Original Article

Institution of MECHANICAL ENGINEERS

Journal of SPORTS ENGINEERING AND TECHNOLOGY

Proc IMechE Part P: J Sports Engineering and Technology
2016, Vol. 230(3) 158–170

DOI: 10.1177/1754337115597335
pip.sagepub.com
SAGE

Performance Analysis Tool for network analysis on team sports: A case study of FIFA Soccer World Cup 2014

Filipe Manuel Clemente[1,2,3], Frutuoso Silva[4,5], Fernando Manuel Lourenço Martins[1,2,4], Dimitris Kalamaras[6] and Rui Sousa Mendes[1,2]

Abstract

The study of teammates' interaction on team sports has been growing in the last few years. Nevertheless, no specific software has been developed so far to do this in a user-friendly manner. Therefore, the aim of this study was to introduce a software called the Performance Analysis Tool that allows the user to quickly record the teammates' interaction and automatically generate the outputs in adjacency matrices that can then be imported by social network analysis software such as SocNetV. Moreover, it was also the aim of this study to process the data in a real-life scenario, thus the seven matches of the German national soccer team in the FIFA World Cup 2014 were used to test the software and then compute the network metrics. A dataset of 3032 passes between teammates in seven soccer matches was generated with the Performance Analysis Tool software, which permitted a study of the network structure. The analysis of variance of centrality metrics between different tactical positions was made. The two-way multivariate analysis of variance revealed that the strategic position ($\gamma = 1.305$; $F = 24.394$; $p = 0.001$; $\eta_p^2 = 0.652$; large effect size) had significant main effects on the centrality measures. No statistical differences were found in the phase of competition ($\gamma = 0.003$; $F = 0.097$; $p = 0.907$; $\eta_p^2 = 0.003$; very small effect size). The network approach revealed that the German national soccer team based their attacking process on positional attacks and not in counter-attack, and the midfielders were the prominent players followed by the central defenders. The Performance Analysis Tool software allowed the user to quickly identify the teammates' interactions and extract the network data for process and analysis.

Keywords

Match analysis, software, graphical application, graphical user interface, German national team

Date received: 24 March 2015; accepted: 30 June 2015

Introduction

The dynamic nature of team sports leads to a cooperation–opposition relationship.[1] For this reason, analysis of both relationships is truly relevant to increase the possibility of overcoming the opponent, exploring their weaknesses and avoiding their strengths.[2] This process of analysis is usually defined in the sports field as match analysis.[3] In recent decades, a set of alternatives have been proposed to process the match analysis and to augment the coaches' perceptions about the individual and collective performance of the players.[4]

The most common analysis used worldwide is traditional notational analysis based on an observational process.[3,5] This kind of process uses the codification approach to classify individual actions or collective behaviors of players and then extract the statistical data.[6–8] Such analysis depends on a human operator to perform the observation and to codify, thus in some cases this takes long periods of time.[9] Despite the time

[1]Department of Education, Coimbra College of Education, Polytechnic Institute of Coimbra, Coimbra, Portugal
[2]ASSERT, RoboCorp, IIA, Polytechnic Institute of Coimbra, Coimbra, Portugal
[3]Faculty of Sport Sciences and Physical Education, University of Coimbra, Coimbra, Portugal
[4]Delegação da Covilhã, Instituto de Telecomunicações, Coimbra, Portugal
[5]University of Beira Interior, Covilhã, Portugal
[6]New Media Network Synapsis S.A., Athens, Greece

Corresponding author:
Filipe Manuel Clemente, Department of Education, Coimbra College of Education, Polytechnic Institute of Coimbra, Rua Dom João III Solum, 3030-329 Coimbra, Portugal.
Email: Filipe.clemente5@gmail.com

taken, the process is relatively user-friendly and over the years different software tools have been created to do this, such as Prozone-Amisco, Noldus, LongoMatch,[10] FutSat[11] or SoccerEye.[12] In the last decade, the advance of technological instruments (multiple camera systems, Global Positioning System (GPS), or electronic transmitting devices (ETDs)) has provided some new alternatives for the match analysis process. Multiple cameras (automatic video tracking (AVT)) may track the players' trajectories (motion analysis) and moreover identify individual actions and collective behavior of team players (notational analysis).[10] There are some commercial systems that employ such technology in the field of football: (1) ProZone-Amisco (http://www.prozonesports.com), (2) Chukyo University (not-commercial), (3) Hiroshima College of Sciences (not-commercial), (4) Datatrax (http://www.datatrax.com), (5) Tracab (http://www.tracab.hegogroup.com) and (6) University of Campinas (not-commercial). GPS may track the players' trajectories (time–motion analysis) but does not provide individual and collective information about the technical and tactical performance of players.[13] The companies that have employed such technology are as follows: (1) Catapult-GPSport (http://www.catapultsports.com) and (2) STATSports Viper Pod (http://www.statsports.ie). Finally, ETDs that can also track the players' trajectories but not codify individual actions and collective behaviors.[14] The companies that have employed such technology are as follows: (1) InMotio Object Tracking BV (http://www.inmotio.eu), (2) Citech Holdings Pty Ltd (http://www.citechholdings.com) and (3) Viper Live Streaming (http://www.statsports.ie). The systems that collect the two-dimensional and three-dimensional coordinates and track the players' trajectories (AVT, GPS, ETD) make time–motion analysis possible.[15] Such analysis gives information about the distance covered, the speed, the time spent in a specific intensity of running speed, the distance covered in a specific intensity of running speed, acceleration, deceleration and changes in direction.[2] Such information gives coaches and analysts the global picture about the physical demands of each tactical position, as well as the specific fitness level of players during the match and training sessions. Moreover, it is now possible to measure teammates' cooperation and synchronization and to classify patterns of play and organizational dynamics.[16] Computational metrics such as *w*Centroid,[17] *w*Stretch Index,[18] Sectorial Lines,[19] Territorial Domain,[20] Attacking Definition Zone[21] and Defensive Play Area[4] were developed in the last 5 years.

Finally, the study of the social network process in team sports such as the match analysis approach has been adopted by the scientific community.[22,23] One of the first studies in team sports was made in the basketball context to identify the structure of cooperation among French players.[24] Then in the football context, social network analysis was used to classify the prominent player in the UEFA Cup 2008 and to associate the network output with the coaches' opinions.[22] Lately, in the FIFA World Cup 2010, social network analysis was used to identify prominent players in the finals and the results were similar; Xavi (a midfielder in the Spanish Team) had the highest centrality values.[25,26] Using a general approach to study network in football, it was found that greater values of density and smaller values of heterogeneity led to the best performance in the Premier Football League.[23] Lately, the centrality metrics of network have been used to identify the patterns of play of teams, identifying the players who had a major participation in building attacks and the players who are more involved.[27,28]

Despite this massive evolution in team sports analysis, no specific software to analyze the teammates' cooperation (one of the main points of analysis) from a network point of view has been developed so far. In fact, there is a general software to compute the social network metrics such as SocNetV, AllegroGraph, Automap, EgoNet, Gephi, GraphStream or NetworkKit. Nevertheless, such software is not specific for sports and needs adjacency matrices that come from teammates' analysis calculated using regular software (such as Excel) or engineering software (MATLAB). Thus, the sports community needs a specific user-friendly software tool that automatically exports the adjacency matrices, reducing the time spent in data processing and making the observational process easier for sports operators.

Therefore, the aim of this study was to introduce the new Performance Analysis Tool (PATO) and describe its possibilities for the scientific community in sports analysis. Moreover, and in order to prove its accuracy, PATO was used in the analysis of the German national soccer team during the FIFA World Cup 2014 matches and the outcomes will be shown in this study. The network metrics will be used to classify the prominent players and to analyze the variance between teammates.

Methods

Sample

Seven matches of the German national soccer team in the FIFA World Cup 2014 were used to test the software and then compute the network metrics. A dataset of 3032 passes between teammates in seven soccer matches enabled a study of network structure. Generally, without PATO and using a simple Excel sheet procedure, it takes around 3 h to calculate adjacency matrices for each match. Using the PATO software, it was possible to reduce the time for the whole of the match (90 min). Nevertheless, it also allows adjacency matrices to be calculated at any time of the game.

Procedures to analyze the teammates' interactions. To process the network analysis in team sports, the first task is to codify the players. In the specific case of football, the positional roles of players can be codified from 1 to 11. The codification allows observers to reduce the time taken to identify the players using the software and to also make the characterization of each positional role. In the present case study, only tactical positions were considered. Thus, when a player was substituted by the other, the only criterion of adjusting it was to follow the codification for each tactical position. In some other cases (like to track the specific players), the codification may change to avoid error of detection. Moreover, the system allows to build the sum adjacency matrix in any time, thus would be recommended to save the adjacency matrix and after substitution to reset the PATO in order to continue the new matrix with the new values of the player who enters in the match.

Once a code has been attributed to each player, the next task is to clearly identify the linkage parameter that connects the players. In the specific case of team sports, such indicators have been the passes.[29] Nevertheless, there are other possibilities such as defensive marking or the momentary positional changes between teammates.[30] In our specific study, the linkage indicator was the pass.

In order to build a proper observation, each unit of attack generated an adjacency matrix. The unit of attack can be defined as the passing sequence of one team without interruption of opponents.[27] In that sense, when a team loses the ball, the unit of attack ends. An interception that did not influence the pass was not considered a ball lost and then the unit of attack goes on. Thus, a unit of attack begins when the team recovers the ball and makes at least one pass to a teammate. Then, the unit of attack ends when the team loses the ball by an opponent interception, ball out of play, or game interruption. After recording the passing sequence, an adjacency matrix is built. This matrix represents the connections between a node (player) and an adjacency node (teammates).[29] In the adjacency matrix, the pass between members is codified by 1 or more in the case of more passes in the same direction (such as Player 3 to Player 4) and 0 for no passes between teammates. It is important to note that Player A passing to Player B is not the same as Player B passing to Player A. For this reason, this is classified as a directed graph (digraph). If the direction is not a criteria (like a connection between A and B), the graph is considered undirected. In our case, the direction is relevant and the number of passes led to a weighted digraph. Generally, the adjacency matrices are not symmetric. We provide an example of the overall process in Figure 1.

Given that in the case of football the passes have directions, this can be classified as weighted directed graphs. For this reason, the links have specific weights

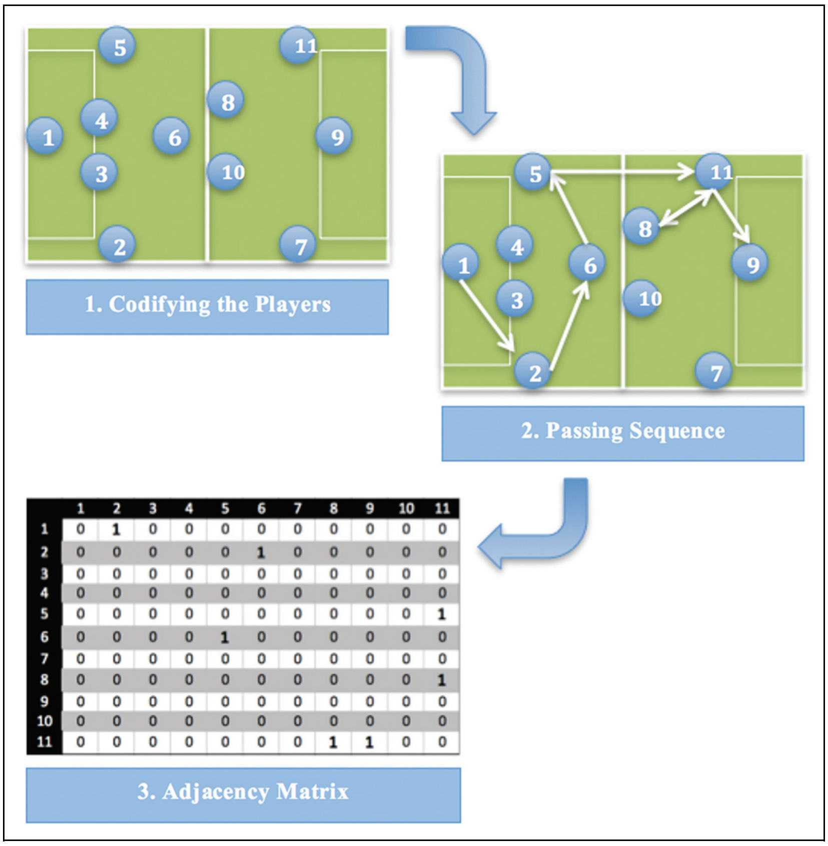

Figure 1. Observational steps on network analysis: (1) codifying the players by the strategic positions, (2) passing sequence of unit of attack and (3) adjacency matrix of the unit of attack.

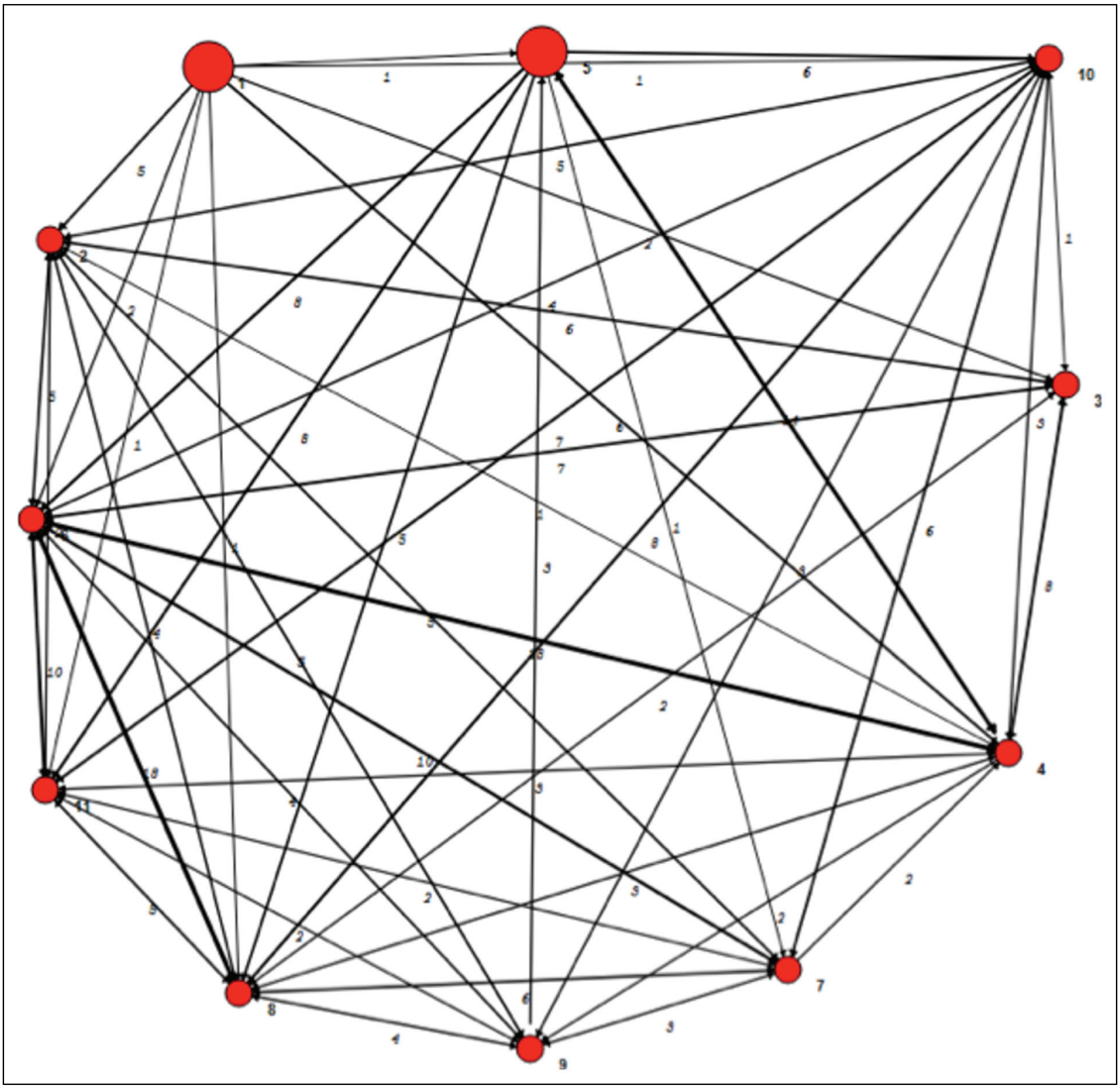

Figure 2. Example of weighted digraph for a single match of a football team.

based on the number of passes. An example can be observed in Figure 2.

For the specific case of this study, a unique researcher with more than 5 years of experience carried out the observational process and codification. The researcher was previously trained and was tested in a test–retest procedure to ensure the reliability of data collected from the FIFA World Cup 2014. Cohen's Kappa test adhered to a 25-day interval for re-analysis to avoid task familiarity issues.[31] For a sample of 15% of the full data, the test–retest had a Kappa value of 0.84, thus enabling the study to proceed.[31]

Introducing the PATO

In order to build specific software to analyze the cooperative behavior of teammates in sports, it was necessary to generate a user-friendly application to be used by coaches and sports researchers in any situation. To do that, the first criterion was to provide the opportunity to the user to choose the number of members to visualize in the layout. This step is represented in Figure 3. The layout was codified based on the positional roles of players, thus it was codified from 1 to 11 in the case of football.

This process increases the possibility of this software being used for any team sport. After choosing the number of players that the observer wants to analyze, the regular layout of analysis emerges as shown in Figure 4.

The PATO format allows two teams to be analyzed by the same observer in a very user-friendly fashion. To do this, it is only necessary to press the top bar (red or blue) to immediately change the color and the team. Only one team can be viewed at a time, nevertheless to

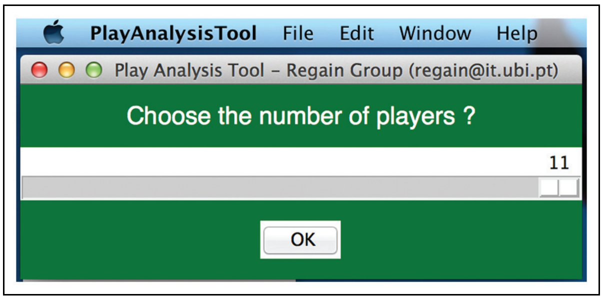

Figure 3. Layout to choose the number of players that observer wants to analyze.

change the codification it is only necessary to change the team in the top bar. This reduces the likelihood of mistakes in the case of two screens for each team. Moreover, the size of the formatting was developed to be adjusted for mobile devices (in future updates), thus only one team is displayed at a time. To begin the analysis, the observer only needs to codify the players in the field with their number and then press the icons to record the passing sequence. Every time that a player icon is pressed, the passing sequence of the unit of attack is displayed in the bar below. Please see the following example in Figure 5.

Figure 4. Layout to perform the observational process (examples for 5 and 11 players).

Figure 5. Recording the passing sequence.

In this example, the passing sequence from player 1 to player 2, then to player 5 and finally to player 10 was observed. As can be verified in Figure 5, the sequence can be immediately viewed by the observer and may also be changed (in the case of mistakes) or even deleted. Once the recording of the passing sequence is finished, it is only necessary to select “add.” Then, the software codifies the passing sequence and converts the information into an adjacency matrix. Every time that the observer wants to codify a given unit of attack based on a special event, he or she only needs to choose the specific event (special or goal) before selecting “add.” The special event may be changed based on observer’s preference.

In order to make exporting the adjacency matrices easier at any time, the software allows the adjacency matrix to be saved as a TXT file and then the user has two possibilities: (1) continue the observation without removing the previous adjacency matrix, thus making it possible to sum all the matrices from the beginning or (2) reset the system in order to build the adjacency matrices from 0. It is also possible to import the adjacency matrices’ files into any software for social network analysis. Note that the save option only requires that the user selects the folder where he or she wants to save the matrices, because the name of the files is created automatically based on the date. This strategy ensures that the observer does not need to divert attention from observing the game to name files. PATO software will generate at least two matrices’ files, one for each team with the following file names: date_Our.txt and date_Vis.txt (our team and visiting team). For more details of PATO options, see their use cases diagram in Figure 6. For example, this software also allows a video of a game to be loaded to be analyzed by the observer (option “Load” in Figure 6).

Computing the network metrics in Social Network Visualizer software

There are some freely available software packages on the market to process the social network analysis. For our specific case and following previous studies,[32] we chose the Social Network Visualizer (SocNetV; version 1.5). This software is a graphical application for the analysis and visualization of social networks.[33] It allows

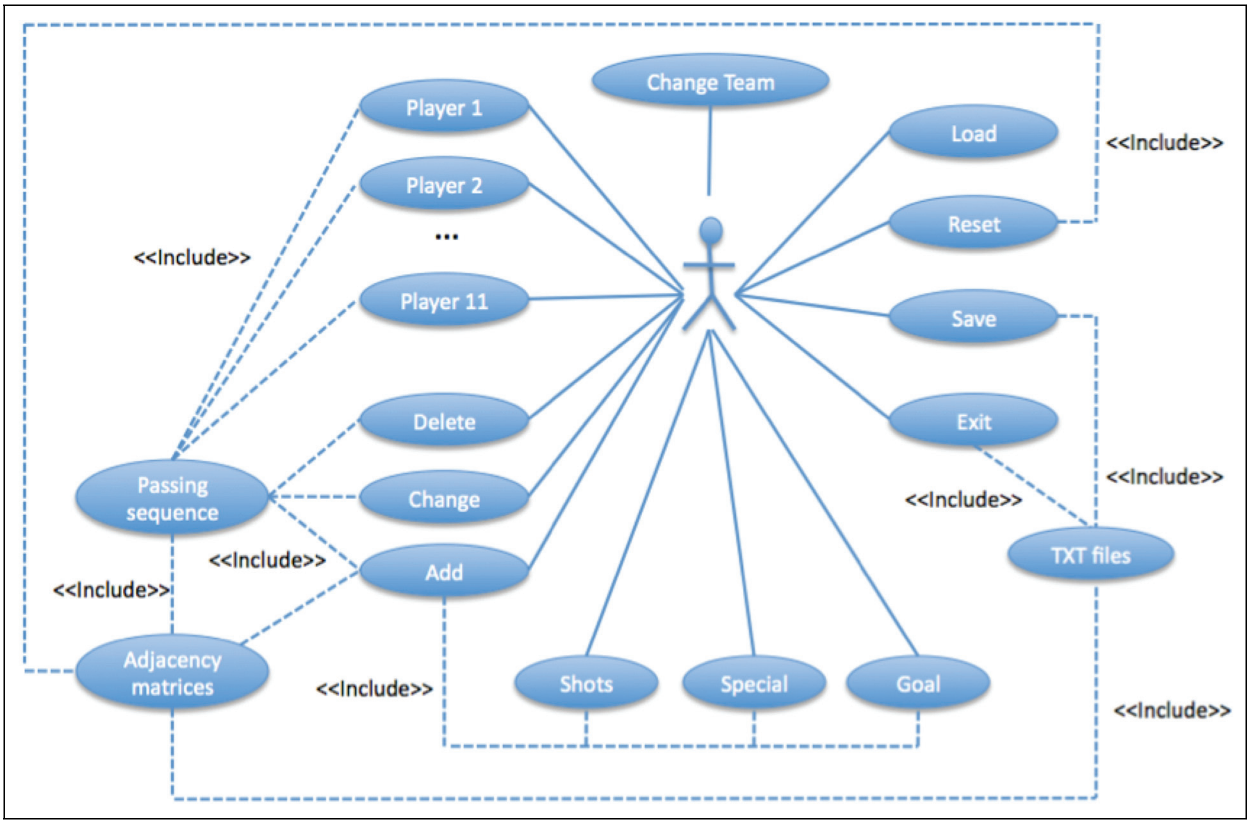

Figure 6. PATO—use cases diagram.

the researcher to load formatted network data such as sociomatrices, to analyze the social and mathematical properties of the corresponding social networks in the form of mathematical graphs, and to compute basic graph properties, such as density, diameter and clustering coefficient, as well as more advanced structural measures such as centrality and prestige indices, which were used in this survey.[33] The software only requires the adjacency matrix to run. Thus, PATO software complements SocNetV for the specific case of team sports.

For the specific case of this study that uses the weighted directed graphs theory, two metrics to characterize the digraph properties were selected (total links and density). For the case of centrality metrics, the degree prestige and degree centrality were chosen. The following gives a brief description of each metric.

Total links. Each element a_{ij} of the adjacency matrix was the number of interactions (passes) from player i to player j and, in terms of the corresponding graph (sociogram) produced by SocNetV, it was represented by a directed line (arc) between player n_i and player n_j.

The sum of the elements of each row of the adjacency matrix $\sum_{\substack{i=1 \\ j \neq i}}^{n} a_{ij}, j = 1, \ldots, n$ was the total number of passes from player i to all the other teammates. The sum of all elements $a_{ij}(i \neq j)$ of the adjacency matrix, $= \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} a_{ij}$, is the first metric used in this study, namely, the *total links* (passes) between each team's players. In the corresponding graph, this number is the total lines between all players (in our case arcs since the graph is directed).

Graph density. Because a digraph consists of a finite number of players (denoted by n), in the case of ordered relations, as in the teammate interactions, the possible directed links in a digraph of n players is $n(n-1)$; the density is computed by

$$\Delta = \frac{L}{n(n-1)} \tag{1}$$

In both cases, the density is a ratio having a minimum of 0 (no lines/arcs present) and a maximum of 1 (all lines/arcs are present).

Degree centrality. A centrality measure for a player in the context of a digraph should be the out-degree of the player, $d_o(n_i)$; thus, it is possible to define $C_D(n_i)$ as a player-level degree centrality index as[34]

$$C_D(n_i) = d_o(n_i) = \sum_{j=1}^{n} x_{ij} = x_{i+} \tag{2}$$

Besides this algorithm, another can be used as a standard measure

$$C'_{(D)}(n_i) = \frac{d_o(n_i)}{(n-1)^2} \tag{3}$$

that is the proportion of players that are adjacent to n_i. $C'_{(D)}(n_i)$ is independent of g and thus can be compared across networks of different sizes.[34] In this analysis, the players with a very high degree of centrality are called hubs since they are connected to many other neighbours.[35]

Degree prestige. The degree prestige measures the in-degree of each player in a digraph context, which can be denoted by $d_I(n_j)$. The degree prestige can be computed as follows[34]

$$P_D(n_i) = d_I(n_i) = x_{+j} \tag{4}$$

In order to standardize the group size g, it is possible to compute as follows

$$P'_D(n_i) = \frac{x_{+j}}{n-1} \tag{5}$$

Betweenness centrality. Betweenness centrality measures the intermediate players between neighbours.[35]

For distinct players $n_i, n_j, n_k \in V(G)$, let g_{ij} be the total number of shortest paths between n_i and n_j and $g_{ij}(n_k)$ be the number of shortest paths from n_i to n_j that pass through n_k. Furthermore, for $n_k \in V(G)$, let $V(n_i)$ denote the set of all ordered pairs, (n_i, n_j) in $V(G)\times V(G)$ such that i, j, k are all distinct.[35] In that sense, the betweenness centrality is calculated as

$$C_b(n_k) = \sum_{(n_i, n_j)\in V(n_k)} \frac{g_{ij}(n_k)}{g_{ij}} \tag{6}$$

In the case of weighted digraphs, we standardize it just like the other actor centrality indices[34]

$$C'_b(n_k) = \frac{C_b(n_k)}{(n-1)(n-2)} \tag{7}$$

Reliability of the software

In order to ensure the reliability of the data collecting process, a small case study was conducted with 10 expert soccer coaches (more than 5 years of experience in elite soccer) using a test–retest protocol with a 10-day interval protocol. Each coach used the PATO to perform the observation, and the video was observed in the same room (with an individual computer with the video of match) and in the same day and time. Such protocol allowed ensuring the same conditions to every coach. Cohen's Kappa test was used to evaluate the intra- and inter-reliability of the data collected by coaches. In the case of intra-reliability (the same coach measured in two occasions with a 10-day interval), the Kappa value was 0.87. On the other hand, the inter-reliability tested between coaches revealed a value of 0.76. Both results suggested a strong value that confirms the reliability of the data collected by using PATO.

Comparison with other sports software

To test the accuracy, reliability and time consumption of the software to perform notational analysis with other sports software, an r-Pearson correlation test was performed. A single match was used for this analysis. The same observer recorded data in PATO and in the *LongoMatch* software. For the *LongoMatch*, a specific PATO code was generated to collect the data in similar conditions to PATO. The results showed that to perform the analysis in post-match mode, the use of PATO consumed 94 min and the *LongoMatch* consumed 112 min (19 more, 15% of the time). The r-Pearson correlation test to compare the sequence of the units of attack (112 in total) showed a positive, nearly perfect value of 0.971. For that reason, the inter-software reliability of the data was achieved in both systems.

Statistical procedures

The influences of phase of competition (group matches and elimination rounds) and positional roles (goalkeeper, external defenders, central defenders, midfielders, external midfielders and forwards) on the total links, graph density, degree centrality, degree prestige and betweenness centrality were analyzed using two-way multivariate analysis of variance (MANOVA) after validating normality and homogeneity assumptions. MANOVA was specifically chosen because it reduces Type I error inflation compared with analysis of variance (ANOVA).[36] Moreover, in many cases, MANOVA can detect statistical differences that many one-way ANOVAs cannot.[37] The assumption of normality for each univariate dependent variable was examined using Kolmogorov–Smirnov tests ($p > 0.05$). The assumption of the homogeneity of each group's variance/covariance matrix was examined with the Box's M test. No homogeneity was shown. When the MANOVA detected significant statistical differences between the two factors, we proceeded to the two-way ANOVA for each dependent variable, followed by Tukey's honest significant difference (HSD) post hoc test.[36] When the two-way ANOVA showed an interaction between factors, it also generated a new variable that crossed the two factors (example: group matches ×

goalkeeper) for each dependent variable to identify statistical significance.[38] Ultimately, the statistical procedures used were one-way ANOVA and Tukey's HSD post hoc. If no interactions were detected in two-way ANOVA, a one-way ANOVA was used for each independent variable. The following scale was used to classify the effect size (partial η^2) of the test:[39] small, 0.14–0.36; moderate, 0.37–0.50 and large, 0.51–1. All statistical analyses were performed using IBM SPSS Statistics (version 22) at a significance level of $p < 0.05$.

Results

The network metrics based on graph and digraph theory allowed obtaining the values of specific kind of interactions between teammates. This is the main core of network. Nevertheless, and in order to improve the knowledge about the variance between different variables, an inferential statistical analysis was made and it is described as follows. The two-way MANOVA revealed that the strategic position ($\gamma = 1.347$; $F = 10.599$; $p = 0.001$; $\eta_p^2 = 0.449$; moderate effect size) had significant main effects on the centrality measures. No statistical differences were found in the phase of competition ($\gamma = 0.004$; $F = 0.087$; $p = 0.967$; $\eta_p^2 = 0.004$; very small effect size). No significant interaction ($\gamma = 0.284$; $F = 1.362$; $p = 0.169$; $\eta_p^2 = 0.0952$; very small effect size) was found between phase of competition and strategic position on centrality measures. Therefore, one-way ANOVA was performed on each independent variable because no interaction was found among factors.

No statistical differences in the phase of competition (Table 1) were observed in %DPrestige ($F = 0.161$; $p = 0.690$; $\eta_p^2 = 0.002$; very small effect size), %DCentrality ($F = 0.197$; $p = 0.658$; $\eta_p^2 = 0.003$; very small effect size) and %BetweennessC ($F = 0.2067$; $p = 0.652$; $\eta_p^2 = 0.003$; very small effect size).

Table 1. One-way ANOVA values in phase of competition in %DPrestige and %DCentrality.

		Mean (SD)	F	p	η_p^2
%DPrestige					
Group matches	Goalkeeper	2.86 (0.39)	0.161	0.690	0.002
	External defenders	7.99 (1.59)			
	Central defenders	9.89 (3.35)			
	Central midfielders	13.51 (2.41)			
	External midfielders	7.75 (1.11)			
	Forwards	5.32 (1.42)			
Elimination matches	Goalkeeper	3.88 (1.54)			
	External defenders	7.62 (3.56)			
	Central defenders	9.47 (1.81)			
	Central Midfielders	12.32 (2.09)			
	External midfielders	9.62 (1.54)			
	Forwards	5.75 (1.88)			
%DCentrality					
Group matches	Goalkeeper	4.95 (0.35)	0.197	0.658	0.003
	External defenders	7.98 (1.25)			
	Central defenders	11.25 (3.83)			
	Central midfielders	13.65 (2.75)			
	External midfielders	6.03 (0.48)			
	Forwards	3.60 (0.72)			
Elimination matches	Goalkeeper	6.01 (1.68)			
	External defenders	9.62 (3.29)			
	Central defenders	10.71 (2.12)			
	Central midfielders	11.79 (2.83)			
	External midfielders	7.05 (1.66)			
	Forwards	3.89 (1.36)			
%BetweennessC					
Group matches	Goalkeeper	1.46 (0.81)	0.206	0.652	0.003
	External defenders	7.72 (5.35)			
	Central defenders	10.86 (5.04)			
	Central midfielders	15.86 (7.08)			
	External midfielders	4.75 (1.88)			
	Forwards	4.31 (1.96)			
Elimination matches	Goalkeeper	3.48 (1.46)			
	External defenders	10.03 (5.63)			
	Central defenders	11.12 (3.23)			
	Central midfielders	11.74 (4.43)			
	External midfielders	7.42 (2.74)			
	Forwards	4.19 (2.09)			

%DPrestige: % degree prestige; SD: standard deviation; %DCentrality: % degree centrality; %BetweennessC: % betweenness centrality.

Table 2. One-way ANOVA values in strategic position in %DPrestige and %DCentrality.

	GK	ED	CD	MF	EMF	FW
%DPrestige	3.45 (1.24)[a,b,c,d]	7.78 (2.80)[c,e]	9.65 (2.47)[c,e,f]	12.83 (2.26)[b,c,d,e,f]	8.82 (1.64)[c,e,f]	5.57 (3.58)[b,c,d]
%DCentrality	5.55 (1.33)[b,c]	8.91 (2.67)[c,f]	10.94 (2.85)[d,e,f]	12.58 (2.89)[a,d,e,f]	6.61 (1.36)[b,c]	3.76 (1.06)[a,b,c]
%BetweennessC	2.61 (1.56)[a,b,c]	9.03 (5.43)[c,e]	11.01 (3.93)[e,f]	13.50 (5.94)[a,d,e,f]	6.28 (2.70)[c]	4.24 (1.86)[b,c]

GK: goalkeeper; ED: external defender; CD: central defender; MF: midfielder; EMF: external midfielder; FW: forward; %DPrestige: % degree prestige; %DCentrality: % degree centrality; % BetweennessC: % betweenness centrality.
Significantly different compared with [a]external defenders, [b]central defenders, [c]central midfielders, [d]external midfielders, [e]goalkeeper and [f]forwards at $p < 0.05$.

Table 3. One-way ANOVA values in phase of competition in total links and graph density.

	Group matches	Elimination matches
Total links	88.00 (1.00)	90.25 (3.77)
Graph density	0.80 (0.01)	0.82 (0.03)

Statistical differences in the strategic position were found in %DPrestige ($F = 26.083$; $p = 0.001$; $\eta_p^2 = 0.667$; large effect size), %DCentrality ($F = 24.243$; $p = 0.001$; $\eta_p^2 = 0.651$; large effect size) and %BetweennessC ($F = 11.058$; $p = 0.001$; $\eta_p^2 = 0.460$; moderate effect size). The post hoc values can be found in Table 2.

No statistical differences in the phase of competition were found in total links ($F = 0.970$; $p = 0.370$; $\eta_p^2 = 0.162$; small effect size) and density ($F = 0.970$; $p = 0.370$; $\eta_p^2 = 0.162$; small effect size). The descriptive statistics can be found in Table 3.

Discussion

This study had two aims. The first was to introduce new software to process the match analysis based on a network approach. The second was to process the data extracted from the software and to compute the network metrics in a specific case study of the German national soccer team in the FIFA World Cup 2014. Two metrics to characterize the digraph properties and two metrics to identify the centrality levels of players were used.

In this study, it was possible to observe that the degree prestige and degree centrality revealed the prominent players in building attacks. Midfielders had the biggest values of degree prestige (in-degree), thus suggesting that these players are the most recruited by teammates to circulate the ball and attack. These findings are in line with previous studies that analyzed all units of attack.[25,28] In particular, cases of analysis centered on quickly attacking transitions or counter-attacks; the most involved players are the wide midfielders or the forwards.[32] Nevertheless, the German national soccer team during the FIFA World Cup 2014 showed a pattern of attack based on ball circulation, thus increasing the intervention of midfielders that links the defenders and forwards. On the other hand, the smallest values, excluding the goalkeeper, were achieved by the German forwards. Once again, these values can be justified by the style of play of Germany that does not exclusively depend on forwards to conclude the attacks. Similar results were found in previous FIFA World Cup tournaments, particularly in the Spanish team.[22,25]

In the case of degree centrality (out-degree), it was possible to observe that midfielders and central defenders achieved the biggest values. These values are also in line with previous studies.[25,27] The degree centrality measures the levels of passes made by players to their teammates. In this case, the central defenders followed the midfielders in the greatest number of passes made by teammates, thus suggesting a pattern based on ball circulation by the center. In other studies, the greatest values followed by midfielders were achieved by wide defenders based on the specific style of play of the teams to build the attacks from the wings.[28]

It was also found in this study that the German national soccer team did not change the recruitment of other positional roles between the group matches and elimination matches. In some cases,[40] after scoring, the teams in elimination matches increase their defensive behavior and change the style of play for a more direct one to exploit the counter-attack. Such a pattern can be represented in the direction of passes by greatest values achieved by forwards in-degree prestige. In the German national soccer team, such changes did not happen, suggesting that the team maintained a consistent style of play of building attack in all types of competition.

In the case of digraph properties, the values did not statistically change between the group matches and the elimination matches. From the highest values of density and total links, it was also found in this study that the German national soccer team did not change the recruitment of other positional roles between the group matches and elimination matches. In some cases,[2] after scoring, the teams in elimination matches increase their defensive behavior and change the style of play for a more direct one to exploit the counter-attack. Such a pattern can be represented in the digraph by greatest values achieved by forwards in-degree prestige. In the German national soccer team, such changes did not happen, suggesting that the team maintained a consistent style of play of positional attack in all types of competition, proof that the style of play is based on positional attack and not in counter-attack. In the cases of teams that use a more direct style of play, the density is smallest.[27] The greatest values achieved by the German national soccer team in this case of digraph properties are also in line with the literature that revealed that the biggest values of density and smallest values of heterogeneity led to the best performance.[23]

The usefulness of the network approach to identify the prominent players and to identify the interaction properties of teammates is undeniable. Moreover, it is important to consider that individual and general properties of network can be analyzed, allowing to identify some patterns of interactions and then contribute for a better understanding of the team's dynamic. Thus, network approach is much more than only to study passes. The software may be used to identify how the marking relationship between defenders and attackers happens or even to identify the positional switching among teammates. Nevertheless, until now there was no specific software for applying such an approach in team sports. With the PATO software, it is possible to quickly codify the interactions and immediately extract the adjacency matrices to import into regular social network analysis software. This advantage reduces the time spent on a paper-and-pencil approach or even Excel solutions. Moreover, by combining the PATO software with SocNetV, it is very user-friendly to process all the data and extract the main properties of the team's digraph and the centrality and prominence levels of team players.

Nevertheless, a major concern is the way in which coaches may receive the feedback. The software is a user-friendly solution to data collection. Although they are utility, it is important to add new functions to produce small reports about the statistical data and to convert such reports into data interpretation in order to increase the coaches' perception about the analysis that was made. Such solution will be integrated in next advances of the software to collect the feedback of coaches about the applicability of the software.

This study had as major limitations the non-comparison of software with other similar software. Nevertheless, the similar are too expensive and not generalizable to the team sports community. Another limitation of our study was the use of a simple case study with only one team. Nevertheless, the main aim was to introduce the software and its applicability for team sports analysis. As a user-friendly tool, the next step will be to convert the computer software for smartphones and tablet applications in order to extend the possibilities to the whole team sports community. Using the network approach, it will be possible to extract the patterns of interactions among teammates and to identify the specific styles of play of teams. It will also be possible to identify the prominent players and their influence during the collective organization making it possible to increase the strategic and tactical information for coaches and their staff.

Practical application and future

In the future, the software will provide updates through a simple visualization of networks based on the level of interactions and inter-dependency. For that reason, the edges and the nodes will have different sizes based on the weight of interactions. If each coach chooses only one player, the connections of that player will only appear between the teammates that cooperated with him. Moreover, the PATO will make it possible to observe the dependency level between the teammates and the players that received more connections to him. During this centralization analysis, it will also be possible to identify the sub-groups inside a team to predict the most common passing sequences in a given context. The user will have the possibility to identify the specific moments (e.g. building attack with ball circulation, counter-attack or quick transition), and based on such organization, the more frequent passing sequences per each type of attack will appear. This identification will help to characterize the style of play and provide such information to avoid or exploit the evidence.

Conclusion

This study introduced the PATO and its compatibility with SocNetV. The PATO software allows the user to quickly codify the interactions between teammates and immediately export the data as adjacency matrices to be imported into SocNetV or similar. The network approach enabled the identification of the prominent players in the German national soccer team during the FIFA World Cup 2014. Using these data, it was possible to characterize that the German national soccer team based their attacking process on building (positional attack) and not on counter-attack. Midfielders had the biggest values of degree prestige and degree centrality, followed by central defenders, revealing the tendency to build the attack using small passes and linking all the sectors. The digraph properties showed great values of density and total links, revealing the capacity of the German national soccer team to involve all the players during the attacking process.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was carried out in the scope of R&D Unit 50008, financed by UID/EEA/50008/2013.

References

1. Gréhaigne JF, Bouthier D and David B. Dynamic-system analysis of opponent relationship in collective actions in football. *J Sport Sci* 1997; 15(2): 137–149.
2. Carling C, Williams AM and Reilly T. *Handbook of soccer match analysis: a systematic approach to improving performance*. London and New York: Taylor & Francis Group, 2005.
3. Sarmento H, Marcelino R, Anguera MT, et al. Match analysis in football: a systematic review. *J Sport Sci* 2014; 32(20): 1831–1843.
4. Clemente FM, Couceiro MS, Martins FML, et al. Practical implementation of computational tactical metrics for the football game: towards an augmenting perception of coaches and sport analysts. In: Murgante B, Misra S, Rocha AMAC, et al. (eds) *Computational science and its applications*. Berlin: Springer, 2014, pp.712–727.
5. Franks IM and McGarry T. The science of match analysis. In: Reilly T (ed.) *Science and football*. Oxon: Spon Press, Taylor & Francis Group, 1996, pp.363–375.
6. Barreira D, Garganta J, Guimarães P, et al. Ball recovery patterns as a performance indicator in elite soccer. *Proc IMechE, Part P: J Sports Engineering and Technology* 2014; 228(1): 61–72.
7. Costa IT, Garganta J, Greco PJ, et al. Influence of relative age effects and quality of tactical behaviour in the performance of youth football players. *Int J Perform Anal Sport* 2010; 10(2): 82–97.
8. Lago-Peñas C and Dellal A. Ball possession strategies in elite soccer according to the evolution of the match-score: the influence of situational variables. *J Hum Kinet* 2010; 25: 93–100.
9. Clemente FM, Couceiro MS, Fernando ML, et al. Measuring tactical behaviour using technological metrics: case study of a football game. *Int J Sport Sci Coach* 2013; 8(4): 723–739.
10. Carling C, Bloomfield J, Nelsen L, et al. The role of motion analysis in elite soccer. *Sport Med* 2008; 38(10): 839–862.
11. Costa IT, Garganta J, Greco PJ, et al. Sistema de avaliação táctica no Futebol (FUT-SAT): desenvolvimento e validação preliminar. *Motricidade* 2011; 7(1): 69–84.
12. Barreira D, Garganta J, Castellano J, et al. SoccerEye: a software solution to observe and record behaviours in sport settings. *Open Sports Sci J* 2013; 6: 47–55.
13. Rampinini E, Alberti G, Fiorenza M, et al. Accuracy of GPS devices for measuring high-intensity running in field-based team sports. *Int J Sports Med* 2015; 36: 49–53.
14. Frencken WGP, Lemmink KAPM and Delleman NJ. Soccer-specific accuracy and validity of the local position measurement (LPM) system. *J Sci Med Sport* 2010; 13(6): 641–645.
15. Couceiro MS, Clemente FM and Martins FML. Analysis of football player's motion in view of fractional calculus. *Cent Eur J Phys* 2013; 11(6): 714–723.
16. Travassos B, Davids K, Araújo D, et al. Performance analysis in team sports: advances from an Ecological Dynamics approach. *Int J Perform Anal Sport* 2013; 13(1): 83–95.
17. Clemente FM, Couceiro MS, Martins FML, et al. Intelligent systems for analyzing soccer games: the weighted centroid. *Ing Investig* 2014; 34(3): 70–75.
18. Clemente FM, Couceiro MS, Martins FML, et al. An online tactical metrics applied to football game. *Res J Appl Sci Eng Tech* 2013; 5(5): 1700–1719.
19. Clemente FM, Martins FML, Couceiro MS, et al. Developing a football tactical metric to estimate the sectorial lines: a case study. In: Murgante B, Misra S, Rocha AMAC, et al. (eds) *Computational science and its applications*. Berlin: Springer, 2014, pp.743–753.
20. Clemente FM, Couceiro MS, Martins FML, et al. Soccer team's tactical behaviour: measuring territorial domain. *Proc IMechE, Part P: J Sports Engineering and Technology* 2015; 229(1): 58–66.
21. Clemente FM, Martins FML, Couceiro MS, et al. Inspecting teammates' coverage during attacking plays in a football game: a case study. *Int J Perform Anal Sport* 2014; 14(2): 1–27.
22. Duch J, Waitzman JS and Amaral LA. Quantifying the performance of individual players in a team activity. *PLoS ONE* 2010; 5(6): e10937.
23. Grund TU. Network structure and team performance: the case of English Premier League soccer teams. *Soc Networks* 2012; 34(4): 682–690.
24. Bourbousson J, Poizat G, Saury J, et al. Team coordination in basketball: description of the cognitive connections among teammates. *J Appl Sport Psychol* 2010; 22(2): 150–166.
25. Peña JL and Touchette H. A network theory analysis of football strategies, http://arxiv.org/abs/1206.6904
26. Cotta C, Mora AM, Merelo JJ, et al. A network analysis of the 2010 FIFA world cup champion team play. *J Syst Sci Complex* 2013; 26(1): 21–42.
27. Clemente FM, Martins FML, Couceiro MS, et al. A network approach to characterize the teammates' interactions on football: a single match analysis. *Cuad Psicol del Deport* 2014; 14(3): 141–148.
28. Clemente FM, Couceiro MS, Martins FML, et al. Using network metrics to investigate football team players' connections: a pilot study. *Motriz* 2014; 20(3): 262–271.
29. Passos P, Davids K, Araújo D, et al. Networks as a novel tool for studying team ball sports as complex social systems. *J Sci Med Sport* 2011; 14(2): 170–176.
30. Duarte R, Araújo D, Correia V, et al. Sports teams as superorganisms: implications of sociobiological models of behaviour for research and practice in team sports performance analysis. *Sports Med* 2012; 42(8): 633–642.
31. Robinson G and O'Donoghue P. A weighted kappa statistic for reliability testing in performance analysis of sport. *Int J Perform Anal Sport* 2007; 7(1): 12–19.
32. Malta P and Travassos B. Characterization of the defense-attack transition of a soccer team. *Motricidade* 2014; 10(1): 27–37.
33. Kalamaras D. *Social Networks Visualizer (SocNetV): social network analysis and visualization software*. Social Networks Visualizer, 2014, http://socnetv.sourceforge.net
34. Wasserman S and Faust K. *Social network analysis: methods and applications*. New York: Cambridge University Press, 1994.
35. Pavlopoulos GA, Secrier M, Moschopoulos CN, et al. Using graph theory to analyze biological networks. *BioData Min* 2011; 4(1): 10.
36. O'Donoghue P. *Statistics for sport and exercise studies. An introduction*. London and New York: Routledge, Taylor & Francis Group, 2012.

37. Pallant J. *SPSS survival manual: a step by step guide to data analysis using the SPSS program*. Crows Nest, NSW, Australia: Allen & Unwin, 2011.
38. Maroco J. *Análise Estatística com utilização do SPSS* [Statistical analysis with SPSS]. Lisbon: Edições Silabo, 2012.
39. Pierce CA, Block RA and Aguinis H. Cautionary note on reporting Eta-squared values from multifactor ANOVA designs. *Educ Psychol Meas* 2004; 64(6): 916–924.
40. Gréhaigne JF. *L'organisation du jeu en football*. Joinville-le-Pont: Éditions Actio, 1992.