex analysis that aims at identifying and characterizing the specificities of team-members' interactions during a match (Bartlett, Button, Robins, Dutt-Mazumder, & Kennedy, 2012). There are some new alternatives such as temporal-patterns (Jonsson, et al., 2006), network analysis (Bourbousson, Poizat, Saury, & Seve, 2010) and spatio-temporal metrics (Clemente, Couceiro, Martins, Mendes, & Figueiredo, 2013). Each type of analysis provides specific information for analysts and depends on different requirements. In the case of temporal-patterns, specific information about more regular plays the team adopts in attack is provided (Jonsson, et al., 2006). In the case of spatio-temporal metrics, the collective organisation of a team (centroid, stretch index, surface area, effective area of play or tactical region) during a match is automatically computed (Bartlett, et al., 2012). In the case of network analysis, the interactions between teammates that characterize the overall relationship is identified (Passos, et al., 2011).

In the case of football analysis, network analysis can provide a specific and important contribution to the characterization of play patterns and specific interactions between teammates (Martins, Clemente, & Couceiro, 2013). Despite this important and interesting contribution, few studies have been published using such an approach. One of the first studies using the network approach was performed to identify the best players at the European 2008 Championships (Duch, Waitzman, & Amaral, 2010). In their study they identified the attacking plays that resulted in shots on goal. Using a centrality approach for each analysed team they found the player with the highest influence (central midfield player, Spanish team). Moreover, eight out of the twenty players from the article list of best performing players were also selected for the twenty-player team of the tournament (Duch, et al., 2010).

Another study inspecting the attacking stage of the game was performed on basketball games (Bourbousson, et al., 2010). Their study aimed at identifying the team coordination network and the team-members' interactions. The number of team-members involved in each attacking play was considered in the study. It was found that for the majority of time (42%) the unit of attack involved all five team-members of the team (Bourbousson, et al., 2010). In water-polo a network analysis of attacking units was performed aiming at identifying how the number of intra-team interactions emerges in a match and the most successful types of intra-team interactions were characterized (Passos, et al., 2011). The main finding suggested that the most successful collective system behaviour required a high probability of each player interacting with the other team players (Passos, et al., 2011). Such evidence was also found in a study that investigated the English Premier League (Grund, 2012). In the study it was found that a great density level of cooperation was associated with a greater number of goals scored. Recently, the defence-attack transition in football was characterized by Malta and Travassos (2014). In their study it was found that the analysed team had patterns of supported play (with a great prominence in the midfield region) and direct play (with a great prominence in the central forward region). Moreover, Malta and Travassos (2014) also found that defensive midfielders had a great level of degree centralization (out-degree) during the build up of attacks and forwards had a great level of degree prestige metric (in-degree, players that mostly receive passes from team-members) representing that such players are the key playmarkers of the team in the moments ball possession.

Generally, in all studies the network analysis based on graph theory was used. This approach allows the use of a semi-automated system where, after observation, some metrics that identify the team properties in a scientific way can be computed (Wasserman & Faust, 1994). Nevertheless, a single use of a network graph immediately allows coaches and analysts to identify some team properties, thus helping them to characterize the linkage between teammates. As previously observed the majority of network analysis of sport was focused on the observation of the processes in attack. Nevertheless, goals do not have a specific study dedicated to understanding how a team scores and concedes goals. Moreover, the majority of studies use a small number of analysed matches such as case studies (5 to 6 matches) or even pilot studies (1 match) (Clemente, Couceiro, Martins & Mendes, 2014). Such an approach can compromise the interpretation of results over a long period of time.

Therefore, the present study aims at characterizing the goals scored and conceded by a professional football team during a whole season. For this approach the team-members' interactions, of the team in focus and of their opponents, through the passes performed as the linkage indicator will be analysed. Moreover, the regions of the field where such interactions occur are going to be identified. A set of network metrics will be computed to characterize the team-members' interaction process. In our approach only the passing sequences resulting in the goals scored or conceded will be considered. This is quite different from the regular studies that use network approach to analyse the style of play of football teams across all passing sequences (Cotta, Mora, Merelo & Merelo-Molina, 2013; Clemente, et al., 2014). Our approach reduces the potential to understand the patterns of play but increases the possibility to identify some efficacy processes that can be masked by a big dataset of full passing sequences.

Methods

Sample

Thirty-six official matches of the same professional team from the Premier Portuguese Football League (Liga Zon Sagres) and the official Cups (Portuguese Cup and Super-Cup) were analysed during a full season. Only the attacking plays that resulted in the goals scored and conceded by the observed team were collected and processed. The scored and conceded goals achieved after free-kicks and penalties, as well as self-goals and one-man attacking plays (individual actions where only one player developed play with no intervention of any other player) were excluded. Using this procedure forty-six goals scored and fifteen goals conceded were collected. Each unit of attack that resulted in the goal scored either for or against the followed team originated one individual adjacency matrix that represented the passing sequence between the team-members.

Procedures

As described before, two analyses were performed in this study: i) team-members' connectivity; and ii) regions' connectivity. Each goal scored or conceded was processed in the analysis. The thirty-six matches were recorded using the national TV transmission. The observation procedure was performed by the same observer. A test-retest reliability for the observations was performed, respecting a 20-day interval for re-analysis, thus trying to avoid task familiarity issues (Robinson & O'Donoghue, 2007). For this, Cohen's kappa test was used. To test the original dataset, 20% of full data for reliability analysis was tested. The kappa value was 0.81, thus being classified as a very good.

For the team-members' connectivity analysis the strategic distribution of players in the team aiming at classifying the position of each player on the field was identified (see Figure 1). The players were classified as follows: GK – goalkeeper (1); RLD – right lateral defender (2); RCD – right central defender (3); LCD – left central defender (4); LLD – left lateral defender (5); MR and ML – midfielder of right (6) and left side (7); F – attacking midfielder (8); S – striker (9); RF – right forward (10); and LF – left forward (11). This was the strategic display adopted by the team for all the observed matches. The same distribution was followed for the opponent teams.

To analyse the regions' connectivity (from the region where the ball had been recovered until the last region from where the goal was scored or ceded) the field was split into eighteen regions (see Figure 2) as proposed by Carling, Williams and Reilly (2005).

To characterize the team's network the criteria was defined by which the players or regions were linked. Only after such a procedure it was possible to compute the adjacency matrix per each unit of attack (attacking play) that resulted in the goal scored or conceded. The adjacency matrix is a means of representing which vertices (players or regions in this case study) in a network are adjacent to other vertices (players or regions) (Passos, et al., 2011).

In this case only the units of attack that resulted in the goals scored or conceded were analysed. Thus, the attack starts at the moment that a given player of a team recovers the ball and interacts with other players until a goal is scored. Therefore, this was the procedure used to collect the data about the play that resulted in the goals scored or conceded. The goals scored represent the preceding passing sequences of the analysed team that resulted in the goals scored. The goals conceded represents the preceding passing sequences of the opposing teams that resulted in the goal conceded by the analysed team.

Team-members' connectivity was defined as the link criteria of the passes performed between team-members until the moment the goal was scored. In the case of regions' connectivity the regions where each player received the ball was considered; thus, the ball's trajectory on the field during the attacking play was tracked.

Network procedures

The criteria to link the players and regions have been defined previously. Thus, from each

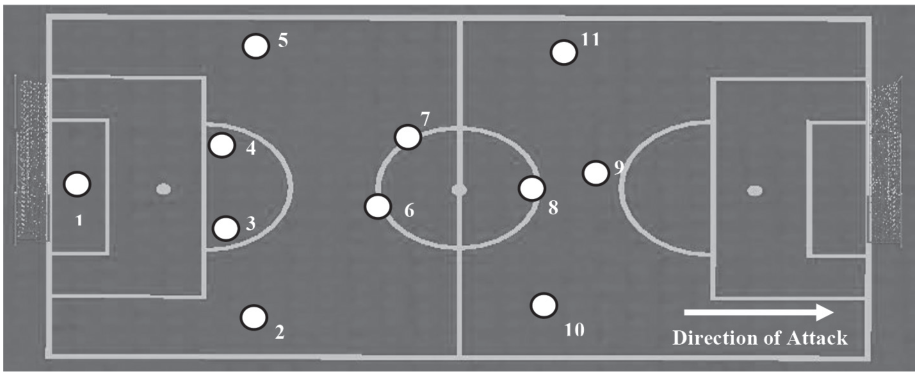

Figure 1. Regular strategic distribution of the players during the season.

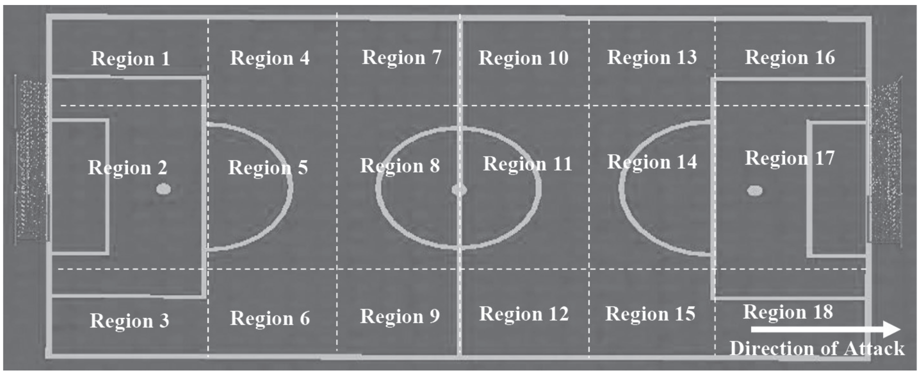

Figure 2. Field split in eighteen regions.

attacking play it is possible to compute the adjacency matrix. The adjacency matrix is used to build a finite network where the entries represent the linkages between the vertices (Couceiro, Clemente, & Martins, 2013) per each attacking play that resulted in the goal scored. To perform the connectivity between teammates the definition used was 0 (zero) value as the non-connectivity and 1 (one) as the connectivity. Let us provide an example in the following Figure 3.

In the case of Figure 3 a single attacking play that resulted in the goal scored was analysed (by either the team or the opponent). Thus, during the observation it is important to establish the linkage between teammates to be able to compute the final graph (network). In this particular case there was only one connection between the same players with the same direction (e.g. player 8 to player 9). Nevertheless, in the cases where such linkage occurs more than once the final number will be *n*. The same procedure was performed for the regions' connectivity.

To generate the forty-six adjacency matrix for the goals scored and fifteen per each goal ceded for both cases (teammates' connectivity and regions' connectivity), the global matrix (sum of all adjacency matrices) was finally performed. Such matrix adds all connections which occurred in all the adjacency matrices. The following will provide an example for the case of only two individual adjacency matrices (Figure 4).

The procedure of the sum of all adjacency matrices was performed for both team-members' connectivity and regions' connectivity. Thus, four final matrices were built: i) team-members' connectivity in attacking plays that resulted in the goals scored; ii) regions' connectivity in attacking plays that resulted in the goals scored; iii) the opponents' connectivity in attacking plays that resulted in the goals conceded (players involved in the attacking play); and iv) connectivity of regions in the opponents' attacking plays that resulted in the goals conceded.

a)

	1	2	3	4	5	6	7	8	9	10	11
1	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	1	0	1	0	0	0
6	0	0	0	0	1	0	0	0	0	0	0
7	0	0	0	0	0	0	0	1	0	0	0
8	0	0	0	0	0	0	1	0	0	1	0
9	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0

b)

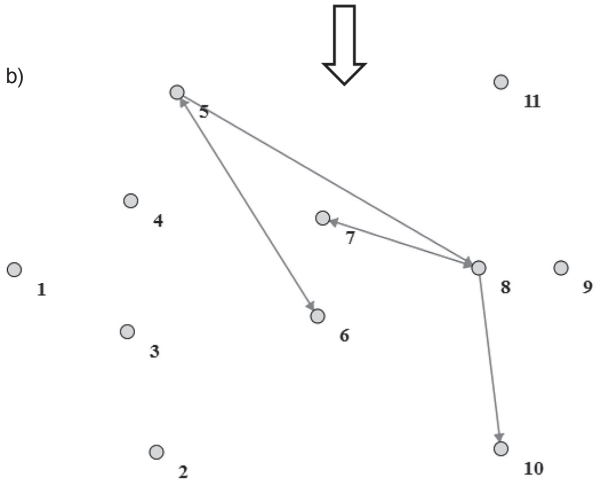

Figure 3. The methodology of data collection: a) Adjacency matrix of one attacking play, and b) Network of the adjacency matrix build.

Network analysis

The network analysis was performed using the free software Social Network Visualizer (SocNetV 0.81). Using this software it was possible to build the graphs based on the matrices previously developed for teammates' interactions and regions' connectivity. Besides the graph building it was possible to compute a set of metrics that mathematically described the connectivity of the graph. All of the following metrics and their algorithms can be observed in detail in the book *"Advances in social network analysis: Research in the social and behavioral sciences"* (Wasserman & Faust, 1994).

Clustering coefficient of the team (CCT)

The clustering coefficient of a player allows the level of interconnectivity with the other teammates to be identified (Horvath, 2011). The higher the values of the CCT the higher the cooperation between the team-members, and vice-versa: the lowest values mean that team-members did not cooperate much with each other.

Indegree centralities (IdC)

The degree of centralities of the node is a measure of the "activity" of the node it represents. In the specific case of the indegree of a node, $d_I(n_i)$, is the number of nodes that are adjacent to node n_i, and thus is the number of arcs terminating at n_i (Wasserman & Faust, 1994, pp. 126). In the particular case of football it denotes the player that received more passes from his team-members than the other players. Values close to 0 mean that all indegrees are equal and the higher values mean that one node is approached from every other node.

Outdegree centralities (OdC)

The outdegree of a node, $d_0(n_i)$ is the number of nodes adjacent to n_i, thus is the number of arcs

Attacking play 1

	1	2	3	4	5	6	7	8	9	10	11
1	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	1	0	1	0	0	0
6	0	0	0	0	1	0	0	0	0	0	0
7	0	0	0	0	0	0	0	1	0	0	0
8	0	0	0	0	0	0	1	0	0	1	0
9	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	1
11	0	0	0	0	0	0	0	1	0	0	0

Attacking play 2

	1	2	3	4	5	6	7	8	9	10	11
1	0	0	0	0	0	0	0	0	0	1	0
2	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	1
6	0	0	0	0	0	0	0	0	0	0	1
7	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	1	0	0	0	0	0
11	0	0	0	0	1	0	0	1	0	0	0

Sum of both attacking plays

	1	2	3	4	5	6	7	8	9	10	11
1	0	0	0	0	0	0	0	0	0	1	0
2	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	1	0	1	0	0	1
6	0	0	0	0	1	0	0	0	0	0	1
7	0	0	0	0	0	0	0	1	0	0	0
8	0	0	0	0	0	0	1	0	0	1	0
9	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	1	0	0	0	0	1
11	0	0	0	0	1	0	0	2	0	0	0

Figure 4. Example of summing up of two attacking plays to build the final matrix.

originating with node n_i (Wasserman & Faust, 1994, pp. 126). In the case of football it denotes the player that originates more passes than the other players. Values closest to 0 means that all outdegrees are equal and the higher values mean that one node nominates and dominates the other nodes.

Closeness centralities (CC)

This measure focuses on how close a node is to all the other nodes in the set of nodes. The idea is that a node is central if it can quickly interact with all the others (Wasserman & Faust, 1994, pp. 181). The variance attains its minimum value of 0 in a network with equal players' indices (equal distances between all players). The highest values mean a star graph situation, where one player chooses all the other players, and the other players interact only with this one, central player (Wasserman & Faust, 1994, pp. 180).

Betweeness centrality (BC)

The interaction between two non-adjacent players might depend on another player (the third player) in the set of players, especially the players who are positioned on the paths between the two, thus these "other players" might possibly have some control over the interactions between the two non-adjacent players (Wasserman & Faust, 1994, pp. 188). In social network analyses, players with higher BC scores might be those who were more often situated between their team-members, thus acting as bridges between team-members. Values closer to 0 occur when all players have the same between index and the highest values occur when one player falls on all other geodesics between all the remaining players (a star graph) (Wasserman & Faust, 1994, pp. 192).

Using all these metrics the individual value of each player/region was computed, thus allowing the identification of individual contributions to attacking plays completed by the goals scored or ceded. The analysis of results was undertaken in a descriptive way.

Results

Teammates' interaction

The analyses performed on the goals scored (Figure 5) led to identifying the attacking midfielder (player 8), left forward (player 11) and right forward (player 10) as the players with the highest values of %IdC (15.38% in all the three cases).

	CCT	%IdC	%OdC	%CC	%BC
Player 1	0.00	0.00	1.92	5.49	0.00
Player 2	0.36	9.62	15.38	12.09	17.86
Player 3	0.00	3.85	1.92	6.04	0.00
Player 4	0.00	3.85	1.92	6.36	0.00
Player 5	0.43	7.69	11.54	10.07	2.68
Player 6	0.35	5.77	9.62	9.30	4.46
Player 7	0.38	11.54	13.46	10.99	9.29
Player 8	0.50	15.38	9.62	9.30	16.13
Player 9	0.50	11.54	9.62	9.30	0.36
Player 10	0.43	15.38	11.54	10.07	22.38
Player 11	0.36	15.38	13.46	10.99	26.85
Group	0.30	0.40	0.40	0.58	0.12

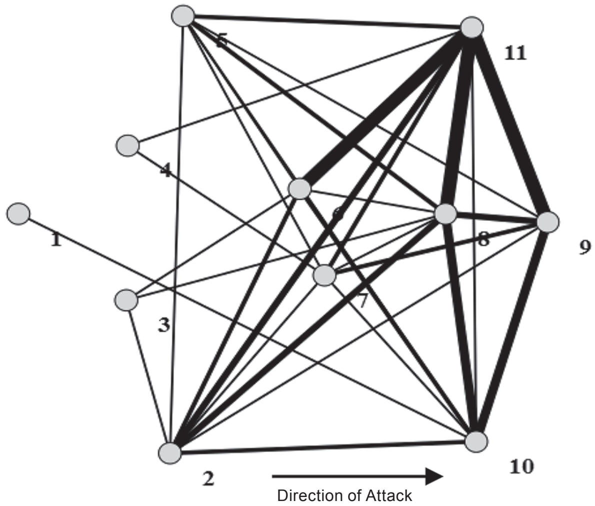

Figure 5. Table with the percentages of each node per each network metric and the respective graph of teammates' interactions preceding the goals scored.

Moreover, it was found that the right defender (player 2) had the higher value of %OdC (15.38%) and %CC (12.09%). In the case of %BC the highest values were found in the left forward (26.85%) and right forward (22.38%). The CCT results showed that the highest values (0.50) were achieved by player 8 (attacking midfielder) and player 9 (striker).

The analyses of the goals conceded (Figure 6) showed that the highest values of %IdC were achieved by the opposing strikers (player 9 – 21.74%) and left forwards (player 11 – 26.09%). On the other hand, the highest values of %OdC (17.39%) were found for the right defenders (player 2), midfielders (player 7) and strikers (player 9).

Moreover, the highest values of %CC and %BC were found for players 2 (15.52% and 15.10%, respectively) and players 9 (15.52% and 32.29%, respectively). Finally, the highest value of CCT was found for the left defenders (players 5 – 1.00).

Interactions between regions

Figure 7 shows the interaction between teammates per regions of the field. Figure 7a shows the goals scored and Figure 7b shows the interactions between players of the opposing teams which resulted in the goals conceded. It is possible to observe a higher interaction tendency in the regions closer to the scoring zone. In the case of statistical analysis it was found that the highest %IdC in the goals scored were found in regions 14 (16.22%) and 17 (16.22%). On the other hand, the highest values of %OdC were found in regions 13 (13.51%), 10, 14 and 15 (10.81% for the last three regions).

In the case of %CC the highest values of plays completed with the goals scored were found for regions 14 and 15 (14.81%) and the lowest for regions 3 and 6 (0%). The highest %BC values were found for regions 10 (25.18%) and 14 (14.46%). Finally in

	CCT	%IdC	%OdC	%CC	%BC
Player 1	0.00	0.00	0.00	0.00	0.00
Player 2	0.50	13.04	17.39	15.52	15.10
Player 3	0.00	4.35	0.00	0.00	0.00
Player 4	0.00	0.00	0.00	0.00	0.00
Player 5	1.00	0.00	8.70	7.76	0.00
Player 6	0.42	8.70	8.70	11.29	3.13
Player 7	0.85	4.35	17.39	12.42	11.98
Player 8	0.67	8.70	8.70	12.42	6.25
Player 9	0.38	21.74	17.39	15.52	32.29
Player 10	0.73	13.04	13.04	13.80	4.17
Player 11	0.27	26.09	8.70	11.29	27.08
Group	0.44	0.43	0.21	1.20	0.18

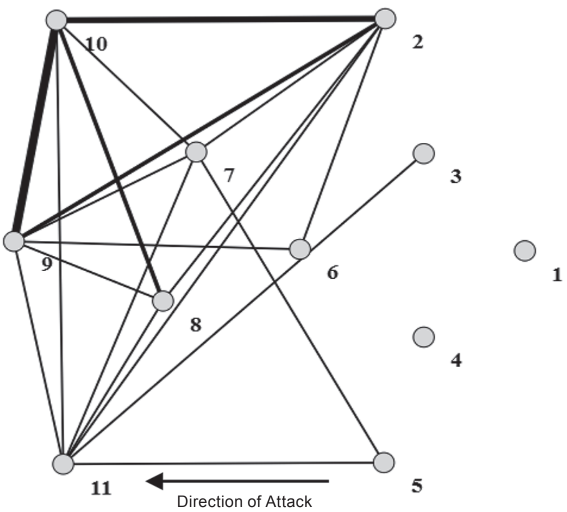

Figure 6. Table with the percentages of each node per each network metric and the respective graph of the opponents' interactions through passes preceding the goals conceded.

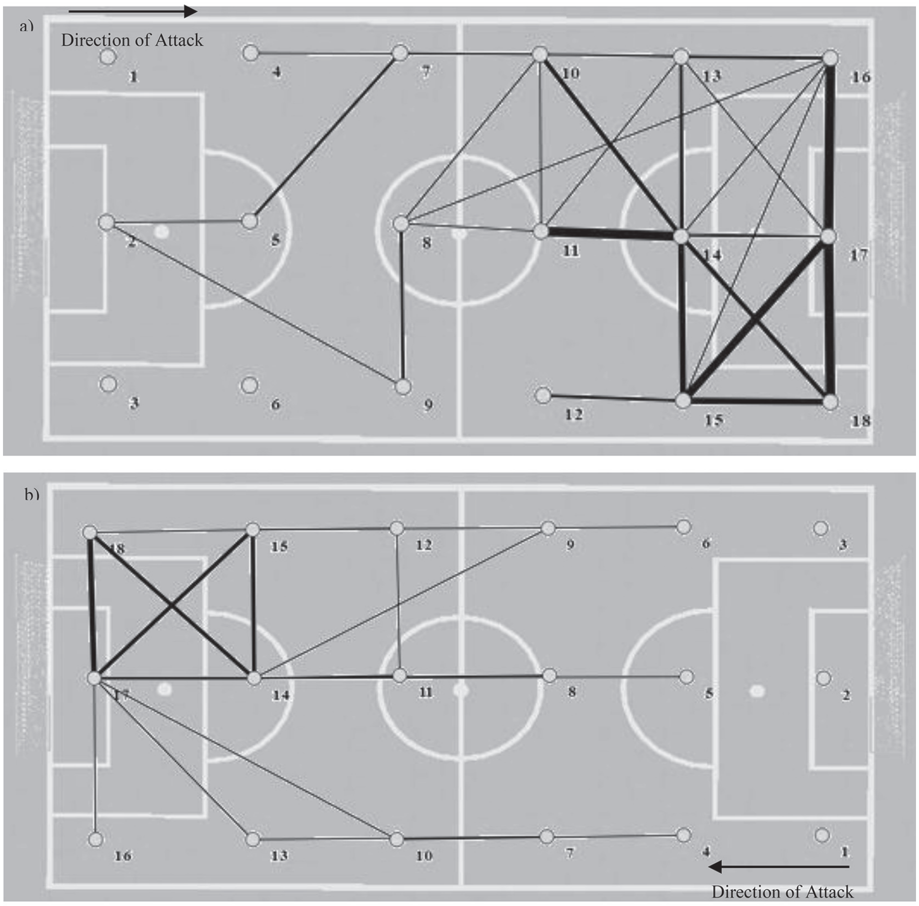

Figure 7. Graph visualization of interactions across field regions preceding the goals scored (a) and goals conceded (b).

Table 1. Percentages of each node per each network metric in the goals scored (a) and goals conceded (b)

a)	CCT	%IdC	%OdC	%CC	%BC
Region 1	0.00	0.00	0.00	0.00	0,00
Region 2	0.00	2.70	5.41	1.60	1.87
Region 3	0.00	0.00	0.00	0.00	0.00
Region 4	0.00	0.00	2.70	3.29	0.00
Region 5	0.00	2.70	2.70	2.19	3.30
Region 6	0.00	0.00	0.00	0.00	0.00
Region 7	0.00	2.70	2.70	3.29	8.91
Region 8	0.33	2.70	8.11	3.95	13.52
Region 9	0.00	2.70	5.41	1.97	6.98
Region 10	0.30	10.81	10.81	5.92	25.18
Region 11	0.55	5.41	8.11	5.38	6.51
Region 12	0.00	0.00	2.70	6.58	0.00
Region 13	0.67	5.41	13.51	6.58	5.82
Region 14	0.43	16.22	10.81	14.81	14.46
Region 15	0.46	10.81	10.81	14.81	5.45
Region 16	0.43	13.51	5.41	9.87	4.76
Region 17	0.39	16.22	5.41	9.87	2.77
Region 18	0.85	8.11	5.41	9.87	0.47
Overall	0.25	0.25	0.18	5.80	0.16

b)	CCT	%IdC	%OdC	%CC	%BC
Region 1	0,00	0,00	0,00	0,00	0,00
Region 2	0.00	0.00	0.00	0.00	0.00
Region 3	0.00	0.00	0.00	0.00	0.00
Region 4	0.00	0.00	4.76	6.17	0.00
Region 5	0.00	0.00	4.76	2.64	0.00
Region 6	0.00	0.00	4.76	2.64	0.00
Region 7	0.00	0.00	4.76	6.17	0.00
Region 8	0.00	0.00	4.76	2.64	0.00
Region 9	0.00	0.00	4.76	2.64	0.00
Region 10	0.00	4.76	4.76	18.50	4.76
Region 11	0.00	4.76	9.52	2.31	2.38
Region 12	0.00	4.76	9.52	2.31	2.38
Region 13	0.00	4.76	4.76	18.50	4.76
Region 14	0.18	23.81	14.29	6.17	54.76
Region 15	0.29	19.05	14.29	6.17	30.95
Region 16	0.00	0.00	4.76	18.50	0.00
Region 17	0.17	28.57	0.00	0.00	0.00
Region 18	0.58	9.52	9.52	4.63	0.00
Overall	0.07	0.30	0.11	25.98	0.08

Appendix

Full adjacency matrix of team-members' interactions in the goals scored.

	1	2	3	4	5	6	7	8	9	10	11
1	0	0	0	0	0	0	0	0	0	1	0
2	0	0	1	0	1	2	2	1	3	4	2
3	0	0	0	0	0	0	0	1	0	1	0
4	0	0	0	0	0	0	0	0	0	0	1
5	0	0	0	0	0	1	2	2	1	1	5
6	0	0	1	0	1	0	0	1	0	1	6
7	0	1	0	1	1	0	0	5	2	2	4
8	0	3	0	0	0	0	1	0	4	4	3
9	0	1	0	0	0	0	2	3	0	2	2
10	0	2	0	0	0	2	1	4	4	0	3
11	0	3	0	1	2	0	2	7	7	1	0

Full adjacency matrix of team-members' interactions in the goals conceded.

	1	2	3	4	5	6	7	8	9	10	11
1	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	2	0	0	2	3	1
3	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	1	0	0	0	1
6	0	1	0	0	0	0	0	0	3	0	0
7	0	1	0	0	0	0	0	0	1	1	1
8	0	1	0	0	0	0	0	0	0	0	1
9	0	0	0	0	0	1	0	1	0	1	2
10	0	0	0	0	0	0	0	2	4	0	1
11	0	0	1	0	0	0	0	0	1	0	0

Full adjacency matrix of interactions across field regions in the goals scored.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0
9	0	1	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	1	0	2	3	0	1	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	1	10	0	0	1	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0
13	0	0	0	0	0	0	0	0	0	1	0	0	0	2	2	2	1	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	4	10	3
15	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	1	5	6
16	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	9	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	9	0

Full adjacency matrix of interactions across field regions in the goals conceded.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
11	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	2	3
15	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	3	1
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	4	0

the goals scored, the highest values of CCT were found for regions 13 (0.67) and 18 (0.85).

The network analysis of the goals conceded showed that the highest values of %IdC were found for regions 14 (23.81%), 15 (19.05%) and 17 (28.57%). Moreover, the highest values of %OdC were found for regions 14 (14.29%) and 15 (14.29%).

The highest values of %CC were found for regions 10 (18.50%), 13 (18.50%) and 16 (18.50%). On the other hand, the highest values of %BC were found for regions 14 (54.76) and 15 (30.95%). Finally in the goals conceded, the results showed that the highest values of CCT were found for regions 15 (0.29) and 18 (0.58).

Discussion and conclusions

This study aimed to use the network approach to increase the understanding of interactions among teammates as manifested in passing sequences. For that reason, this study used network metrics to identify tendencies of the team during their own and their opponents' attacking plays that resulted in the goals scored and goals conceded, respectively. In the case of goals scored and goals conceded only the open-plays were analysed, excluding the goals scored from free-kicks, penalties or corners.

In the case of team-members' interactions and regions' connectivity the passing sequences that resulted in the goals scored were investigated. Thus, the first step was to classify the strategic distribution of players from the observed team on the field and then the interactions were established based on the passes between players' positions. From the analysis performed on the goals scored it was possible to identify the highest values of %IdC. The results showed that the forward (player 8), left forward (player 11) and right forward (player 10) were the players with the highest values of %IdC. In fact, such results make sense because the level of IdC reveals that such players received more passes from their teammates. In the case of football, a higher proximity to the opponent's goal is desirable. Thus, the forwards are the players with the higher tendency to receive the ball in the scoring zone. These results are partially verified by the study that analysed the attacking transition of football (Malta & Travassos, 2014). In the case of attacking transition analysis the defensive midfielder was the player with the highest IdC, maybe due to the specific tendency of the team and due to the sample used (all the transitions, with and without goals or shots). In our case, only the goals scored were considered, thus the forwards took a greater role as the targets of passes sequences. In the case of regions' connectivity, that determines the sequence of passes performed through the field, the highest %IdC in the goals scored were found for regions 14 and 17 (the central regions closest to the opponent's goal). Such results are in accordance with the notational analysis performed in previous years (Barreira, Garganta, Castellano, Prudente, & Anguera, 2014; Carling, et al., 2005; Pollard, Ensum, & Taylor, 2004). A retrospective study analysed the evolution of goals scored in the UEFA European and FIFA World Cup to assess the evolution of attacking patterns of play from 1982 until 2010 (Barreira, Garganta, Castellano, et al., 2014). In their study it was found that the majority of goals were scored from the central regions of the attacking sector. Such evidence seems to be usual in a set of studies that analysed the goals scored (Carling, et al., 2005; Pollard, et al., 2004). Moreover, it has been found that a higher proximity to the goal (mainly using the penalty area) increases the opportunity to score (Althoff, Kroiher, & Hennig, 2010). As an example, in the 2002 FIFA World Cup it was found that the majority of the goals were scored from inside the penalty area (Carling, et al., 2005).

In the case of OdC the highest values were found for the right defender (player 2), midfielder (player 7) and left forward (player 11). In this case, the OdC means that given players originated more passes or sequence of passes. Such values reveal that the team had a tendency to use the sides of the field and the main midfielder to promote the attacking plays that resulted in the goals scored. Once again, such results are in accordance with the values found in the attacking transitions (Malta & Travassos, 2014). Such results can suggest that the main contributors to promoting the attacking plays that resulted in the goals scored are the wing defenders and midfielders. Thus, coaches can use such information to avoid the opponent's attacking.

In the case of the regions the highest values of %OdC were found for regions 13 (13.51%), 10, 14 and 15 (10.81% for the last three regions). Such values are not in line with a similar study that analysed the attacking transition (Malta & Travassos, 2014). In the case of the study conducted by Malta and Travassos (2014) the main OdC regions were the central midfield. The same protocol to split the field in regions was adopted, which allowed comparison with the present results. In our study that analysed the goals scored it was possible to identify that the main OdC regions were the wing midfield and the attacking midfield closest to the penalty area. Maybe these values could be explained by the goals that emerged from the unbalanced defensive organization of the opposing teams that led to their loss of the ball in the lateral and midfield regions. In the cases of analysis of passing sequences it was found that the majority of sequences of play that resulted in goals scored involved 1 to 4 passes (Carling, et al., 2005). Similar results were found in the 1990 (80% of goals scored) and 1994 World Cups (77% of goals scored) (Hughes & Franks, 2005).

The network analysis revealed that the right and left defenders (players 2 and 5), midfielder (player

7) and left and right forwards (players 10 and 11) achieved the highest values of closeness centralities. Globally, the network had a value of 0.58. Such results suggest that the network tends to have some players that team-members use more frequently. In this case, the team uses the wing players to build the attacking plays that resulted in the goals scored. This can suggest a pattern of play of the team, which is highly efficient when using the wing players. Such behaviour can be used to avoid the clustered opponents in the middle area and to stretch the opposing players, thus increasing the opportunities to penetrate the scoring zone (Costa, Garganta, Greco, Mesquita, & Seabra, 2010). In the case of the regions, the highest values of %CC were found for regions 10, 13 and 16 – the three left regions of the attacking midfield. An evident tendency to build up the attacking plays completed with the goals scored emerges in this analysis. Such information can be extremely important to help coaches to identify the patterns of play of their team and even of the opposing team.

For the betweeness centralities (BC) it was found that the highest values belong to the right defender, forward, right forward and left forward. Such values means that these players are the ones that most promote connections with the other players of the network. In fact, as previously observed, the majority of these players had the highest values of closeness centralities. Therefore, the main tendency of the team could be the participation of these players in attacking plays performance. Moreover, all of these players play on the sides of the field or to the central region closest to the scoring zone. Therefore, the attacking efficacy results from a greater tendency to use the wings and reduce the exposure to the central region of the field before the last pass. To confirm this observation the %BC per regions of the field was inspected. It was found that the highest values were for regions 8 (central midfield), 10 (left midfield) and 14 (forward midfield). Such results reveal that these regions can be associated with three main moments of building up attack. The first moment (ball recovery) occurs in general on the central midfield, then the players opt to exploit the wings and finally the ball is returned to the central region (closest to the scoring zone) to increase the possibility of scoring.

Until now, the discussion has focused on the goals scored. It was found that there was a clear tendency to exploit the wing players and to use the forwards to score. Nevertheless, it is also important to identify how the team conceded goals. Therefore, a similar analysis was performed on all goals conceded during the season. It is important to highlight that the goals conceded were fewer than goals scored. The analyses of the goals conceded showed that the highest values of %IdC were achieved by the opposing striker and right and left forwards. Such findings are similar to the goals scored. Therefore, the similarities between the properties of attacking play completed by a goal of the analysed team and the opponents can be discussed. In fact, during attacking plays forwards are targeted to finalize sequences of passes. Moreover, the regions with the highest %IdC were 14 (forward region), 15 (right midfield) and 17 (penalty area), that is the regions closest to the scoring zone. Once again, such results are similar to the analyses performed on the goals scored.

In the case of %OdC for the goals conceded it was found that the opposing right defender, midfielder and striker had the highest values. While in the case of the right defender and midfielder such values are in line with the findings for the goals scored, in the case of attacking transition (Malta & Travassos, 2014) the values of striker also suggests a specific action that originates in goals conceded. In fact, the OdC reveals that these opposing players originated more passes or sequences of passes; thus, the great value of strikers can suggest that the team conceded goals to counter-attacks and maybe after losing the ball in regions closest to their defensive area (in the first third of the pitch). After the analysis of the %OdC was performed in the regions of play it was found the highest values were in regions 14 (forward midfield) and 15 (right midfield). Such results justify the discussion about the moments of losing the ball that originated the goals conceded had started with the opposing striker, if he recovered the ball and quickly initiated the counter-attack that resulted in a goal.

In the cases of %CC and %BC it was found that the highest values were for the opposing players 2 (right defenders) and players 9 (strikers) in the plays preceding the goals conceded. Moreover, in the case of %BC the left forwards had the second highest value. Such values reveal that these specific positions were greatly used by the opposing team-members and also promoted connections with their team-members, which must be taken into account by the team that conceded goals. An interesting result is the %CC per region that showed the highest values in regions 10, 13 and 16 (all the left regions in the opponents' attacking half of the field). Thus, such results suggest a specific pattern of play by the team exposed to the attacking plays from those regions. On the other hand, the highest values of %BC were found in regions 14 and 15. Such values indicate that both regions must be taken into account by the defending team in order to avoid the opponent's players using their team-members from there.

Throughout this study it was possible to characterize the patterns of play of one team in their goals scored and conceded. The network analysis provided a set of outcomes that mathematically determined specific connections between teammates. However, there are some limitations to the current study. The fact of only one team being analysed caused that the main results cannot be compared with other teams nor generalized. This is an issue that must be fixed in future studies by a bigger range of data on different teams even with different final scores. Another limitation of our study was the codification of team tactical position. In a game a given player may assume different tactical positions associated with their specific mission and role. Nevertheless, for a network classification a specific codification must be done, thus maybe in the future players should be codified in alternative ways. The main advantage of our study was the proposition of a new approach: application of the social network analysis to the game of football. Moreover, such information will allow coaches to find a user-friendly solution when analysing their players' interactions during a game. The system depends on the user's observation and recording using semi-automated systems. Nevertheless, it provides different outcomes from the traditional notational analysis, providing coaches and sports analysts with information about the team's processes during attack building up. In the specific case of this study, it was possible to determine the patterns of team-members' interactions leading to the goals scored. Also, it was possible to delineate how the opponents' players scored. Such information can provide a bigger picture when reconciled with the notational analysis. In sum, the network analysis can be a useful tool for the future of match analysis, giving specific information about the tactical behaviour of the team using the characterization of team-mates interactions during attacking processes. The network analysis usefulness is clearly evident in classifying the centrality level of players and tactical positions that contribute most to the identification of the team style of play. Thus, network analysis can be useful as research method and as a scouting tool as well.

The performance analysis of football teams aims to increase the opportunity to understand the specific properties of team-members' interactions. In this study a network analysis was performed of team-members' interactions and regions of attacking plays that resulted in goals scored and conceded. It was possible to identify that wing midfielders and attacking midfielder were the most used players and lateral defenders were the players that initiated most attacking plays. Moreover, it was possible to identify that the attacking midfield zone was the main region that contributed most to goals scored and conceded. Such results indicate the potential of network analysis for a new vision of match analysis. Network metrics may complement the traditional notational analysis by giving specific insights into the processes of play of the team. That allows coaches to identify the team's play properties. In sum, network analysis, as a semi-automated system, can be a user-friendly and useful method.

References

Althoff, K., Kroiher, J., & Hennig, E. (2010). A soccer game analysis of two World Cups: Playing behavior between elite female and male soccer players. *Footwear Science, 2*(1), 51-56.

Barreira, D., Garganta, J., Castellano, J., Prudente, J., & Anguera, M.T. (2014). Evolución del ataque en el fútbol de élite entre 1982 y 2010: Aplicación del análisis secuencial de retardos. [Evolution of attacking patterns in elite-level soccer between 1982 and 2010: The application of lag sequential analysis. In Spanish.] *Revista de Psicología del Deporte, 23*(1), 139-146.

Barreira, D., Garganta, J., Guimarães, P., Machado, J., & Anguera, M.T. (2014). Ball recovery patterns as a performance indicator in elite soccer. *Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology, 228*(1), 61-72.

Bartlett, R., Button, C., Robins, M., Dutt-Mazumder, A., & Kennedy, G. (2012). Analysing team coordination patterns from player movement trajectories in football: Methodological considerations. *International Journal of Performance Analysis in Sport, 12*(2), 398-424.

Bourbousson, J., Poizat, G., Saury, J., & Seve, C. (2010). Team coordination in basketball: Description of the cognitive connections among teammates. *Journal of Applied Sport Psychology, 22*(2), 150-166.

Carling, C., Williams, A.M., & Reilly, T. (2005). *Handbook of soccer match analysis: A systematic approach to improving performance*. London & New York: Taylor & Francis Group.

Clemente, F.M., Couceiro, M.S., Martins, F.M.L., & Mendes, R.S. (2014). Using network metrics to investigate football team players' connections: A pilot study. *Motriz*, *20*(3), 262-271.

Clemente, F.M., Couceiro, M.S., Martins, F.M.L., Mendes, R., & Figueiredo, A.J. (2013). Measuring collective behaviour in football teams: Inspecting the impact of each half of the match on ball possession. *International Journal of Performance Analysis in Sport, 13*(3), 678-689.

Costa, I.T., Garganta, J., Greco, P.J., Mesquita, I., & Seabra, A. (2010). Influence of relative age effects and quality of tactical behaviour in the performance of youth football players. *International Journal of Performance Analysis in Sport, 10*(2), 82-97.

Cotta, C., Mora, A.M., Merelo, J.J., & Merelo-Molina, C. (2013). A network analysis of the 2010 FIFA World Cup champion team play. *Journal of Systems Science and Complexity*, *26*(1), 21-42.

Couceiro, M.S., Clemente, F.M., & Martins, F.M.L. (2013). Towards the evaluation of research groups based on scientific co-authorship networks: The RoboCorp case study. *Arab Gulf Journal of Scientific Research, 31*(1), 36-52.

Duarte, R., Araujo, D., Freire, L., Folgado, H., Fernandes, O., & Davids, K. (2012). Intra- and inter-group coordination patterns reveal collective behaviors of football players near the scoring zone. *Human Movement Science, 31*(6), 1639-1651.

Duch, J., Waitzman, J.S., & Amaral, L.A. (2010). Quantifying the performance of individual players in a team activity. *PloS One, 5*(6), e10937.

Grund, T.U. (2012). Network structure and team performance: The case of English Premier League soccer teams. *Social Networks*, *34*(4), 682-690.

Horvath, S. (2011). *Weighted network analysis: Applications in genomics and systems biology*. New York: Springer.

Hughes, M., & Franks, I. (2005). Analysis of passing sequences, shots and goals in soccer. *Journal of Sports Sciences, 23*(5), 509-514.

Jonsson, G.K., Anguera, M.T., Blanco-Villaseñor, Á., Losada, J.L., Hernández-Mendo, A., Ardá, T., ... Castellano, J. (2006). Hidden patterns of play interaction in soccer using SOF-CODER. *Behavior Research Methods, 38*(3), 372-381.

Malta, P., & Travassos, B. (2014). Characterization of the defense-attack transition of a soccer team. *Motricidade, 10*(1), 27-37.

Martins, F.M.L., Clemente, F.M., & Couceiro, M.S. (2013). From the individual to the collective analysis at the football game. Paper presented at the *Proceedings Mathematical Methods in Engineering International Conference*, Porto.

Passos, P., Davids, K., Araújo, D., Paz, N., Minguéns, J., & Mendes, J. (2011). Networks as a novel tool for studying team ball sports as complex social systems. *Journal of Science and Medicine in Sport, 14*(2), 170-176.

Pollard, R., Ensum, J., & Taylor, S. (2004). Estimating the probability of a shot resulting in a goal: The effects of distance, angle and space. *International Journal of Soccer and Science, 2*(1), 50-55.

Robinson, G., & O'Donoghue, P. (2007). A weighted kappa statistic for reliability testing in performance analysis of sport. *International Journal of Performance Analysis in Sport, 7*(1), 12-19.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. New York: Cambridge University Press.

Submitted: May 30, 2014
Accepted: February 17, 2015

Correspondence to:
Filipe Manuel Clemente
Instituto Politécnico de Viana do Castelo, Escola Superior de Desporto e Lazer, Melgaço, Portugal
Complexo Desportivo e Lazer de Melgaço – Monte de Prado 4960-320, Melgaço, Portugal
E-mail: Filipe.clemente5@gmail.com

Acknowledgments

This study was carried out within the scope of R&D Unit 50008, financed by UID/EEA/50008/2013.