International Journal of Performance Analysis in Sport

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

General network analysis of national soccer teams in FIFA World Cup 2014

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

To cite this article: Filipe Manuel Clemente, Fernando Manuel Lourenço Martins, Dimitris Kalamaras, P. Del Wong & Rui Sousa Mendes (2015) General network analysis of national soccer teams in FIFA World Cup 2014, International Journal of Performance Analysis in Sport, 15:1, 80-96, DOI: 10.1080/24748668.2015.11868778

To link to this article: https://doi.org/10.1080/24748668.2015.11868778

Published online: 03 Apr 2017.

International Journal of Performance Analysis in Sport
2015, 15, 80-96.

General network analysis of national soccer teams in FIFA World Cup 2014

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

[1]Polytechnic Institute of Coimbra, ESEC, Department of Education, Portugal
*Corresponding author: Filipe.clemente5@gmail.com (Filipe M. Clemente)
[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]New Media Network Synapsis S.A.
[6]Human Performance Laboratory, Technological and Higher Education Institute of Hong Kong, Hong Kong, China

Abstract

This study analyzed the network characteristics of successful and unsuccessful national teams that participated in FIFA World Cup 2014. The relationship between the variables of overall team performance and the network characteristics measured on the basis of the passes between teammates was also investigated. A dataset of 37,864 passes between teammates in 64 soccer matches enabled the study on network structure and team performance of 32 national soccer teams. Our results showed significant differences in the dependent variables of network density ($F_{4,123} = 2.72$; $p = 0.03$; $\eta_p^2 = 0.04$; small effect size) and total links ($F_{4,123} = 2.73$; $p = 0.03$; $\eta_p^2 = 0.04$; small effect size) between the teams that reached the later stages of the tournament. Goals scored presented a small positive correlation with total links ($r = 0.24$; $p = 0.001$), network density ($r = 0.24$; $p = 0.001$), and clustering coefficient ($r = 0.17$; $p > 0.050$). High levels of goals scored were associated with high levels of total links, network density, and clustering coefficient. This study showed that successful teams have a high level of network density, total links, and clustering coefficient. Thus, large values of connectivity between teammates are associated with better overall team performance.

Keywords: Social network analysis; Match analysis; Soccer; Connectivity; Performance.

1. Introduction

The interactions of teammates in team sports games are a consequence of the rules and dynamics of the games (Gréhaigne, Bouthier, & David, 1997). In fact, the cooperation between teammates emerges from the need to overcome the opponent team through a strong collective organization (Duarte, Araújo, Correia, & Davids, 2012). The main idea is that the team is necessarily different than the sum of all its players; thus, a strong team seeks for a high level of efficacy in the synchronization and interaction between teammates (Fewell, Armbruster, Ingraham, Petersen, & Waters, 2012; Grund, 2012). Such cooperation between teammates in team sports can be understood as a social relationship that emerges in a micro-social system (Lusher, Robins, & Kremer, 2010). Having this idea in mind, the game of soccer may be regarded as an interesting and specific place to apply social network analysis in order to identify the properties of team network (Malta & Travassos, 2014).

Studies on social network analysis have highlighted the benefits of decentralized cooperation for team performance and the success of tasks (Balkundi & Harrison, 2006; Grund, 2012). The main rationale behind such results are associated with the benefits of exploiting the individual advantages of each member and generating the best possible result that integrates the contribution of each member, thus optimizing the final result (Martins, Clemente, & Couceiro, 2013). This rationale is in line with a meta-analysis that summarized these two main conclusions (Balkundi & Harrison, 2006): i) a network with the highest density values leads to the best performances, and ii) networks with high centralized tendencies are associated with poor performances.

These outcomes seem to be generalized for all social teams. Nevertheless, in the specific field of sports science, the number of research using social network approach is small. In this paper, the main limitations of previous research will be discussed after a brief review of the literature on team sports.

1.1. Brief Review of Social Network Analysis in Team Sports

One of the first published articles that used Social Network Analysis in team sports was conducted by Bourbousson, Poizat, Saury, and Seve (Bourbousson, Poizat, Saury, & Seve, 2010). The study was performed with basketball players; the goal was to understand how the players were connected with the activities of their teammates. The authors found four typical forms of team coordination network (Bourbousson et al., 2010): i) a network that involves a certain number of players who are associated with one or several other players having no connections with the remaining players; ii) a network in which the five players (all players in the field) are linked by dyadic coordination; iii) a network of two units, with all players in dyadic coordination; and iv) a network with no connections between any of the players.

One of the first published articles on soccer games studied the Union of European Football Associations (UEFA) Euro 2008 tournament (Duch, Waitzman, & Amaral, 2010). In the study, the application of flow centrality metric to classify collective and individual performances was observed. The metric was used along with an approach using weighted network based on the accuracy of passes and shots, which is a different approach to the regular network metric that only applies linking criteria.

In one application of social network analysis, a water polo study used only the passes to link the nodes (players) (Passos et al., 2011). In the study, two main parameters were associated with the successful patterns of play: i) the number of interactions between teammates, and ii) the probability of each player to interact with each teammate in subsequent phases of attack. The main conclusions highlighted that the most successful collective system behavior requires a high probability of each player interacting with other players in a team.

Recently, a network analysis of the English Premier League teams was also conducted (Grund, 2012). A dataset of 283,529 passes between teammates in 760 soccer matches was used. The goals scored (team performance) were associated with density and centralization metrics. The study showed that high levels of interactions (density) lead to increased team performance. In contrast, a centralized interaction was associated with a decrease in team performance.

Social network analysis has several applications in team sports. Nevertheless, some limitations can be identified. The majority of studies fixed their focus on the player but disregard their tactical positioning and variation during matches. This methodological approach reduces the possibility to understand how strategic position influences the team network. Moreover, the studies performed only attributed a mean value of network metric, and did not consider the evolution during the tournaments. Such studies mainly investigated competitions within a small period of time [e.g., FIFA World Cup and UEFA Europe Cup)]. Other interesting analyses that a few studies performed involve the association between team success and specific kinds of network properties and the difference between successful and unsuccessful teams.

1.2. Statement of Contribution

A small number of studies have identified the network properties of a team and compared the success of teams in soccer competitions. To classify network properties in a macro-analysis, the total links (the number of connections between teammates) were assessed. Density metric (which measures the overall affection between teammates), network diameter (which quantifies the distance between the farthest two players in the graph), and clustering coefficient (which measures the degree of interconnectivity in the neighborhood of a player) of a team network were used. These metrics were computed per team in each game. The present study has three main goals as follows:

(1) To analyses the differences between the teams that reached different stages in the competition and then compare their network performances.
(2) To compare the final scores of the teams with the network properties computed using the metrics previously described.
(3) To associate the number of goals scored and the number of goals conceded per match with the network values achieved per national team during their matches.

2. Methods

2.1. Sample

Sixty-four official matches from FIFA World Cup 2014 were analyzed in this study. All the matches of all 32 national teams that participated in the tournament were analyzed. Thus, a total of 128 adjacency matrices were generated on the basis of teammate interactions and converted into 128 network graphs. A total of 37,864 passes were analyzed.

2.2. Observation and Data Coding

An adjacency matrix must be generated to perform network analysis, which represents the connections between a node (player) and an adjacency node (teammate). To generate an adjacency matrix for network analysis, the criteria that characterize the connection must be defined. This study defined the passes between teammates as linkage criteria. An adjacency matrix per attacking unit must be generated to identify the attacking properties of the team. An attacking unit starts at the moment that a team player recovers ball possession for the team and makes a successful pass to a teammate who receives and controls the ball. The attacking unit ends when the team loses ball possession (e.g., out of bounds, shot, unsuccessful pass to the teammate). The passes between teammates in an adjacency matrix were recorded during each attacking unit. Each pass between nodes was given a code of 1. For no passes between teammates, a code of 0 was given. Figure 1 presents an example.

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

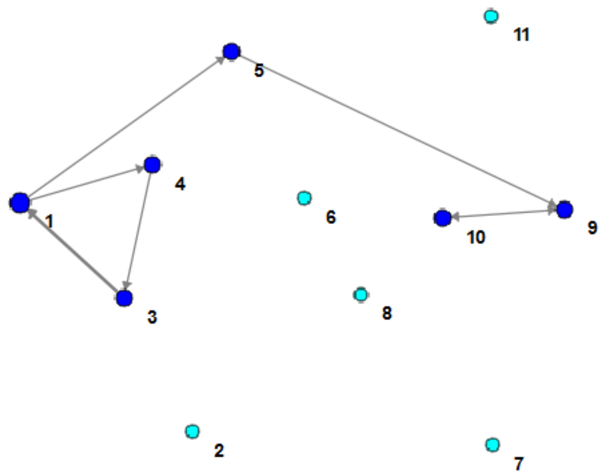

Figure 1. Example of adjacency matrix per attacking unit. The order of connections (passes) between teammates (order between numbers 1 and 11) was 3-1-4-3-1-5-9-10-9. In the table, the row represents that player n performed a given number of passes for the remaining teammates. The column represent that the player n received a given number of passes from their teammates.

Figure 1 shows that player 3 made two passes to player 1. Thus, in an adjacency matrix, the code is 2 for the interaction. This procedure was performed for all attacking units of each national team during the tournament. At the end of each match, an overall adjacency matrix that sums up all attacking units of the team in a single matrix was developed. The players were numbered from 1 to 11 to classify the interactions. Each number was coded on the basis of the tactical position of players. When a player was replaced by another, a new number was given in accordance with the tactical criteria.

The procedures of data collecting, which mainly pertains to the sequence of passes, and data inserting in adjacency matrices were performed by the same researcher, who has more than five years of experience in match analysis. To ensure the reliability of the data collecting and codification processes, a test-retest reliability was conducted using Cohen's Kappa test by adhering 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 number of matches. The Kappa value ensured a recommended margin for these kinds of procedures (Robinson & O'Donoghue, 2007).

2.3. Network Analysis

The 128 overall adjacency matrices were generated based on the passes between teammates, and then imported into Social Networks Visualizer (SocNetV) for analysis. SocNetV (Kalamaras, 2014) is a graphical application for the analysis and visualization of social networks. It allows a researcher to load formatted network data such as sociomatrices, and analyze the social and mathematical properties of the corresponding social networks in the form of mathematical graphs. The application also computes basic graph properties, such as density, diameter, and clustering coefficient, which were used in this study, as well as advanced structural measures, such as centrality and prestige indices, which were out of the scope of this study.

The network analysis of the 32 national teams in FIFA World Cup 2014 was focused on the following four measures based on the connections between teammates: i) total links, ii) density, iii) diameter, and iv) clustering coefficient.

2.4. Total Links

For each of the 64 matches played in World Cup 2014, we developed two adjacency matrices. Each element *(i, j)* of the adjacency matrix is the number of interactions (passes) from players *i* to *j*. In terms of the corresponding graph (sociogram) produced by SocNetV, the adjacency matrix was represented by a directed line (arc) between nodes *i* and *j*.

The sum of the elements of each row of the adjacency matrix $\sum_{j \neq i}^{11}(i,j)$ is the total number of passes from player *i* to all its other teammates, which is the Nodal outDegree in graph theoretical terms. Total outbound links from the corresponding node represent the player to the other node players. A node with a high outDegree is a player who made considerable passes to most of its teammates. This player was thus involved in the attacking development of his team.

The total sum of each adjacency matrix row sum $L = \sum_{i}^{11}\sum_{j \neq i}^{11}(i,j)$ was the first metric used in this study, that is, the *Total Links* (passes) between each team player. In the corresponding graph, this number is the total lines between all nodes (in our case, arcs, because the graph is directed).

The Total Links measure is useful in comparing the 32 teams because it is the absolute number of the total interactions conducted between teammates during the match. Thus, a higher than average Total Links index of a team is an indicator of strong cooperation between team players. This index may also be correlated with a high probability of the players to interact successfully with one another, which may result in long ball possession, good performance, and generally strong collective organization against the opponent team.

2.5. Network Density

Total Links metric is an absolute number of interactions from one player to another player. The density of team network is a relative index that also measures the overall affection between teammates.

In graph theory, the density of a (directed) graph is the proportion of the maximum possible lines present between nodes. A graph consists of a finite number of nodes (denoted by *n*). In case of an undirected graph, maximum $n\,(n-1)/2$ possible pairs between nodes and $n\,(n-1)/2$ possible links [divided by 2 because the link (*i*, *j*) is the same to (*j*, *i*)] can exist. The density *Δ* of the graph is defined as the ratio of the total links that present *L* to the maximum possible number of links as follows:

$$\Delta = \frac{L}{n(n-1)/2} \text{or } \Delta = \frac{2L}{n(n-1)}. \quad (1)$$

In case of ordered relations, as in the teammate interactions investigated in this study, the possible directed links in a digraph of *n* nodes are $n(n-1)$. Thus, the density is computed by

$$\Delta = \frac{L}{n(n-1)}. \quad (2)$$

In both cases, the density is a fraction with a minimum of 0 (no lines/arcs present) and a maximum of 1 (all lines/arcs are present). However, Wasserman and Faust (Wasserman & Faust, 1994) suggested that for weighted (or valued) graphs or digraphs, the notion of density can be generalized by averaging the values attached to the lines/arcs across all lines/arcs. Given that each national team network corresponds to a valued digraph [(*i*, *j*) element of the adjacency denotes the number of passes between players *i* and *j* and can be larger than unity], we could also measure the average strength of the arcs in it and compute its density D using the following formula:

$$\Delta = \frac{\Sigma w_k}{n(n-1)}. \quad (3)$$

This condition applied for all *k* values of the adjacency matrix. However, the valued digraph formula of density was not used in this study.

2.6. Network Diameter

The first two metrics (total links *L* and density *Δ*) focus on the number of links inside a given social network. The *diameter d* of the corresponding graph is related to the distance between nodes. In graph theory, two nodes are *connected* if a sequence of nodes exists and their links (*walk*) are adjacent. In the example above, nodes 3 and 9 are connected through the walk 3-1-4-3-1-5-9. If a walk consists only of distinct nodes and lines, then this walk is called a *path*. In the same example above, the path 3-1-5-9 exists from nodes 3 to 9.

The *distance d* (or *geodesic distance*) between two nodes of a graph is the length of the shortest path (called *geodesic*) between them. A pair of connected nodes may have more than one shortest path that connects them. In any case, the distance between the two nodes is the length of any shortest path between them.

The *diameter* of a graph is the maximum distance (the length of the largest geodesic) between any two connected nodes and is computed by the formula $diameter = max_i max_j\, d(i, j)$.

The diameter has a minimum of 1 when all nodes are directly connected with each other, and a maximum of $n - 1$ when a path between a pair of nodes passes by every other node of the network. Therefore, the diameter of a network is an important metric because it reflects how far, at most, two nodes in the graph are. In case of national team players, a small diameter reflects a low maximum distance between teammates, which may reveal that the team's passing game was diffused among most of its players (rather than a few acting as central ones).

2.7. Clustering coefficient

Clustering coefficient, which was introduced by Watts and Strogatz (Watts & Strogatz, 1998), quantifies how close a node and its neighbors in a graph are to become a clique (a complete subgraph). Watts and Strogatz used the local version of clustering coefficient to determine whether a graph is a small-world network (a network of small average distance but relative large number of cliques).

In case of directed graphs, the (local) clustering coefficient of a node is the proportion of links present between nodes directly connected to it (*neighborhood N* of the node). Thus, the local clustering coefficient of each node i is computed as the fraction of the number of all arcs a_{jk} between k_i nodes in its neighborhood divided by the maximum number $k_i(k_i - 1)$ of links that could exist among them. That is,

$$C_i = \frac{|\{a_{jk}, a_{jk} \in E\}|}{k_i(k_i - 1)}. \tag{4}$$

This condition applied for all j, k nodes in the neighborhood N of i.

Thus, the local clustering coefficient measures the degree of interconnectivity in the neighborhood of a node. A high degree means that the node and its neighbors are close to become a clique.

We use a variant of the global version of clustering coefficient, which measures the level of clustering in the entire network. This variant is the network average of the local clustering coefficients:

$$\overline{C} = \frac{1}{n} \sum_{i=1}^{n} C_i. \tag{5}$$

A network with a high cluster coefficient means its nodes tend to form cliques, which may suggest poor overall cooperation and low interconnectivity because actors seem to communicate more with certain nodes than with all. In the case of national team players, higher than average values of clustering coefficient may be an indicator of decreased performance. Therefore, teammates tend to limit their passing options to a few players.

All aforementioned four metrics were computed per team in each game.

2.8. Variables and Objectives of this Study

Besides the graph characteristics measured by network metrics, several performance variables that define the overall performance of a team and characterize the attacking performance of the team were used in this study. The overall performance of a team was considered as the maximum stage that each team reached in FIFA World Cup 2014. The variables of maximum stage in the competition were as follows: i) group stage, ii) round of 16, iii) quarterfinals, iv) semifinals, and v) finals. The final score per match was the second overall performance variable considered in this study. The variables were as follows: i) lose, ii) draw, and iii) win. The following attacking variables were also considered: i) number of scored goals per match, ii) number of shots per match, and iii) number of shots on goal per match. This information about team performance was directly extracted from the official website of FIFA World Cup 2014[1], which provided the team statistics per match.

Considering the network graph performance variables (density, centrality, and clustering coefficient) and the team performance variables (maximum stage in competition, score, goals, shots, and shots on goal) during the competition, three objectives were defined for this study:

(i) Analyze the differences (if any) on network graph variables between the teams that reached different stages during the competition. Teams that achieved the highest stages in competition were expected to show high values of density and low values in clustering coefficient and diameter.

(ii) Analyze the differences (if any) on network graph variables between the teams that achieved different final scores in the tournament. The highest values of density were expected in teams that won, whereas the highest values of diameter and clustering coefficient were expected in teams that lost.

(iii) Analyze if the team variables of scored goals, shots, and shots on goal are associated with the network graph variables. The highest values of density were expected to lead to increased team performance, whereas the highest values of clustering coefficient and diameter were expected to lead to decreased team performance.

2.9. Statistical Procedures

The influences of maximum stage on competition and final score per match factor on the density, diameter, and clustering coefficient were analyzed using one-way ANOVA

[1] http://www.fifa.com/worldcup/

after validating normality and homogeneity assumptions (O'Donoghue, 2012; Pallant, 2011). The assumption of normality for each univariate dependent variable was examined through Kolmogorov–Smirnov tests ($p > 0.05$). Although the distributions are not normal in the dependent variables, since $n > 30$ and considering the Central Limit Theorem, the assumption of normality was assumed (O'Donoghue, 2012). The assumption of homogeneity of the variance/covariance matrix of each group was examined with Box's M test. When ANOVA detected significant statistical differences between factors, we proceeded to utilize Tukey's HSD post-hoc test (Maroco, 2012). All statistical analyses were performed using IBM SPSS Statistics (version 22) at a significance level of $p < 0.05$. The following scales were used to classify the effect size 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; nearly perfect, 0.81–1.0.

The relationship between network metrics (density, centrality, and clustering coefficient) and team performance variables (goals scored, shots, and shots on goal) was investigated using Pearson product moment correlation coefficient. Preliminary analysis was performed to ensure no violation of the assumptions of normality, linearity, and homoscedasticity, as suggested by Pallant (Pallant, 2011). The following scales were used to classify the correlation strength (Hopkins et al., 1996): very small, 0–0.1; small, 0.1–0.3; moderate, 0.3–0.5; large, 0.5–0.7; very large, 0.7–0.9; 0.9–1, nearly perfect; 1, perfect. All statistical analyses were performed using IBM SPSS Statistics (version 22) at a significance level of $p < 0.05$.

3. Results

A defining feature of a team is how teammates interact in the moments when they have possession of the ball. Thus, this study analyzed the network differences between national teams that reached different stages in FIFA World Cup 2014 and compared the network characteristics in three possible scores. All national teams were analyzed. The network results are shown in Table 1.

Table 1. Descriptive statistics (mean and standard deviation) of network performance per national team

	Total Links		**Density**		**Clustering Coefficient**		**Diameter**	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Algeria	70.50	7.30	0.64	0.07	0.68	0.08	2.25	0.43
Argentina	82.71	3.95	0.75	0.04	0.78	0.03	2.29	0.45
Australia	83.00	2.16	0.75	0.02	0.76	0.01	2.00	0.00
Belgium	82.20	6.08	0.75	0.06	0.79	0.05	2.40	0.49
Bosnia and Herzegovina	81.67	2.49	0.74	0.02	0.78	0.03	2.00	0.00
Brazil	80.86	4.58	0.74	0.04	0.76	0.02	2.43	0.49
Cameroon	72.67	1.70	0.66	0.02	0.70	0.02	2.33	2.33
Chile	87.50	4.82	0.80	0.04	0.80	0.04	2.00	0.00
Colombia	75.60	5.08	0.69	0.05	0.72	0.03	2.60	0.49
Costa Rica	76.20	4.31	0.69	0.04	0.71	0.03	**2.80**	**0.40**
Côte d'Ivoire	78.33	7.72	0.71	0.07	0.72	0.06	2.00	0.00
Croatia	82.00	6.68	0.75	0.06	0.76	0.06	2.67	0.47
Ecuador	74.00	2.94	0.67	0.03	0.71	0.05	**3.00**	**0.00**
England	88.50	0.03	0.80	0.03	**0.82**	**0.01**	2.67	0.47
France	82.00	3.74	0.75	0.03	0.76	0.05	2.20	0.40
Germany	**89.29**	**2.76**	**0.81**	**0.03**	**0.82**	**0.02**	2.00	0.00
Ghana	77.33	1.89	0.70	0.02	0.77	0.04	2.33	0.47
Greece	73.00	17.62	0.66	0.16	0.70	0.11	2.75	0.83
Honduras	77.67	4.50	0.71	0.04	0.75	0.03	2.00	0.00
Iran	55.00	8.83	0.50	0.08	0.60	0.03	**3.33**	**0.47**
Italy	82.67	7.13	0.75	0.06	0.79	0.03	2.33	0.47
Japan	79.67	4.11	0.72	0.04	0.80	0.01	2.33	0.47
Korea Republic	77.00	0.82	0.70	0.01	0.75	0.01	2.00	0.00
Mexico	81.75	4.02	0.74	0.04	0.77	0.04	2.25	0.43
Netherlands	82.57	10.70	0.75	0.10	0.75	0.09	2.29	0.45
Nigeria	75.00	4.06	0.68	0.04	0.73	0.03	2.50	0.50
Portugal	**89.33**	**0.03**	**0.81**	**0.03**	0.81	0.03	2.00	0.00
Russia	81.33	4.71	0.74	0.04	0.77	0.03	2.33	0.47
Spain	**91.67**	**3.40**	**0.83**	**0.03**	**0.83**	**0.03**	2.00	0.00
Switzerland	76.50	8.79	0.70	0.08	0.75	0.05	2.50	0.50
Uruguay	75.00	10.27	0.68	0.09	0.74	0.13	2.50	0.50
USA	77.75	4.32	0.71	0.04	0.73	0.04	2.00	0.00

Values in bold text represent the highest values. Values highlighted in gray represent the lowest values

Table 1 shows that, on one hand, Germany (the winner of FIFA World Cup 2014) had the highest mean values of total links between teammates (89.29 ± 2.76), network density (0.81 ± 0.03), and clustering coefficients (0.82 ± 0.02). On the other hand, Germany had the lowest value of network diameter (2.00 ± 0.00). Iran national team had the lowest mean values of total links between teammates (55.00 ± 8.83), network density (0.50 ± 0.08), and clustering coefficient (0.60 ± 0.03), but had the highest value of network diameter (3.33 ± 0.47).

The aforementioned results generated an example of a network graph (a single match) of the team with the highest values of total links, network density, and cluster coefficient (Figure 3a), and an example of the team with the lowest values of such network metrics (Figure 3b). The node size represents the weight of passes performed and received (large nodes mean high values of passes performed and received). The size of edges represents the weight of connection between nodes (large edges means high values of interactions between teammates).

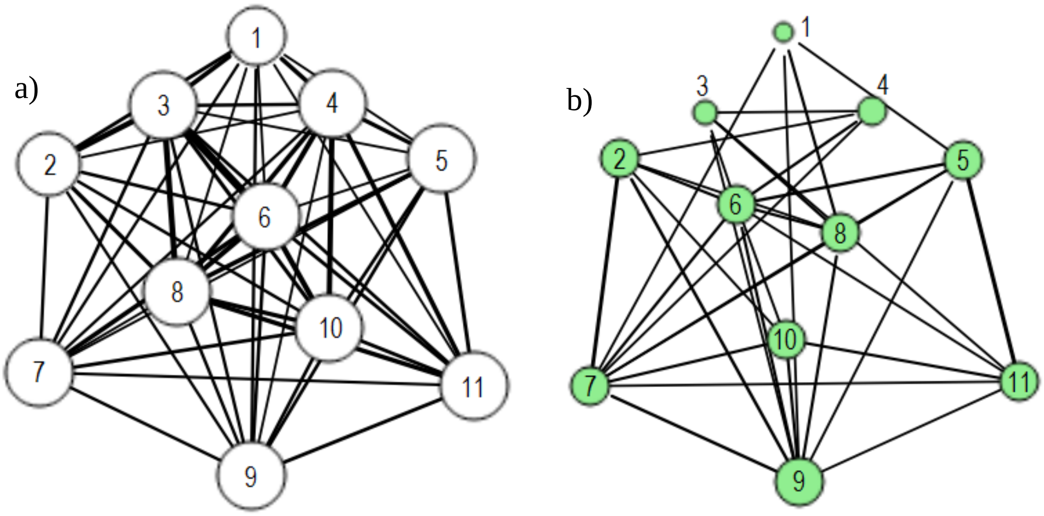

Figure 3. Network graph of a single match: a) Graph (white) with the highest values of total links, network density, and clustering coefficient; b) Graph (green) with the lowest values of total links, network density, and clustering coefficient.

Figure 3 shows that the interactions between teammates in the white graph are greater than the network in the green graph. Such evidence can be confirmed by the highest value of total links (94 of white graph vs. 43 of green graph), network density (0.85 of white graph vs. 0.39 of green graph), and clustering coefficient (0.84 of white graph vs. 0.58 of green graph). However, the green graph had the highest value of graph diameter (4) in comparison with the white graph (2).

ANOVA test and a post-hoc test were performed to identify if differences existed between the maximum stage in competition and the network characteristics of teams. The results are presented in Table 2.

Table 2. Descriptive table (mean and standard deviation) and statistical comparison between factors (maximum stage reached in tournament)

	Group Stage	**Round of 16**	**Quarterfinals**	**Semifinals**	**Finals**
Total Links	79.49 (9.46)	77.13 (10.29)[e]	79.00 (5.94)	81.71 (8.59)	86.00 (4.91)[b]
Density	0.72 (0.09)	0.70 (0.09)[e]	0.72 (0.05)	0.74 (0.08)	0.78 (0.04)[b]
Clustering Coefficient	0.76 (0.06)	0.74 (0.08)[e]	0.74 (0.05)	0.76 (0.07)	0.80 (0.03)[b]
Diameter	2.33 (0.52)	2.34 (0.55)	2.50 (0.51)	2.36 (0.50)	2.14 (0.36)

Significantly different compared with Group Stage[a], Round of 16[b], Quarterfinals[c], Semifinals[d], and Finals[e] at $p < 0.05$

Statistical differences were found between the maximum stages achieved in tournament and the dependent variables of network density ($F_{4,123}$ = 2.723; p = 0.033; η_p^2 = 0.039; small effect size) and total links ($F_{4,123}$ = 2.725; p = 0.032; η_p^2 = 0.039; small effect size). No differences were found in the clustering coefficient ($F_{4,123}$ = 2.234; p = 0.069; η_p^2 = 0.030; small effect size) and network diameter graph ($F_{4,123}$ = 1.026; p = 0.397; η_p^2 = 0.008; very small effect size).

The one-way ANOVA and the following post-hoc test were conducted to analyze the variance between the final scores per match. The descriptive statistics are presented in Table 3.

Table 3. Descriptive table (mean and standard deviation) and statistical comparison between factors (final score per match)

	Lose	**Draw**	**Win**
Total Links	77.69 (8.57)^c	78.90 (10.70)	82.33 (7.99)^a
Density	0.71 (0.08)^c	0.72 (0.10)	0.75 (0.07)^a
Clustering Coefficient	0.74 (0.06)	0.75 (0.08)	0.77 (0.07)
Diameter	2.37 (0.53)	2.40 (0.58)	2.29 (0.46)

Significantly different compared with Lose^a, Draw^b, and Win^c at p < 0.05.

Statistical differences were found between the final score per match (lose, draw, and win) and the dependent variables of network density ($F_{2,125}$ = 3.729; p = 0.027; η_p^2 = 0.065; small effect size) and total links ($F_{3,125}$ = 3.731; p = 0.027; η_p^2 = 0.065; small effect size). No differences were found in the clustering coefficient ($F_{2,125}$ = 2.307; p = 0.104; η_p^2 = 0.063; small effect size) and network diameter graph ($F_{2,125}$ = 439; p = 0.646; η_p^2 = 0.033; small effect size).

The relationship between team performance variables (goals scored, overall shots, and shots on goal) and the characteristics of the network graphs (total links, network density, clustering coefficient, and diameter) was investigated using Pearson product-moment correlation coefficient. The values of the coefficients are shown in Table 4.

Table 4. Correlation values between the team performance variables and the network values provided by the metrics

	GS	OS	SG	TL	ND	CC	D
Team Attacking Performance							
(1) GS: Goals Scored	1	0.238**	0.375**	0.240**	0,240**	0.172	−0.142
(2) OS: Overall Shots		1	0.884**	0.204*	0.204*	0.237**	−0.141
(3) SG: Shots on Goal			1	0.197*	0.196*	0.188*	−0.184*
Network Performance							
(4) TL: Total Links				1	1.000**	0.905**	−0.593**
(5) ND: Network Density					1	0.905**	−0.593**
(6) CC: Clustering Coefficient						1	−0.465**
(7) D: Diameter							1

* Correlation is significant at p ≤ 0.050. ** Correlation is significant at p = 0.001.

The goals scored showed a small positive correlation with total links ($r = 0.240$; $p = 0.001$), network density ($r = 0.240$; $p = 0.001$), and clustering coefficient ($r = 0.172$; $p > 0.050$). High levels of goals scored were associated with high levels of total links, network density, and clustering coefficient. A small negative correlation existed between goals scored and graph diameter ($r = -0.142$; $p > 0.050$).

The overall shots indicated a small positive correlation with total links ($r = 0.204$; $p \leq 0.050$), network density ($r = 0.204$; $p \leq 0.050$), and clustering coefficient ($r = 0.237$; $p = 0.001$). High levels of overall shots were associated with high levels of total links, network density, and clustering coefficient. A small negative correlation existed between overall shots and graph diameter ($r = -0.141$; $p > 0.050$).

The shots on goal had a small positive correlation with total links ($r = 0.197$; $p \leq 0.050$), network density ($r = 0.196$; $p \leq 0.050$), and clustering coefficient ($r = 0.188$; $p \leq 0.050$). High levels of shots on goal were associated with high levels of total links, network density, and clustering coefficient. A small negative correlation was observed between overall shots and graph diameter ($r = -0.184$; $p \leq 0.050$).

4. Discussion

Can great cooperation between teammates increase team performance? This question applies for any biological collective organization. In animals (even in prey or predators), collective organizations are formed to increase the defensive or offensive performance to ensure survival (Couzin, Krause, Franks, & Levin, 2005). A similar kind of cooperation occurs in human organizations to improve the efficacy of action and achieve the ultimate goal (Balkundi & Harrison, 2006). Despite such evidence in social network analysis, the knowledge about the network characteristics that occur in team sports during a match is not solid (Passos et al., 2011). Some studies have been published in the last couple of years (Bourbousson et al., 2010; Duch et al., 2010; Grund, 2012), yet the multiple and disperse uses of the network metrics studied do not allow strong evidence on the importance of a specific kind of network characteristic to the overall performance of a team.

Recent scientific literature suggests that the play patterns of soccer teams vary and emerge from the self-organization between players (teammates and opponents) (McGarry, Anderson, Wallace, Hughes, & Franks, 2002). Despite such evidence, some stable procedures regulate the teammate interactions and tactical actions in soccer teams (Couceiro, Clemente, Martins, & Tenreiro Machado, 2014). Thus, this study analyzed if differences exist between teammate interactions on the basis of the overall team performance (maximum stage achieved in FIFA World Cup 2014 and the final score per match).

Some studies have suggested that the passing strategies for successful and unsuccessful teams do not differ (passes performed), which indicates that the performance variable does not easily discriminate teams at FIFA World Cup 2002 (Scoulding, James, & Taylor, 2004). Similar results have been found in FIFA World Cup 2010 (Clemente, 2012). Despite such general evidence, the passes performed between teammates are always one of the main indicators of a team (Carling, Williams, & Reilly, 2005). Nevertheless, traditional notational analysis does not consider the characteristic of interactions between teammates. Thus, the current study applied several well-known metrics used on social network analysis to identify the properties of teammate interactions (assessed by passes), as well as to analyze the variance between successful and unsuccessful national teams in FIFA World Cup 2014.

Statistical differences were found between the maximum stages achieved in tournament and the dependent variables of network density (overall affection between teammates) and total links (number of connections between teammates). The results showed that teams that achieved the highest stages in competition had great values of total links and network density. Thus, the successful teams had a greater distribution of interaction between teammates than the unsuccessful teams. In fact, the ability to decentralize the patterns of passes may be a specific characteristic to increase variability of action and decrease exposure to the opponent (Gréhaigne et al., 1997). Not only did the analysis of national teams that reached the highest stages in competition showed differences in network density and total links, but the analysis between final scores also confirmed the differences between the teams that won and lost the matches. The greatest values of such metrics were identified in the winning teams. These values are in line with literature on social network analysis that suggest an increase in mutual interdependence when team members have strong interactions with one another (Sparrowe, Liden, Wayne, & Kraimer, 2001). This finding thus calls for cooperation and coordination of actions (Molm, 1994).

In addition to the total links and network density, the clustering coefficient and diameter of network graph were tested. The diameter allows for identifying the distance between the farthest two players in a graph. The clustering coefficient measures the interconnectivity in the neighborhood of a player. High values of cluster coefficients mean that every player and its neighborhood are complete cliques, and high values of diameters mean that the players are farthest from each other. Statistical differences in clustering coefficient were found between successful and unsuccessful teams that reached different stages in tournament. The highest values were found in national teams that reached the final stages of competition. Such results suggested that great cooperation and interconnectivity can lead to a high efficacy to achieve the best performance in soccer. In fact, the team with the highest levels of clusters may decrease the ability to act as one, which is the main priority of a collective organization (Davids, Araújo, & Shuttleworth, 2005). No statistical differences were found in diameter analysis. Nevertheless, the lowest mean value was found in teams that reached the final stages and in the winning teams. Such analysis of proximity between connections may not be crucial in football because of the regularity of players who use crossing passes to penetrate in the opponent organization. In other sports where ball travelling is controlled and the proximity makes a difference (such as basketball and handball), statistical differences may easily be found.

This study also analyzed the association between several regular performance variables, such as goals scored, shots, and shots on goal. A previous study that analyzed 760 matches in the English Premier League identified that the number of goals scored is strongly associated with the highest levels of network density (Grund, 2012). In our study, significant positive correlations were found between the total links and network density with the number of goals scored, shots, and shots on goal. Only two network metrics had significant correlations with all performance variables (e.g., goals and shots). Thus, such results confirmed the previous findings of Grund (Grund, 2012), and provided an interesting idea about the ability of teams to maximize cooperation to increase the possibilities of play and overcome the defensive organization of opponents.

The highest levels of clustering coefficient led to increased overall shots and shots on goal. The possibility of ensuring a strong connection between teammates could increase the possibility of ensuring ball possession and providing a support for the player with the ball to make accurate passes to forward players. The diameter metric had a negative correlation with all performance variables. Thus, high levels of diameter led to decreased team performance. The farthest connection between teammates may increase the predictable attacking process, thus decreasing the possibility of creating instability on the defensive organization of opponents.

The important contribution of this present study lies in differentiating the successful and unsuccessful national teams present in FIFA World 2014, extending previous studies on associated network metrics with performance variables, and increasing the knowledge on the connectivity behavior of teammates in soccer. Despite these contributions, this study had its limitations. The main limitation could be associated to differentiating the efficacy of a network. Two teams can have similar network density and teammate connections, but their ability to make efficient passes may be significantly different. In one team, the travelling of the ball can be performed in longitudinal axis (goal-to-goal), thus exploring the width of the field to increase proximity to the opponent goal. In the other team, the play style may tend to explore the width of the field (side-to-side), which does not move forward in the field. Another limitation was the unidentified patterns of relationship within a team. The network metrics used in this study only gave the overall characteristics of the graph but did not provide information about specific interactions in a team. Further studies may use several centrality measures that provide information about the most important players who determine the network graph structure, thereby identifying the players who generate the attacking process and the players who are the main targets of their teammates. Such information can provide relevant information about a specific team-playing style and establish a possibility to apply network metrics in match analysis, thus giving a different approach to soccer coaches.

5. Conclusion

The main findings of this study suggest that successful teams have the highest levels of network density, total links, and clustering coefficient. Thus, the ability to increase the connection between all teammates may result in excellent overall team performance in a tournament. High levels of total links, density, and clustering coefficient lead to increased team performance in matches (goals scored, overall shots, and shots on goal). The farthest relationship between teammates (measured by the network graph diameter) may decrease the possibility of scoring and shooting.

6. Acknowledgments

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

7. References

Balkundi, P., & Harrison, D. A. (2006). Ties, leaders, and time in teams: Strong inference about network structure's effects on team viability and performance. **Academy of Management Journal**, 49(1), 49-68.

Bourbousson, J., Poizat, G., Saury, J., & Seve, C. (2010). Team Coordination in Basketball: Description of the Cognitive Connections Among Teammates. **Journal of Applied Sport Psychology**, 22(2), 150-166.

Carling, C., Williams, A. M., & Reilly, T. (2005). **Handbook of Soccer Match Analysis: A Systematic Approach to Improving Performance**. London & New York: Taylor & Francis Group.

Clemente, F. M. (2012). Study of Successful Teams on FIFA World Cup 2010 through Notational Analysis. **Pamukkale Journal of Sport Sciences**, 3(3), 90-103.

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.

Couzin, I. D., Krause, J., Franks, N. R., & Levin, S. A. (2005). Effective leadership and decision-making in animal groups on the move. **Nature**, 433, 513-516.

Davids, K., Araújo, D., & Shuttleworth, R. (2005). Applications of Dynamical Systems Theory to Football. In T. Reilly, J. Cabri & D. Araújo (Eds.), **Science and Football V** (pp. 556-569). Oxon: Routledge Taylor & Francis Group.

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.

Fewell, J. H., Armbruster, D., Ingraham, J., Petersen, A., & Waters, J. S. (2012). Basketball teams as strategic networks. **PloS One**, 7(11), e47445.

Gréhaigne, J. F., Bouthier, D., & David, B. (1997). Dynamic-system analysis of opponent relationship in collective actions in football. **Journal of Sports Sciences**, 15(2), 137-149.

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.

Kalamaras, D. (2014). The SocNetV Manual Social Networks Visualizer (2014)

Lusher, D., Robins, G., & Kremer, P. (2010). The application of social network analysis to team sports. **Measurement in physical education and exercise science**, 14(4), 211-224.

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.
Martins, F. M. L., Clemente, F. M., & Couceiro, M. S. (2013). From the individual to the collective analysis at the football game. Paper presented at the Proceedings Mathematical Methods in Engineering International Conference, Porto.
McGarry, T., Anderson, D., Wallace, S., Hughes, M., & Franks, I. (2002). Sport competition as a dynamical self-organizing system. **Journal of Sports Sciences**, 20(10), 771-781.
Molm, L. D. (1994). Dependence and risk: Transforming the structure of social exchange. **Social Psychology Quarterly**, 57(3), 163-176.
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.
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.
Scoulding, A., James, N., & Taylor, J. (2004). Passing in the Soccer World Cup 2002. **International Journal of Performance Analysis in Sport**, 4(2), 36-41.
Sparrowe, R. T., Liden, R. C., Wayne, S. J., & Kraimer, M. L. (2001). Social networks and the performance of individuals and groups. **Academy of Management Journal**, 44(2), 316-325.
Wasserman, S., & Faust, K. (1994). **Social network analysis: Methods and applications**. New York, USA: Cambridge University Press.
Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. **Nature** 393, 393, 440–442.