
International Journal of Performance Analysis in Sport

ISSN: 2474-8668 (Print) 1474-8185 (Online) Journal homepage: https://www.tandfonline.com/loi/rpan20

# Midfielder as the prominent participant in the building attack: A network analysis of national teams in FIFA World Cup 2014

**Filipe Manuel Clemente, Fernando Manuel Lourenço Martins, P. Del Wong, Dimitris Kalamaras & Rui Sousa Mendes**

**To cite this article:** Filipe Manuel Clemente, Fernando Manuel Lourenço Martins, P. Del Wong, Dimitris Kalamaras & Rui Sousa Mendes (2015) Midfielder as the prominent participant in the building attack: A network analysis of national teams in FIFA World Cup 2014, International Journal of Performance Analysis in Sport, 15:2, 704-722, DOI: 10.1080/24748668.2015.11868825

**To link to this article:** https://doi.org/10.1080/24748668.2015.11868825

Published online: 03 Apr 2017.

International Journal of Performance Analysis in Sport
2015, 15, 704-722.

# Midfielder as the prominent participant in the building attack: A network analysis of national teams in FIFA World Cup 2014

Filipe Manuel Clemente[1,2,3*], Fernando Manuel Lourenço Martins[1,2,4], Del P. Wong[5], Dimitris Kalamaras[6] and Rui Sousa Mendes[1,2]

[1]Polytechnic Institute of Coimbra, ESEC, Department of Education, Portugal
[2]Polytechnic Institute of Coimbra, RoboCorp, ASSERT, Portugal
[3]Faculty of Sport Sciences and Physical Education, University of Coimbra, Portugal
[4]Instituto de Telecomunicações, Delegação da Covilhã, Portugal
[5]Human Performance Laboratory, Technological and Higher Education Institute of Hong Kong, Hong Kong, China
[6]New Media Network Synapsis SA

**Abstract**

***This study aimed to analyze the most prominent players' positions that contributed to the build of attack in football during FIFA World Cup 2014. The connections among teammates in all matches of the tournament were analyzed, and the tactical lineup and players' positions of players were codified as independent variables. Four centrality network metrics were used to identify the pertinence of each players' position. A total of 37,864 passes between teammates were recorded. Each national team was analyzed in terms of all their matches, thus all 64 matches from the FIFA World Cup 2014 tournament were analyzed and codified in this study. A total of 128 adjacency matrices and corresponding network graphs were generated and used to compute the centrality metrics. Results revealed that the players' position (p = 0.001; $\eta_p^2$= 0.143; Power = 1.00; moderate effect size) showed significant main effects on centrality measures. The central midfielders possessed the main values in all centrality measures in the majority of analyzed tactical lineups. Therefore, this study showed that independent of the team strategy, the players' position of a central midfielder significantly contributed to the build of attack in football, for example, greater cooperation and activity profile.***

**Keywords:** Match Analysis, Football, Social Network Analysis, Graph Theory, Adjacency Matrix, Tactics.

## 1. Introduction

Numerous analyses of the structural properties of interactions among teammates have been conducted to identify whether these properties contribute to team performance outcomes (Cotta, Mora, Merelo, & Merelo-Molina, 2013). To achieve this objective, some techniques and methods from social and exact sciences have been used in the specific field of sport sciences to increase the possibility of match analysis (Duarte, Araújo, Correia, & Davids, 2012). Such a multidisciplinary approach enables the properties of a team to be rapidly identified and the individual and collective behavior of players to be characterized (Couceiro, Clemente, Martins, & Tenreiro Machado, 2014).

Several approaches are used to identify and classify teams and their properties. One approach is using the traditional notational analysis of singular events (such as passes, recovers, or shots) that quantify the actions of players and provide some information on the type of team performance (Hughes & Bartlett, 2002). Such approach has become rarely used to facilitate a scientific understanding of team performance, even among coaches (Vilar, Araújo, Davids, & Bar-Yam, 2013). Another approach is the observational method based on semi-computational systems that identifies the tactical behavior of players on the basis of the fundamental principles of play (Barreira, Garganta, Guimarães, Machado, & Anguera, 2014). Even if qualitative analysis is performed by observers, a great deal of time is spent to accomplish such massive analysis. Alternatives to the observational method are tactical metrics, which quantify some spatiotemporal relationships of teammates using computational methods (Clemente, Couceiro, Martins, Mendes, & Figueiredo, 2013). Such method enables the rapid identification of the collective organization of a team. However, this approach does not enable the characterization of the interconnectivity among teammates because the majority of metrics do not consider individual action with a ball. Another alternative is the social network approach that uses the graph theory and provides some mixed information between individual action and tactical behavior (Passos et al., 2011). Moreover, the social network approach is a mixed method because the observational method is first used, followed by the computational method to classify the properties of a network graph (Duch, Waitzman, & Amaral, 2010).

Some studies have used social network analysis to characterize the connections among players within a team (Bourbousson, Poizat, Saury, & Seve, 2010; Duch et al., 2010; Grund, 2012). These studies can be categorized as a network analysis of the global structure of the team and that of the individual connections among players (Clemente, Couceiro, Martins, & Mendes, 2014). In football, one study of the general analysis to the network graph of teams was conducted using 760 matches from the English Premier League (Grund, 2012). Using network metrics, high overall density levels of the network graph were found to enhance team performance. By contrast, high-centralized interaction degraded team performance.

Another study that analyzed the structure of a team investigated the number of passes per minute and the clustering coefficient of teams in FIFA World Cup 2010 (Cotta et al., 2013). This study concluded that classifying some styles of play was possible.

In the individual analysis of players, one of the first studies that used the graph theory was applied in the European Cup 2008 tournament. The study aimed to quantify the contributions of each player and the overall team performance (Duch et al., 2010). That study used a network approach to identify the best individual performance of players and the best team performance.

Another work studied the FIFA World Cup 2006 Finals and an international "A" match in Japan (Yamamoto & Yokoyama, 2011). This study analyzed the probability distribution that emerged in the passing behavior of players. The main results showed that a player who touched the ball many times frequently changed the player to whom he was connected by passes.

Despite these important studies, an in-depth research on the network analysis of the individual classification of players and their contribution to the team process remains lacking. In fact, the information on how players connect with one another and build the attacking process is decreased. Such information is mainly considered in a tactical point of view. Moreover, studies that used the centrality metrics (such as between centralities and in-degree and out-degree network metrics) to determine the prominent players who contribute to the overall structure of a team's network graph are scarce (Malta & Travassos, 2014; Peña & Touchette, 2012). The studies that used a closer approach only focused on the analysis in a small sample, such as case studies, and did not differentiate the *players'* position that contributes most in building the attacking process in football.

Based on the above-mentioned reasons, the present study aimed to identify the connections among teammates in all matches of FIFA World Cup 2014 and analyze the variance between *players'* positions and the tactical lineup of national teams in their levels of in-degree, out-degree, closeness, and betweenness centralities. This study also aimed to identify the most prominent players and *players'* positions that contribute to the building attack in national teams.

## 2. Methods

### 2.1. Sample

A total of 64 matches from the FIFA World Cup 2014 tournament were analyzed and codified in this study. A total of 37,864 passes between teammates were recorded and processed. Each national team was analyzed in terms of all their matches. Thus, a total of 128 adjacency matrices and corresponding network graphs were generated and used to compute the centrality metrics.

### 2.2. Data Collecting and Processing

All matches in FIFA World Cup 2014 were examined. The players of national teams were codified by their *players'* position on the basis of the tactical lineup of each national team. The tactical lineup of each team was classified by three football coaches with more than five years of experience. During the matches, the tactical lineup of some teams changed, and in these situations, the national teams were classified on the basis of tactical lineup on which each team spent more time. To guarantee the reliability of classification, the same three coaches classified the teams in two occasions (during and one month after the FIFA tournament). These two classifications were tested by Cohen's Kappa test, adhering to a 30 day interval for re-analysis to avoid task familiarity issues (Robinson & O'Donoghue, 2007). A Kappa value of 0.92 was obtained after testing the full data (tactical lineup), thus ensuring a recommended margin for this type of procedures (Robinson & O'Donoghue, 2007). Based on the global observation of the national teams, the tactical lineup variable was generalized in four factors: (i) 1-4-3-3, (ii) 1-4-2-3-1, (iii) 1-4-4-2, and (iv) 1-3-5-2.

The tactical lineup was used in codifying the position of each player. A techno-tactical assignment was adopted to positional roles (Di Salvo et al., 2007), and the tactical position of goalkeeper and striker was added. The tactical assignment can be verified in Figure 1.

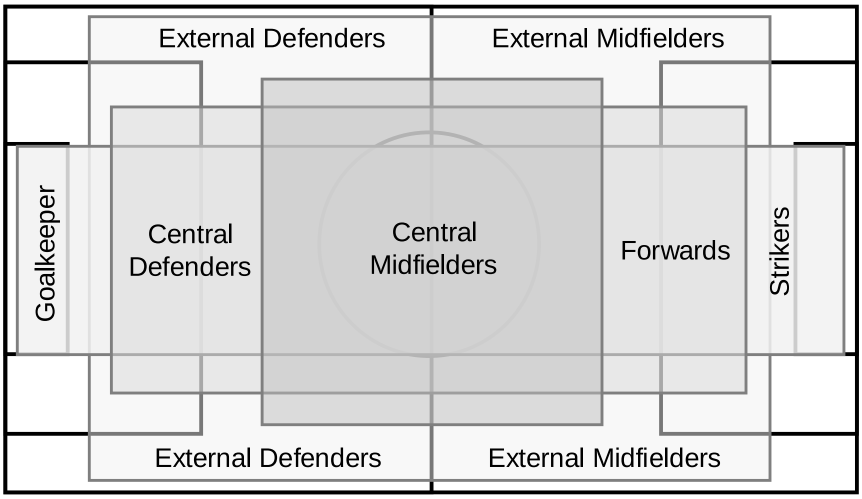


Figure 1. *players'*positions codified based on match analysis.

To study the connections between *players'* positions, the linkage indicator of ball passes between teammates was defined. Thus, during all matches, all attacking instants with more than one pass were observed. Each sequence of passes was classified as a unit of attack. An attacking unit started at the moment that a team player made a successful pass to a teammate and finished when the team lost possession of the ball (e.g., ball out of boundaries, ball out of shot, and unsuccessful pass to a teammate). An adjacency matrix was generated per unit of attack. This matrix represents the connections between a node (player) and an adjacency node (teammate) (Passos et al., 2011). In the adjacency matrix, each pass between nodes was codified as 1 (one), and no passes between teammates were codified as 0 (zero). More than one pass between the same nodes were codified with the number of passes. Each player was classified with a number between P1 and P11 for easy codification. Figure 2 shows an example of all steps performed in data collecting and processing.

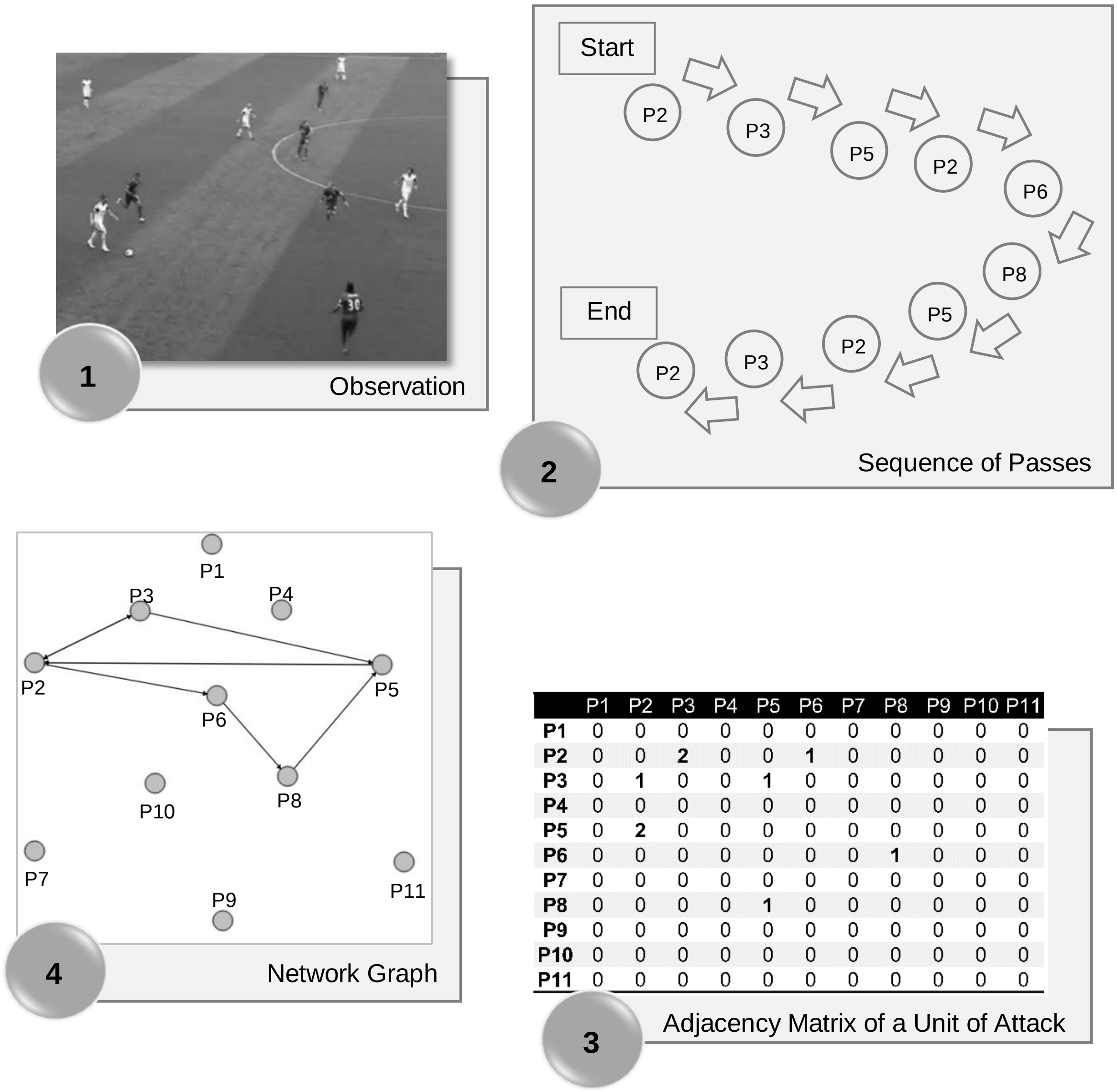


Figure 2. Sequence of data collecting and processing.

The procedure described in Figure 2 only represents one unit of attack. Nevertheless, the final adjacency matrix that comprised the sum of all adjacency matrices from all units of attack was computed at the end of each match. The network metrics were computed on the basis of the final adjacency matrix, which will be further described in this study. The observation and codification of sequences of passes were performed by the same researcher with more than five years of experience in match analysis to minimize inter-reliability error. The observer was previously trained and was tested in a test-retest procedure to ensure the reliability of data. Cohen’s Kappa test adhered to a 20-day interval for re-analysis to avoid task familiarity issues (Robinson & O'Donoghue, 2007). A Kappa value of 0.76 was obtained after testing 15% of the full data. The Kappa value ensured a recommended margin for this type of procedures (Robinson & O'Donoghue, 2007).

## 2.3. Centrality Measurement

Once created, the adjacency matrices were imported in Social Networks Visualizer (SocNetV), which is an application for the visualization and analysis of social networks (Kalamaras, 2014).

Social network analysts use a variety of measures (metrics or indices) to quantify the prominence or importance of each actor (node) inside a given social network (called "graph" in terms of graph theory). Considering that prominence may have different meanings for different types of network data, social network analysts have proposed and developed specific metrics or variants of older metrics, with each often focusing on different graph notions and thus being suitable for application to special graph types. For instance, the so-called out-degree centrality (ODC) considers the number of outbound edges from each network actor and may be applied to every network type, whereas the closeness centrality (CC) counts the total "distance" from the actor to all others but can be applied only to connected graphs and strongly connected directed graphs.

The data in this study were analyzed using four widely known actor centrality metrics: out-degree, in-degree, closeness, and betweenness. The following sections briefly describe each of these metrics along with their common interpretation and meaning in the context of network analyses.

## 2.4. Out-Degree Centrality

The ODC, also known as "degree centrality" (Nieminen, 1974), of each node $u$ in an unvalued or directed graph G (V,E), with a set of nodes V and a set of edges E, is simply the count of outbound edges (arcs) from that node to all nodes that are connected to it (neighbors). This count is called the "out-degree" of a node.

$$ODC_u = \sum_{v=1,u\neq v}^{g} e_{uv} \text{ where } e_{uv} = 1 \text{ if } e_{uv} \in E, \text{ otherwise } e_{uv} = 0 \tag{1}$$

Apparently, the theoretical maximum value of $ODC_u$ in a non-valued graph is $(g-1)$, where $g = |V|$. That is, a node will have an absolutely maximum ODC score if it is outbound connected to every other node in the network. Thus, a standardized ODC index can be computed as follows:

$$ODC'_u = \frac{\sum_{v=1,u\neq v}^{g} e_{uv}}{g-1} \tag{2}$$

In valued graphs and digraphs, that is, graphs where each edge has a value or weight, the ODC of each node is computed by summing the weights of all arcs from that node to its neighbors.

$$ODC_u = \sum_{v=1,u\neq v}^{g} a_{uv} \tag{3}$$

where $a_{uv}$ is the weight of $e_{uv} \in E$.

Given that valued graphs can have edges with arbitrarily low or high weights, a theoretical maximum ODC cannot be computed in those cases.

For both valued and unvalued graphs, the ODC score of a node can be easily computed by summing the elements of the corresponding row of the adjacency matrix.

$$ODC_u = \sum_{v=1, u \neq v}^{g} A(u, v) \tag{4}$$

where $A(u, v)$ is $(u, v)$ element of the adjacency matrix A.

By considering only the number of outbound edges, ODC is a simple metric for computation that is usually interpreted as a measure of the activity of each node. Nodes with higher ODC are connected to more nodes than those with lower ODC. Thus, such nodes are believed to be more important for the overall network structure.

In the context of the soccer in this study, where each edge between nodes signifies the passes between the relevant teammates, the players with larger ODC scores are those who contributed more to their team's offensive attempts through their passes to the other players of their team.

**2.5. In-Degree Centrality**

Whereas the ODC focuses on outbound edges from nodes, the "sister" metric called in-degree centrality (IDC, also known as "degree prestige") considers only the inbound links to each node from other nodes. Thus, the IDC score of each node $u$ is the total number of its inbound edges (arcs) from all neighboring nodes that connect to it

$$IDC_u = \sum_{v=1, u \neq v}^{g} e_{vu} \tag{5}$$

where $e_{vu} \in E$.

Similar to ODC, the theoretical maximum value of $IDC_u$ in a non-valued graph is $(g - 1)$, where $g = |V|$. Thus, a node may have an absolutely maximum IDC score among the network if it is inbound connected from every other node. Again, a standardized IDC index can be computed as follows:

$$ODC'_u = \frac{\sum_{v=1, u \neq v}^{g} e_{uv}}{g - 1} \tag{6}$$

The IDC can be computed for valued graphs and digraphs as well. In these cases, the IDC of each node is the sum of weights of all inbound arcs to that node from its neighbors

$$IDC'_u = \sum_{v=1, u \neq v}^{g} a_{vu} \tag{7}$$

where $a_{vu}$ is the weight of $e_{vu} \in E$.

Similar to ODC, the IDC score of a node can be easily computed by summing the elements of the corresponding column of the adjacency matrix.

Given that the IDC index considers only inbound links, it is often used as indication of the “prestige” of each node among its peers. Nodes with high IDC scores are those that receive many inbound links from other nodes. These links can be interpreted as “choices” or “nominations” to a specific actor from others. Thus, a larger IDC score of an actor indicates that this actor is more prestigious or important among its peers.

Analysis of the data in this study shows that the players with higher IDC scores are obviously those to whom their teammates preferred to pass the ball more often. These players might possibly be the ones crucial for their team’s offensive development because they receive the ball more often than other players during their team’s attempt to attack.

**2.6. Closeness Centrality**

The CC (Freeman, 1979; Sabidussi, 1966) is a more complex index than both the above-mentioned degree centrality metrics. Rather than merely counting edges, this index attempts to quantify actor importance in terms of their total graph theoretic distance in the social network.

In graph theory, the geodesic distance $d$ (or just distance) of two nodes $u, v$ in a non-directed unvalued graph is the length of the shortest path (called geodesic) between them. That is, the distance of two nodes is the minimum number of edges one has to traverse to move from the first node to the second. For instance, if node $u$ is connected with an edge to $v$, which in turn is connected to $w$, then the distance between $u$ and $v$ is $d(u, v) = 1$, whereas the distance between $u$ and $w$ is $d(u, w) = 2$. In undirected graphs, the distance of two nodes is reciprocal: if $d(u, w) = x$, then by definition, $d(w, u) = w$ as well.

Apparently, graphs can have multiple geodesics that connect a pair of nodes with the same length. If no path connects two nodes, then these nodes are called not connected or not reachable, and their distance is infinite: $d(u, v) = \infty$.

In directed graphs, the distance $d$ of two nodes $u, v$ is the length of a shortest path starting from node $u$ and ending at node $v$. In this case, their distance is not always reciprocal. Whether $d(u, w) = d(w, u)$ is uncertain because there can be one or more directed geodesics of length $l$ from $u$ to $v$, but an “opposite” geodesic from $v$ to $u$ can go through different intermediate nodes or it cannot exist at all.

The CC of each node in an undirected graph is the inverse sum of its distances to all other nodes:

$$IDC_u = \sum_{v=1,u\neq v}^{g} e_{vu} \tag{8}$$

Apparently, in unvalued graphs and digraphs, the CC index of each node has a maximum value of $1/(g-1)$. Thus, a standardized CC can be computed by the following formula:

$$CC'_u = \frac{g-1}{\sum_{v=1,u\neq v}^{g} d(u,v)} \tag{9}$$

The computation of CC is more complex than that of degree-based centralities because the matrix of distances should be computed first using either Breadth-First Search for unvalued graphs or Dijkstra's algorithm for valued networks. In essence, the CC score of each node in a graph is the sum of the corresponding rows of the distance matrix of the graph.

At this point, the computation of CC in weighted graphs has even more complications because the notion of distance and shortest path should be extended first to accommodate edge weights. The distance of two nodes $u$,$v$ in a valued graph can be computed as the sum of the weights of the edges in a shortest path from $u$ to $v$. However, clarifying what a shortest path is would also be necessary. For instance, if the edge weights denote "cost," then the shortest path could naturally be that of minimum total sum of weights. By contrast, if the edge weights denote "votes" or "nominations", then a shortest path might also have the maximum total sum of weights. In the latter case, social network analysts usually invert the distances before computing distance-based indices. In this study, the CC index (or betweenness centrality [BC] described below) was not extended to consider edge weights. Rather, distances were computed by omitting weights.

Given that the distance of two unconnected nodes is infinite, the CC index cannot be used straightforwardly in not-connected graphs or not-strongly connected digraphs. In not-connected undirected graphs, the CC metric can be calculated either by dropping isolated nodes or by computing a variant of CC, which considers the distance of each node $u$ to all nodes in its influence range $J$ (the set of nodes that are reachable from $u$):

$$CC_u = \frac{|J|/g-1}{\sum_{v\in I,u\neq v} d(u,v)/|J|} \tag{10}$$

where $J$ the set of all nodes reachable from $u$.

The last formula can also compute CC in the case of not-strongly connected directed graphs.

In any case, the CC score of a node quantifies the proximity of how close is such node to its peers. Nodes with higher CC scores can reach more nodes in fewer steps than those with lower CC scores. This metric can also be interpreted as an index of the capability of a node to access or pass information to other nodes in the network.

In the context of the dataset used in this study (passes between teammates in FIFA World Cup 2014), the CC index of each player denotes how close, in terms of passes, that player has been to all other teammates during the development of the team's attack. High CC scores of a player might indicate that this player not only participated in the attacks successfully passing to other players, but was also closer to the final outcome of the attack (i.e., shoot and out).

**2.7. Betweenness Centrality**

The last actor centrality index used in this study is BC (Anthonisse, 1971; Freeman, 1979), which is the most complex yet the most meaningful. Rather than focusing only on the distances between actors, the BC index attempts to measure the extent of the control that each node holds over the network by considering the shortest paths between all pairs of nodes.

The BC index of each node $u$ is computed by counting the relative number of shortest paths between pairs of other nodes that pass through $u$. In graph-theoretic terms, the BC score of a node $u$ is the ratio of all geodesics between pairs of nodes that run through it. Let denote the total number of geodesics between nodes $s$ and $t$ and $\sigma_{st}(u)$ the geodesics that pass through node $u$ (also called the *dependency* of $s, t$ on $u$), then the actor BC index is given by the following formula:

$$BC_u = \sum_{s \neq t \neq u \in V} \frac{\sigma_{st}(u)}{\sigma_{st}} \tag{11}$$

In directed graphs each node $u$ can apparently lie at most at $(g-1)(g-2)$ geodesics, which is the number of pairs of nodes, except for $u$. In undirected graphs, where pair ordering does not matter, $\frac{(g-1)(g-2)}{2}$ pairs exist. Thus, the BC score of each node $u$ can be standardized by dividing by $(g-1)(g-2)$ and $\frac{(g-1)(g-2)}{2}$ for directed and undirected graphs, respectively:

$$BC'_u = \frac{\sum_{s \neq t \neq u \in V} \frac{\sigma_{st}(u)}{\sigma_{st}}}{(g-1)(g-2)} \tag{12}$$

for diagraphs, and

$$BC'_u = \frac{2 \cdot \sum_{s \neq t \neq u \in V} \frac{\sigma_{st}(u)}{\sigma_{st}}}{(g-1)(g-2)} \tag{13}$$

for undirected graphs.

The computation of BC is the most complicated among the centrality metrics used in this study because the shortest paths between every pair of nodes $(s, t)$ should be computed first. Thereafter, the number of paths that pass through $u$ should be counted. Finally, all the fractions for all pairs $(s, t)$ should be summed.

The BC index is often considered the most meaningful measure among other centrality indices because it successfully quantifies how often each node lies between other nodes of the network, perhaps acting as a mediator or "bridge" for them. In essence, the BC score of each node can be explained as a measure of the relative control that node has on other nodes. In social network analyses, nodes with higher BC scores are commonly assumed to have a higher probability to exert control on the information flow between other nodes in the same network.

As regards the passing game data, players with higher BC scores might be those who more often were situated between their teammates. For instance, a player with high BC score could be important in passing the ball to others.

**2.8. Variables of the Study**

This study aimed to analyze the variance of centrality measures between *players'* positions and attempted to identify the most prominent *players'* positions that contribute to building the attack. The *players'* position of players depends on the tactical lineup adopted by each national team. Thus, both variables (*players'* position and tactical lineup of national teams) were used as independent variables in this study. To identify the most prominent *players'* positions in all national teams, the social network analysis approach, particularly the network metrics of centrality, was used with the following considerations: (i) IDC (%IdC—to identify the *players'* position that most players receive the ball from teammates), (ii) ODC (%OdC—to identify the *players'* position that starts the build of attack), (iii) CC (%CC—to identify the simplest manner of reaching a particular player within a team), and (iv) BC (%BC—to identify how the ball flows between other tactical positions depending on a particular *players'* position).

**2.9. Statistical Procedures**

The influences of tactical lineup and *players'* position factors on the %IdC, %OdC, %CC, and %BC were analyzed using two-way MANOVA after validating the normality and homogeneity assumptions. MANOVA was specifically selected because it reduces Type I Error Inflation compared with ANOVA (O'Donoghue, 2012, p. 242; Pallant, 2011, p.283). In many cases, MANOVA can detect statistical differences that many one-way ANOVAs cannot (Maroco, 2011, p. 276; Pallant, 2011, p. 283). The assumption of normality for each univariate-dependent variable was examined using Kolmogorov–Smirnov tests ($p$-value $< 0.05$). The assumption of the homogeneity of the variance/covariance matrix of each group was examined using the Box's M Test (Pallant, 2011). No homogeneity was shown. When the MANOVA detected significant statistical differences between the two factors, the two-way ANOVA was used for each dependent variable, followed by Tukey's HSD post-hoc test (O'Donoghue, 2012). When the two-way ANOVA showed an interaction between factors, it also generated a new variable that crossed the two factors (e.g., 1-4-3-3*Goalkeeper; 1-4-2-3-1*Central Midfielder) for each dependent variable to identify statistical significance (Maroco, 2012). Ultimately, the statistical procedures used were one-way ANOVA and Tukey's HSD post-hoc test. If no interactions were detected in the two-away ANOVA, one-way ANOVA was used for each independent variable. All statistical analyses were performed using IBM SPSS Statistics (version 21) at a significance level of $p < 0.05$.

The following scale was used to classify the effect size and the power of the test (Hopkins, Hopkins, & Glass, 1996): very small, 0–0.01; small, 0.01–0.09; moderate, 0.09–0.25; large, 0.25–0.49; very large, 0.49–0.81; and nearly perfect, 0.81–1.0.

## 3. Results

The two-way MANOVA revealed that the *players'* position ($p$ = 0.001; $\eta_p^2$= 0.143; *Power* = 1.00; moderate effect size) had significant main effects on the centrality measures. No statistical differences were found in the tactical lineup ($p$ = 0.643; $\eta_p^2$= 0.002; *Power* = 1.00; very small effect size). Significant interaction (Pillai's Trace = 0.071; $F_{52,5540}$ = 1.917; $p$ = 0.001; $\eta_p^2$= 0.018; *Power* = 1.00; small effect size) was found between tactical lineup and *players'* position on centrality measures. As previously indicated in the statistical procedures, two-way ANOVA was conducted for each dependent variable after the confirmation of the interaction (O'Donoghue, 2012, p. 243).

Interaction was found between factors (tactical line up and players' positions) for %IdC ($F$ = 3.764; $p$ = 0.001; $\eta_p^2$= 0.034; *Power* = 0.999; small effect size), %OdC ($F$ = 4.322; $p$ = 0.039; $\eta_p^2$= 0.006; *Power* = 1.00; small effect size), and %CC ($F$ = 2.412; $p$ = 0.003; $\eta_p^2$= 0.022; *Power* = 0.978; small effect size). No interaction was found between factors for %BC ($F$ = 1.383; $p$ = 0.160; $\eta_p^2$= 0.013; *Power* = 0.800; small effect size).
One-way ANOVA tested the crossing among factors. Statistical differences were found between the new variable (cross between tactical lineup and *players'* position) and the dependent variables of %IdC ($F$ = 38.983; $p$ = 0.001; $\eta_p^2$= 0.382; *Power* = 1.00; large effect size), %OdC ($F$ = 44.018; $p$ = 0.001; $\eta_p^2$= 0.411; *Power* = 1.00; large effect size), and %CC ($F$ = 28.477; $p$ = 0.001; $\eta_p^2$= 0.311; *Power* = 1.00; large effect size). The post-hoc results are shown in Table 1.

Table 1. Descriptive table (mean and standard deviation) and statistical comparison between crossing factors.

| | Goalkeeper (TP1) | External Defenders (TP2) | Central Defenders (TP3) | Central Midfielders (TP4) | External Midfielders (TP5) | Forwards (TP6) | Striker (TP7) |
|---|---|---|---|---|---|---|---|
| 1-4-3-3 | | | | | | | |
| %IdC | 2.71(0.20) | 8.98(0.29) | 8.67(0.28) | **11.53(0.26)** | 9.57(0.33) | - | 8.26(0.54) |
| %OdC | 4.99(0.26) | 10.95(0.34) | 10.06(0.30) | **11.43(0.34)** | 6.95(0.30) | - | 4.79(0.25) |
| %CC | 8.24(0.13) | 9.22(0.09) | 9.57(0.10) | 9.69(0.09) | 8.53(0.10) | - | 8.05(0.11) |
| 1-4-2-3-1 | | | | | | | |
| %IdC | 2.11(0.18) | 10.12(0.33) | 8.67(0.26) | **11.45(0.28)** | 9.25(0.26) | - | 7.48(0.38) |
| %OdC | 4.38(0.24) | **11.46(0.23)** | 10.01(0.27) | **11.29(0.29)** | 6.95(0.31) | - | 4.92(0.28) |
| %CC | 8.01(0.13) | 9.39(0.07) | 9.35(0.08) | 9.88(0.09) | 8.47(0.09) | - | 7.95(0.13) |
| 1-4-4-2 | | | | | | | |
| %IdC | 2.66(0.38) | 10.57(0.62) | 8.44(0.81) | **12.32(0.77)** | 9.81(0.53) | 7.53(0.66) | - |
| %OdC | 4.41(0.71) | **13.46(0.82)** | 9.84(0.89) | **12.67(0.97)** | 7.20(0.54) | 4.62(0.46) | - |
| %CC | 8.41(0.39) | 9.64(0.29) | 9.62(0.27) | **10.17(0.27)** | 8.50(0.18) | 7.87(0.26) | - |
| 1-3-5-2 | | | | | | | |
| %IdC | 3.78(0.46) | 9.64(0.38) | **9.89(0.49)** | 9.55(0.42) | - | 9.31(0.60) | - |
| %OdC | 5.39(0.50) | 10.46(0.35) | **11.20(0.47)** | 8.87(0.48) | - | 6.72(0.45) | - |
| %CC | 8.05(0.22) | 9.01(0.15) | **9.61(0.13)** | 9.33(0.13) | - | 8.57(0.17) | - |

In the case of %BC, one-way ANOVA was performed on each independent variable because no interaction was found among factors (Table 2). The results for %BC (Table 2) showed statistical differences in the tactical lineup of 1-4-3-3 ($F$ = 35.109; $p$ = 0.001; $\eta_p^2$ = 0.264; *Power* = 1.000; large effect size), 1-4-2-3-1 ($F$ = 51.391; $p$ = 0.001; $\eta_p^2$ = 0.293; *Power* = 1.000; large effect size), 1-4-4-2 ($F$ = 6.650; $p$ = 0.001; $\eta_p^2$ = 0.357; *Power* = 1.000; large effect size), and 1-3-5-2 ($F$ = 10.570; $p$ = 0.001; $\eta_p^2$ = 0.164; *Power* = 1.000; moderate effect size).

Table 2. One-way ANOVA values in *players'* position in each tactical lineup in %BC.

| | | M(SD) | F | $p$ | $\eta_p^2$ | Power |
|---|---|---|---|---|---|---|
| 1-4-3-3 | Goalkeeper | 2.05(0.26)[b,c,d,e,g] | 35.109 | 0.001 | 0.264 | 1.000 |
| | External Defenders | 10.20(0.63)[a,d,e,g] | | | | |
| | Central Defenders | 10.79(0.62)[a,e,g] | | | | |
| | Central Midfielders | 12.39(0.54)[a,b,e,g] | | | | |
| | External Midfielders | 6.59(0.50)[a,b,c,d] | | | | |
| | Forwards | - | | | | |
| | Striker | 5.60(0.61)[a,b,c,d] | | | | |
| 1-4-2-3-1 | Goalkeeper | 1.89(0.21) [b,c,d,e,g] | 51.391 | 0.001 | 0.293 | 1.000 |
| | External Defenders | 10.98(0.56)[a,e,g] | | | | |
| | Central Defenders | 11.02(0.64)[a,e,g] | | | | |
| | Central Midfielders | 12.37(0.47)[a,e,g] | | | | |
| | External Midfielders | 5.82(0.38)[a,b,c,d] | | | | |
| | Forwards | - | | | | |
| | Striker | 5.37(0.44) [a,b,c,d] | | | | |
| 1-4-4-2 | Goalkeeper | 1.37(0.28)[b,c,d] | 6.650 | 0.001 | 0.357 | 1.000 |
| | External Defenders | 11.61(1.50)[a] | | | | |
| | Central Defenders | 10.49(2.09)[a] | | | | |
| | Central Midfielders | 14.40(1.77)[a,e,f] | | | | |
| | External Midfielders | 7.54(1.53)[d] | | | | |
| | Forwards | 5.28(1.25)[d] | | | | |
| | Striker | - | | | | |
| 1-3-5-2 | Goalkeeper | 2.48(0.51)[b,c,d,f] | 10.570 | 0.001 | 0.164 | 1.000 |
| | External Defenders | 9.37(0.89)[a] | | | | |
| | Central Defenders | 11.47(0.78)[a,f] | | | | |
| | Central Midfielders | 9.77(0.81)[a] | | | | |
| | External Midfielders | - | | | | |
| | Forwards | 7.52(0.83)[a,c] | | | | |
| | Striker | - | | | | |

Significantly different compared with Goalkeeper[a], External Defenders[b], Central Defenders[c], Central Midfielders[d], External Midfielders[e], Forwards[f], and Striker[g] at $p < 0.05$.

In the *players'* position analysis, the results for %BC (Table 3) showed statistical differences in the Central Midfielders ($F$ = 3.522; $p$ = 0.015; $\eta_p^2$ = 0.027; *Power* = 0.781; small effect size). No differences were found in Goalkeeper ($F$ = 0.860; $p$ = 0.464; $\eta_p^2$ = 0.020; *Power* = 0.233; small effect size), External Defenders ($F$ = 0.953; $p$ = 0.415; $\eta_p^2$ = 0.011; *Power* = 0.259; small effect size), Central Defenders ($F$ = 0.167; $p$ = 0.918; $\eta_p^2$ = 0.002; *Power* = 0.081; very small effect size), External Midfielders ($F$ = 1.318; $p$ = 0.270; $\eta_p^2$ = 0.012; *Power* = 0.283; small effect size), Forwards ($F$ = 1.835; $p$ = 0.182; $\eta_p^2$ = 0.035; *Power* = 0.264; small effect size), and Striker ($F$ = 0.096; $p$ = 0.757; $\eta_p^2$ = 0.001; *Power* = 0.061; very small effect size).

Table 3. One-way ANOVA values in tactical lineup in each *players'* position in %BC.

| | | M(SD) | F | $p$ | $\eta_p^2$ | Power |
|---|---|---|---|---|---|---|
| Goalkeeper | 1-4-3-3 | 2.05(0.26) | 0.860 | 0.464 | 0.020 | 0.233 |
| | 1-4-2-3-1 | 1.89(0.21) | | | | |
| | 1-4-4-2 | 1.37(0.28) | | | | |
| | 1-3-5-2 | 2.48(0.51) | | | | |
| External Defenders | 1-4-3-3 | 10.20(0.63) | 0.953 | 0.415 | 0.011 | 0.259 |
| | 1-4-2-3-1 | 10.98(0.56) | | | | |
| | 1-4-4-2 | 11.61(1.50) | | | | |
| | 1-3-5-2 | 9.37(0.89) | | | | |
| Central Defenders | 1-4-3-3 | 10.79(0.62) | 0.167 | 0.918 | 0.002 | 0.081 |
| | 1-4-2-3-1 | 11.02(0.64) | | | | |
| | 1-4-4-2 | 10.49(2.09) | | | | |
| | 1-3-5-2 | 11.47(0.78) | | | | |
| Central Midfielders | 1-4-3-3 | 12.39(0.54)$^{d}$ | 3.522 | 0.015 | 0.027 | 0.781 |
| | 1-4-2-3-1 | 12.37(0.47)$^{d}$ | | | | |
| | 1-4-4-2 | 14.40(1.77) | | | | |
| | 1-3-5-2 | 9.77(0.81)$^{a,b}$ | | | | |
| External Midfielders | 1-4-3-3 | 6.60(0.50) | 1.318 | 0.270 | 0.012 | 0.283 |
| | 1-4-2-3-1 | 5.82(0.38) | | | | |
| | 1-4-4-2 | 7.54(1.53) | | | | |
| | 1-3-5-2 | - | | | | |
| Forwards | 1-4-3-3 | - | 1.835 | 0.182 | 0.035 | 0.264 |
| | 1-4-2-3-1 | - | | | | |
| | 1-4-4-2 | 5.28(1.25) | | | | |
| | 1-3-5-2 | 7.52(0.83) | | | | |
| Strikers | 1-4-3-3 | 5.60(0.61) | 0.096 | 0.757 | 0.001 | 0.061 |
| | 1-4-2-3-1 | 5.37(0.44) | | | | |
| | 1-4-4-2 | - | | | | |
| | 1-3-5-2 | - | | | | |

Significantly different compared with 1-4-3-3$^{a}$; 1-4-2-3-1$^{b}$; 1-4-4-2$^{c}$; and 1-3-5-2$^{d}$ at $p<0.05$

## 4. Discussion

Identifying the most prominent players that build the attack of a football team is one of the key indicators of match analysis by the opposing teams. In fact, the strategy of opponents is often to block the prominent player to prevent a successful attack. Moreover, knowledge on the individual contribution of each player for the overall connection of the team can be an important indicator that may increase the possibility of optimizing the tactical behavior of football players. Thus, this study focused on identifying the *players'* positions that contribute more to the attacking process of national teams in FIFA World Cup 2014.

The pass between players is one of the main indicators that determines the connections among teammates in the attacking process. To build a network approach, the players were considered as the node and the pass as the linkage indicator. After collecting such indicators, social network analysis was used to determine the most prominent tactical positions in building the attack on football. During the attacking phase, players who act as a link in the moments between recovering the ball until the proximity to shot. Thus, determining the *players'* positions that most often start the unit of attack is important. The ODC (%OdC) was used in the analysis (Wasserman & Faust, 1994). The results revealed that the highest levels of %OdC were found in the central midfielders, specifically in the 1-4-3-3 and 1-4-2-3-1 tactical lineups. The highest mean values of %OdC were found in external defenders and central defenders in the 1-4-4-2 and 1-3-5-2 tactical lineups, respectively. These results suggest that the central midfielders act as the priority link of the team to build the attack in football. In fact, the central midfielder shows higher centrality based on network results. This result is consistent with that of previous studies on football (Duch et al., 2010; Malta & Travassos, 2014). The results found on 1-3-5-2 can be justified by the higher relevance of the central defender in this tactical lineup. Malta and Travassos (2014) found that the player with the highest %OdC was the defensive midfielder because he links the defense and the middle in the first phase of building attack. In the case of 1-3-5-2, the central defender acts as a defensive midfielder in the attacking moment. Thus, player in this playing position is a relevant player in building the attack, that is, mainly starts the unit of attack with passes for the playmakers and forwards. Excluding the goalkeeper, the forwards showed the lowest levels of %OdC. Once again these results were consistent with a study that analyzed the attacking transition of one team during four matches (Malta & Travassos, 2014). These results can be justified by the tactical role and mainly being the final targets of the players in building the attack. These players were also not active in the majority of the time in building the attack. Nevertheless, it is important to highlight that this may not always occur and depends from contextual variables.

Another important indicator that should be considered in match analysis is the *players'* position in which most players receive passes from their teammates. These *players'* positions can be considered as the final targets of the attacking process. The network metric of IDC (%IdC) was used for such analysis, as suggested by Wasserman and Faust (1994). The results showed that the *players'* position with the highest levels of %IdC was the central midfielder in the tactical lineup with four defenders (1-4-3-3, 1-4-2-3-1, and 1-4-4-2). In the case of 1-3-5-2, the highest mean value was found in the central defender. The results also showed that the forwards and strikers were the *players'* positions with the lowest %OdC in all tactical lineups. These results differ from the findings of Malta and Travassos (2014), who only studied the specific phase of the defense–attack transition. In the present study, all units of attack were investigated. Thus, most of the time the process of building the attack from behind ends in the middle without an active participation of the forward players. Normally, the players in the middle of the field increase their participation at the start of building the attack and in the final phase of the attacking process when they lose the ball. Therefore, these results suggest that in the majority of the tactical lineup, the central midfielder is the prominent *players'* position at the start and end of building the attack. By contrast, the forwards have the lowest participation in building the attack. Finally, the central and external defenders could also contribute to building the attack through passing and receiving the ball, mainly in the moments of the game in which controlling the match with passes between defenders without moving forward is necessary.

Aside from IDC and ODC, CC (%CC) and BC (%BC) of *players'* positions were also studied. They respectively represent the players who showed highest connections with their teammates, and the ball flow between other players depends on that particular *players'* position (Peña & Touchette, 2012). The results of this study showed that the central midfielders had the highest levels of %CC and %BC in all tactical lineups with four defenders, with the exception of 1-3-5-2. In this case, the highest levels were found in central defenders. The highest values found in midfielders are consistent with the study performed in the FIFA World Cup 2010 Finals, despite of no tactical-lineup have been studied (Peña & Touchette, 2012). They found that in the Spanish team, the players with greatest values of closeness were Xavi (central midfielder) and Busquets (defensive midfielder). Meanwhile, in the German team, the highest values of betweenness and closeness were found in Lahm (external defender), Mertesacker (central defender), and Schweinsteiger (central midfielder). These results on central midfielders and central defenders are consistent with those from %IdC and %OdC. The highest volume of passes between such *players'* positions, mainly in the first phase of building the attack, increases the connection among teammates in the defensive and middle sectors. Thus, the lowest values of %CC and %BC in forwards and strikers are justified. The results in forwards and strikers in the present study are consistent with the study conducted in the FIFA World Cup 2010 Finals (Peña & Touchette, 2012). The authors noted that forwards almost always can be identified as the players who have the lowest closeness and betweenness values explained by their patterns of play farthest from the defenders and midfielders who have the highest volume of play.

The present study had some limitations. First, each player was not identified. Only their *players'* position, not their name, was considered. This limitation is important mainly because some players have different levels of importance in teams, and the overall codification can decrease the individual perception of their contribution. Second, a particular national team was not analyzed because this study aimed to understand the global participation of *players'* positions and not to consider each national team. Surely, the results in some national teams may vary mainly because each national team has a specific style of play and a strategy to act. Despite these limitations, the findings of this study may increase the importance of network analysis in match analysis. The network metrics used in this study can provide useful information on the specific characteristics of each team and help the coaches to understand the most relevant and prominent players in building the attacking process. The results of this study can be a great practical application for the future of match analysis. This study also provides some opportunities for future works. The use of these tactical metrics may increase the knowledge on the specific tactical behavior of players in specific moments, such as only in counter-attack or even only in the first stage of building the attack during the defense–attack transition. Further studies should be conducted using different indicators, such as shots and goals, or defensive indicators, such as recovering balls and passing interceptions.

## 5. Conclusion

This study extends the previous research on network and team performance in football by incorporating repeated observations of top national teams that competed in FIFA World Cup 2014. Applied network centrality metrics measured the importance of each tactical position in building the attacking process. The results reveal that central midfielders are the prominent players in the attacking process in the majority of tactical lineups. These players show the highest levels of connection with their teammates and are significantly relevant in making passes and linking the sectors of the team. Goalkeepers, forwards, and strikers show the lowest contribution to building the attacking process, mainly owing to their specific tactical role. Their specific position decreases their participation in the first stages of building the attack.

## 6. Acknowledgments

This work was supported by the FCT project PEst-OE/EEI/LA0008/2013.

## 7. References

Anthonisse, I. M. (1971). **The Rush in a Graph**. Amsterdam, Netherlands: Mathematisch Centrum.

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.

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.

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.

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., & Tenreiro Machado, J. A. (2014). Dynamical Stability and Predictability of Football Players: The Study of One Match. **Entropy**, 16(2), 645-674.

Di Salvo, V., Baron, R., Tschan, H., Calderon Montero, F. J., Bachl, N., & Pigozzi, F. (2007). Performance characteristics according to playing position in elite soccer. **Int J Sports Med**, 28, 222-227.

Duarte, R., Araújo, D., Correia, V., & Davids, K. (2012). Sports Teams as Superorganisms: Implications of Sociobiological Models of Behaviour for Research and Practice in Team Sports Performance Analysis. **Sports Medicine**, 42(8), 633-642.

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

Freeman, L. C. (1979). Centrality in Social Networks Conceptual Clarification. **Social Networks**, 1, 215-239.

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

Hopkins, K. D., Hopkins, B. R., & Glass, G. V. (1996). **Basic statistics for the behavioral sciences**. Boston: Allyn and Bacon.

Hughes, M. D., & Bartlett, R. M. (2002). The use of performance indicators in performance analysis. **Journal of Sports Sciences**, 20(10), 739-754.

Kalamaras, D. (Producer). (2014). Social Networks Visualizer (SocNetV): Social network analysis and visualization software. Social Networks Visualizer.

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

Maroco, J. (2012). Análise Estatística com utilização do SPSS [Statistical analysis with SPSS]. Lisbon, Portugal: Edições Silabo.

Nieminen, J. (1974). On the centrality in a graph. **Scandinavian Journal of Psychology**, 15(1), 332-336.

O'Donoghue, P. (2012). **Statistics for sport and exercise studies: An introduction**. London and New York, UK and USA: Routledge Taylor & Francis Group.

Pallant, J. (2011). **SPSS Survival Manual: A Step by Step Guide to Data Analysis Using the SPSS Program**. Australia: Allen & Unwin.

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.

Peña, J. L., & Touchette, H. (2012). A network theory analysis of football strategies. Paper presented at the arXiv preprint arXiv.

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.

Sabidussi, G. (1966). The centrality index of a graph. **Psychometrika**, 31(581–603).

Vilar, L., Araújo, D., Davids, K., & Bar-Yam, Y. (2013). Science of winning football: emergent pattern-forming dynamics in association football. **Journal of Systems Science and Complexity**, 26, 73-84.

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

Yamamoto, Y., & Yokoyama, K. (2011). Common and unique network dynamics in football games. **PloS One**, 6(12), e29638.

**Corresponding author:**

Filipe Manuel Clemente

Polytechnic Institute of Coimbra, ESEC, Department of Education, Portugal
Filipe.clemente5@gmail.com