http://pjss.pau.edu.tr
http://psbd.pau.edu.tr

PJSS

ISSN: 1309-0356

Pamukkale Journal of Sport Sciences
2017, Vol. 8, No. 3, 45-57

accepted: 21.06.2017

A network-science approach: how different are the attacks of football?

Necmi Gursakal[1], Halil Orbay Cobanoglu[2] and Sendi Ipeker Cagliyor[3]

[1] Uludag University, Faculty of Economics and Administrative Sciences, Bursa, Turkey.
[2] MEB, Muzaffer Cil Anadolu High School, Eskisehir, Turkey.
[3] Halic University, Faculty of Business Administration, Istanbul, Turkey.

orbay_79@hotmail.com

Abstract

Network science is an emerging field. The purpose of this study is to investigate soccer attacks by using network science. In this study, by applying network science approach, four Turkish National Football Team's attacks analyzed with an open-source NodeXL program. We have focused on two types of attacks: the attacks that end with goals and the ones that don't. Our main aim is to see whether there is a difference between the network metrics of these two types of attacks? Using network metrics, for attacks in a same match we couldn't find important differences but we have found real differences for networks' metrics when opponent team changes. Our findings also support that micro measures can be used for new line-up's. First of all, it should be mentioned that our study is a case study and the results of this study should not be generalized. However, our findings can be the start point for further researches with larger samples sizes. With the help of network science approach, the most effective players could be found, the most compatible line-up for the future games could be chosen and the opponent team's key players could be analyzed.

Keywords: Network science, network metrics, football, pass data

INTRODUCTION

Football is one of the most popular sports in the world. Especially in the last decades, with the increasing interest towards to it, football has become something more than just a game; it has evolved to an industry. For instance, a professional soccer club in a league from a smaller European country is like a small commercial or medium-sized enterprise (SME) in terms of its turnover and number of employees and has fans and members of extraordinary loyalty (Dolles and Söderman, 2005). Despite to its popularity all around the globe, or to its financial value as an industry, football has not received much attention from the scientific community (Kooij et al., 2009). The need for "scientific approaches to sports both for performance evaluation or predictions has been highlighted by various sources in the literature.

To respond this need, in this study, we suggest using network science to analyze football attacks. Network science is a new field that has been developed at the dawn of the 21st century (Barabasi, 2002). It is a highly interdisciplinary field that is concerned with the study of networks, which can be biological, technological or social. Before network science it hasn't been known that the structure and the evolution of the networks behind each system is driven by a common set of fundamental laws and principles (Barabasi, 2014).

Even though it is not very common, there are examples of networks science applications for sports analytics. For example, the article "Basketball isn't a sport. It's a statistical network" (Mossop, 2012) is about ta research done by researches in Arizona State University, where network science is applied to analyze basketball games. Another example, "Can Complex Network Metrics Predict the Behavior of NBA Teams?" highlights that box score statistics is significant but not sufficient to predict the success of a team. To fulfill this gap, they suggest new models for predicting a team success based on complex network metrics, such as clustering coefficient and node degree.

Based on its structure, where players can be considered as the nodes and passes as the edges of a network, a football team in a match can be seen as a network too (Pena and Touchette, 2012). As a result of the rapid development in network sciences, and due to its dynamic nature, which is more suitable to evaluate sports than the more static analysis methods, analysis of passes network gained interest among researches after millennium.

Using network analysis for evaluation of football teams and matches provide substantial information that could be useful for coaches. For instance, statistics gained by network analysis of previous games can offer supplementary decision tools for coaches to choose the optimal line-up upcoming games. As an example, based on the topological metrics of players in terms of participation, coaches can choose a line-up, where as many players as possible have already played together, if it is assumed that a team becomes better when enough players have played together before (Kooij et al., 2009).

Similarly, network science gives information about the cliques and motifs that dominate the game. For example, with closeness centrality, how well connected a player to the team can be calculated. Out-degree of a football player can be used as a metric to measure the passes he has given. The clustering coefficient can be used as a sign of "possession" in the pass network. And, betweenness, which naturally capture the hubs and essential associations in the distribution of the ball, has crucial importance to analyze network metrics (Cotta et al., 2011).

There are different researches, where network science approach is applied to examine football games. To summarize, first of all it has been found that the power law in degree distribution emerged in passing behavior in the 2006 FIFA World Cup Final and an international "A" match in Japan. In the same research, the exponent value γ~3.1 has been found similar to the values which occur in many real-world networks. In a different study, the frequencies of passing interactions within the same team in 5-min intervals have been analyzed and found that when a player who touched the ball many times changes the player to whom he was connected by passes and this is called hub-switching behavior (Yamamoto and Yokoyama, 2011). Another research suggests that network in the second half of the game network density decreases, heterogeneity increases and centralization decreases (Clemente et al., 2015). However, we know that if all players have the same centrality, the homogeneity level will be high. Hence, in second half of games that we analyzed, the increase in, heterogeneity shows that all players have different centralities. Beside that, in general centralizations decreased in the second half. Also different centralities lead us to think about "a direct play" rather than "team work".

Also another study shows that, high levels of interaction between teammates (density) led to increased team performance" (Clemente et al., 2015). We have also conflicting relations between density and success, **"**…which eventually leads to victory…" where in our findings the density of the pass network decreases with time (Cotta et al., 2013). Finally, the amount of correlations between degree, betweenness, closeness, and eigenvector indicates that these measures are distinct, yet conceptually related, the only network variable that has positively and significantly associated with correlations between all centrality measures was reciprocity (Valente et al., 2008).

In light of all this information, we can expect that attacks of networks, which end with a goal, should have higher densities. And if increases in network centralization lead to decreased team performance, then we should expect that attacks of networks, which do not end with a goal, should have high centralizations (Grund, 2012).

METHODS

Turkish National Football Team's attack networks and metrics as a case study

In this paper, we focus on analyzing temporal pass data. The data is from a PhD dissertation thesis, which studies Turkish National Football Team's matches for a six sigma application. In this thesis, "e-analysis soccer program"' is used to determine the sequence of the passes (Çobanoğlu, 2015). After we analyzed Turkish National Football Team's attacks with an open-source NodeXL program.

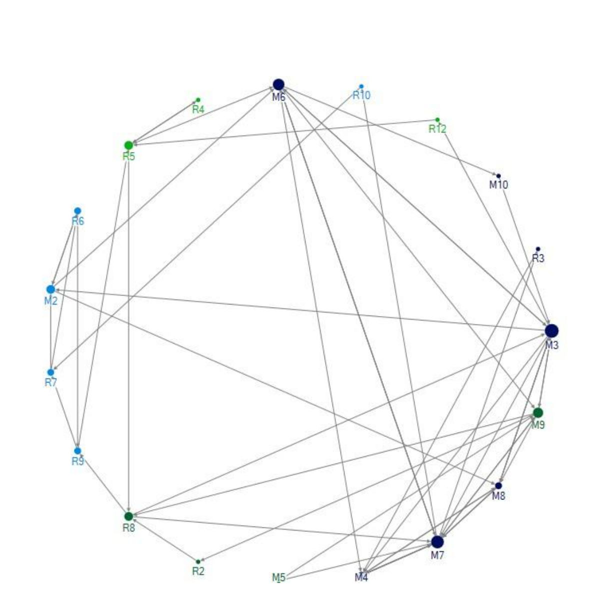
Our sample consists of four attacks from two different matches: a friendly game between Finland and Turkey on 26.05.2012, a Group Elimination game from 2014 World Cup between Hungary and Turkey on 16.10.2012. From Finland match we choose three attacks, where two of them end with goals and one does not. The attack that is chosen from the Hungary game also did not end with a goal.

RESULTS

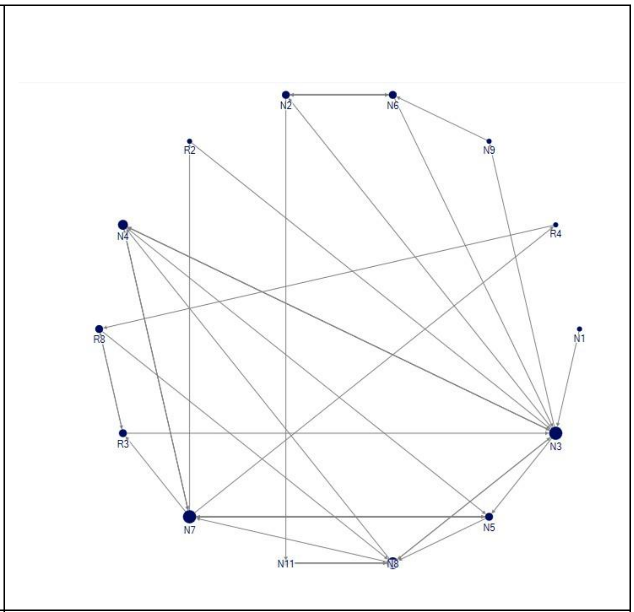
We studied four attacks within the boundaries of macro and micro network measures.

Macro Network Metrics

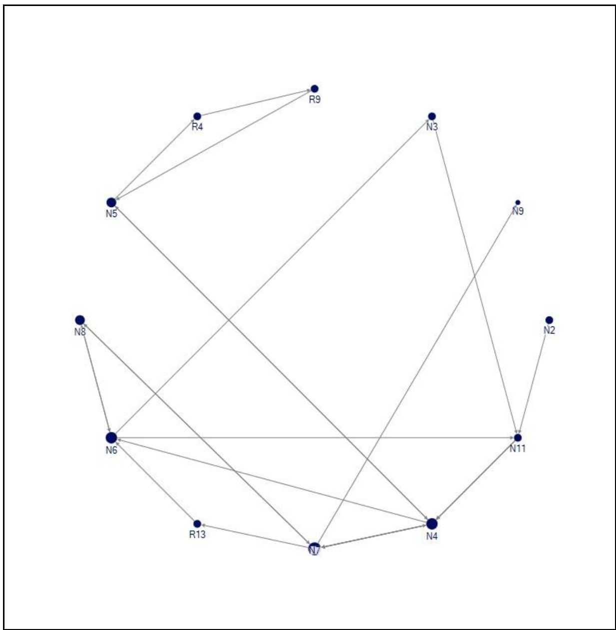
Macro network metrics are listed in Table 1 and Networks of four attacks are drawn as in Figure 1.



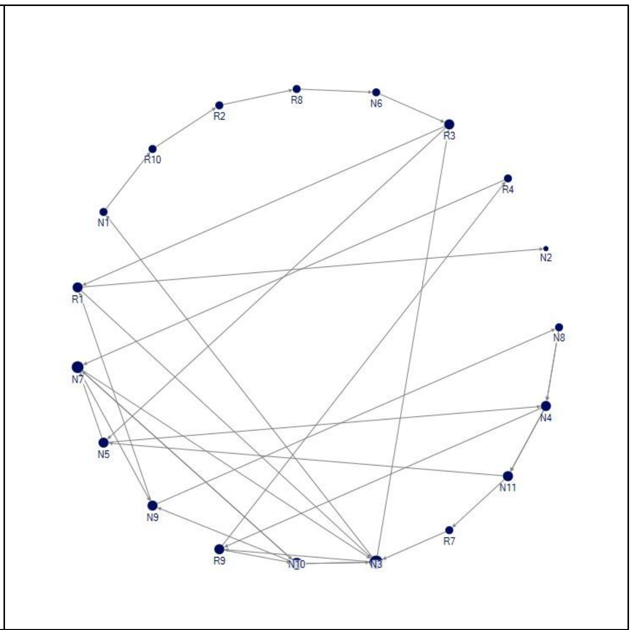
Attack-1, goal scored, 3 groups



Attack-2 , no goal, one group



Attack-3, goal scored, one group



Attack-4, no goal, one group, different opponent team, different network metrics

Figure-1. Networks of four attacks drawn by using out-degrees

*In Table 1 the numbers in the first raw stands for:

1. Is there a goal in the end?
2. Number of vertices
3. Number of edges

4. Diameter: Diameter is the greatest distance between any pair of vertices.
5. Average geodesic distance: It is the mean geodesic (i.e., shortest-path) distance between nodes
6. Density: Actual connections /Potential connections
7. Average in-degree
8. Average out-degree
9. Average Betweenness Centrality: Betweenness centrality is an indicator of a node's centrality in a network. It is equal to the number of shortest paths from all vertices to all others that pass through that node.
10. Average Closeness Centrality: closeness centrality measures how close a vertex is to all other vertices in the graph.
11. Average Eigenvector Centrality: Eigenvector centrality is one method of computing the "centrality", or approximate importance, of each node in a graph.
12. Average PageRank: The underlying assumption is that more important websites are likely to receive more links from other nodes.
13. Average Clustering Coefficient: a clustering coefficient is a measure of the degree to which nodes in a graph tend to cluster together.

Table 1. Macro Network Metrics*

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
Attack1 (Finland)	Yes	19	41	4	1,98338	0,140350	2,526	2,526	19,684	0,027	0,053	1	0,317
Attack2 (Finland)	No	14	40	4	1,83673	0,175824	2,286	2,286	12,714	0,040	0,071	1	0,193
Attack3 (Finland)	Yes	12	24	4	2,18055	0,151515	1,667	1,667	15,167	0,040	0,083	1	0,168
Attack4 (Hungary)	No	19	34	6	2,43213	0,093567	1,684	1,684	28,211	0,022	0,053	1	0,070

At the same time, we studied groups and group by motifs for these four attacks using group by motifs (Dunne and Shneiderman, 2013) feature of NodeXL and they are drawn as Table 2.

Table 2. Groups and group by motifs

	Goal	**Group**	**Group by motifs**
Attack1 (Finland)	Yes	3 Groups	Clique (N3,N6,N7,N9)
Attack2 (Finland)	No	1 Group	Clique (N3,N4,N5,N8)
Attack3 (Finland)	Yes	1 Group	2 Connector (N8, R13)
Attack4 (Hungary)	No	1 Group	No motifs

Similarities between these four attacks

Using some R codes we have calculated a similarity matrix of these four attacks:

```
> A1=c( 1.98338, 0.140350, 2.526, 2.526, 19.684, 0.027, 0.053,1, 0.317)

> A2=c( 1.83673, 0.175824, 2.286, 2.286, 12.714, 0.040, 0.071, 1, 0.193)

> A3=c(2.18055, 0.151515, 1.667, 1.667, 15.167, 0.040, 0.083,1, 0.168)

> A4=c(2.43213, 0.093567, 1.684, 1.684, 28.211, 0.022, 0.053, 1, 0.070)

>  A=rbind(A1,A2,A3,A4)

> dist(A)

          A1        A2        A3
A2  6.981027

A3  4.684156  2.627376

A4  8.625095 15.532510 13.046992
```

Results show that similarity measures changes between 2.62 and 6.98 for first three attacks but if Attack-4 (A4) is taken into consideration, we see that measures changes between 8.62 and 15.53.

There are examples suggesting that change in the network measures over time and how opponent team's effectiveness changes everything. For example, after applying network analysis to the champion team of 2010 FIFA World Cup, it has been found that**,** the effectiveness of the opposing team in negating the Spanish game is reflected in the change of several network measures over time (Cotta et al., 2013). Our findings support this statement.

Micro network metrics

The whole team's triadic relations also can be seen in Figure 2 and this table shows us the triadic relationships in a micro basis. Table 3 also gives us the answer of "which player had the most co-players?" Within this context, N3 has got the most co-players in Attack-1 and Attack-2.

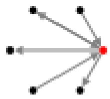
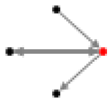
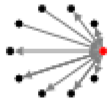
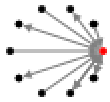
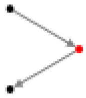
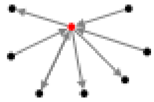
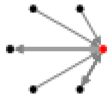
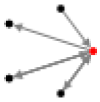
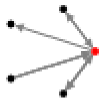
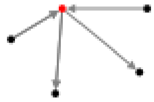
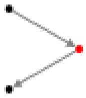
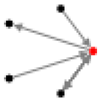
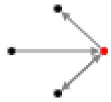
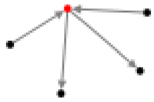
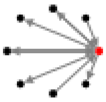
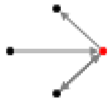
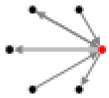
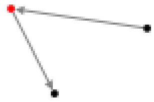
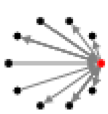
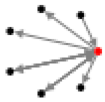
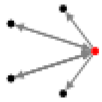
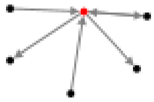
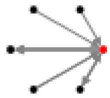
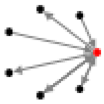
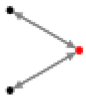
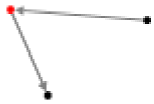
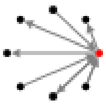
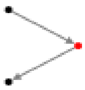
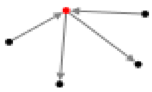
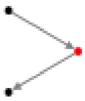
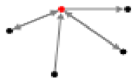
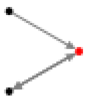
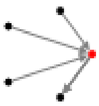
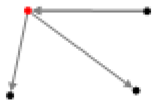
Player	Attack-1 (Goal)	Attack-2 (No goal)	Attack-3 (Goal)	Attack-4 (No Goal)
N2				
N3				
N4				
N5				
N6				
N7				
N8				
N9				
N10				
N11				

Figure 2. Turkish National Team's players' triadic relations(Players may change)

In Attack-1 average betweenness centrality is 19.684. So in Turkish team players 4, 5, 8 and 10 have betweenness scores under this level (see Table 3). We can conclude that regarding to the betweenness centrality they are below the average performance level.

Table 3. Micro metrics for attack 1

Label	In-Degree	Out-Degree	Betweenness Centrality	Closeness Centrality	Eigenvector Centrality	PageRank	Clustering Coefficient	Reciprocated Vertex Pair Ratio
N2	3	3	29,878	0,029	0,058	1,121	0,250	0,200
N3	5	7	54,544	0,036	0,114	1,896	0,208	0,333
N4	4	2	8,928	0,028	0,076	1,095	0,450	0,200
N5	1	1	0,000	0,023	0,036	0,514	0,500	0,000
N6	4	5	44,967	0,034	0,093	1,505	0,214	0,286
N7	5	6	68,283	0,034	0,110	1,923	0,208	0,222
N8	4	2	8,078	0,029	0,080	1,070	0,400	0,200
N9	4	4	28,578	0,031	0,094	1,506	0,286	0,143
N10	1	1	0,000	0,024	0,037	0,512	1,000	0,000
R2	1	1	0,000	0,023	0,030	0,523	0,500	0,000
R3	1	1	0,000	0,022	0,033	0,518	1,000	0,000
R4	1	1	0,000	0,019	0,007	0,367	0,000	1,000
R5	3	3	44,833	0,029	0,040	1,276	0,050	0,200
R6	2	2	1,400	0,023	0,019	0,762	0,333	0,333
R7	2	2	9,606	0,025	0,022	0,997	0,250	0,000
R8	3	3	44,028	0,033	0,073	1,342	0,200	0,000
R9	2	2	21,328	0,027	0,027	0,985	0,167	0,000
R10	1	1	7,772	0,025	0,023	0,543	0,000	0,000
R12	1	1	1,778	0,025	0,027	0,546	0,000	0,000

In Attack-2 average betweenness centrality is 12.714. So in the Turkish team, excluding the goal keeper, players 2, 4, 5, 6, 9 and 11 (see Table 4) have betweenness scores under this level and we may conclude that regarding to the betweenness centrality they are below the average performance level.

Table 4. Micro metrics for Attack 2

Label	In-Degree	Out-Degree	Betweenness Centrality	Closeness Centrality	Eigenvector Centrality	PageRank	Clustering Coefficient	Reciprocated Vertex Pair Ratio
N1	0	1	0,000	0,033	0,031	0,377	0,000	0,000
N2	2	2	6,500	0,037	0,051	0,898	0,167	0,333
N3	6	5	81,689	0,056	0,140	2,408	0,083	0,222
N4	3	3	3,298	0,045	0,105	1,042	0,583	0,500
N5	3	2	3,298	0,045	0,105	1,042	0,667	0,250
N6	2	2	1,667	0,037	0,052	0,901	0,333	0,333
N7	3	5	25,081	0,042	0,108	1,573	0,100	0,333
N8	4	4	31,741	0,050	0,121	1,557	0,267	0,333
N9	1	1	0,000	0,034	0,043	0,633	0,500	0,000
N11	2	1	2,444	0,034	0,038	0,625	0,000	0,500
R2	1	1	3,298	0,038	0,055	0,600	0,000	0,000
R3	2	2	10,797	0,042	0,066	0,848	0,000	0,333
R4	1	1	1,067	0,029	0,035	0,621	0,000	0,000
R8	2	2	7,119	0,036	0,049	0,875	0,000	0,333

In Attack-3 average betweenness centrality is 15.666. So in the Turkish players 2, 3, 8 and 9 (see Table 5) have betweenness scores under this level and we may conclude that regarding to the betweenness centrality they are below the average performance level.

Table 5. Micro metrics for Attack 3

Label	In-Degree	Out-Degree	Betweenness Centrality	Closeness Centrality	Eigenvector Centrality	PageRank	Clustering Coefficient	Reciprocated Vertex Pair Ratio
N2	0	1	0,000	0,032	0,042	0,456	0,000	0,000
N3	1	1	0,000	0,036	0,093	0,747	0,500	0,000
N4	3	3	59,333	0,056	0,144	1,387	0,083	0,500
N5	2	2	36,000	0,042	0,061	1,147	0,167	0,333
N6	3	3	28,000	0,050	0,166	1,711	0,100	0,200
N7	2	4	29,000	0,045	0,105	1,482	0,000	0,500
N8	2	2	2,333	0,038	0,083	0,756	0,000	1,000
N9	1	0	0,000	0,031	0,032	0,465	0,000	0,000
N11	3	1	25,000	0,048	0,137	1,441	0,167	0,000
R4	1	1	0,000	0,030	0,027	0,826	0,500	0,000
R9	1	1	0,000	0,030	0,027	0,826	0,500	0,000
R13	1	1	2,333	0,038	0,083	0,756	0,000	0,000

In Attack-4 average betweenness centrality is 28.210. So in the Turkish players 2, 4, 8, 9, 10 and 11 (see Table 6) have betweenness scores under this level and we may conclude that regarding to the betweenness centrality they are below the average performance level.

Table 6. Micro metrics for Attack 4

Label	In-Degree	Out-Degree	Betweenness Centrality	Closeness Centrality	Eigenvector Centrality	PageRank	Clustering Coefficient	Reciprocated Vertex Pair Ratio
N1	1	1	46,833	0,023	0,034	0,747	0,000	0,000
N2	1	0	0	0,018	0,018	0,417	0,000	0,000
N3	4	4	125,167	0,031	0,125	1,995	0,095	0,143
N4	2	2	20,833	0,023	0,056	1,179	0,083	0,000
N5	2	2	32,333	0,026	0,068	1,162	0,083	0,000
N6	1	1	37,167	0,021	0,019	0,771	0,000	0,000
N7	3	3	32,333	0,026	0,103	1,394	0,150	0,200
N8	1	1	4,333	0,019	0,033	0,649	0,000	0,000
N9	2	2	24,500	0,023	0,076	1,168	0,167	0,000
N10	3	3	7,667	0,025	0,096	1,122	0,250	0,500
N11	1	2	6,000	0,021	0,041	0,926	0,167	0,000
R1	2	2	48,833	0,026	0,073	1,257	0,083	0,000
R2	1	1	6,000	0,015	0,004	0,864	0,000	0,000
R3	2	2	71,500	0,028	0,071	1,234	0,083	0,000
R4	1	1	0,667	0,019	0,046	0,631	0,000	0,000
R7	1	1	9,667	0,023	0,042	0,654	0,000	0,000
R8	1	1	13,167	0,017	0,006	0,845	0,000	0,000
R9	2	2	26,167	0,026	0,081	1,149	0,167	0,000
R10	1	1	22,833	0,018	0,009	0,835	0,000	0,000

DISCUSSION and CONCLUSION

First of all it should be mentioned that our study is a case study and the results of this study should not be generalized. However our findings can be the start point for further researches with larger samples sizes. With the help of network science approach, the most effective players could be found, the most compatible line-up for the future games could be chosen and the opponent team's key players could be analyzed.

Another interesting results of our study is that, although most of the sources emphasize the importance of central players in terms of scoring goals, we have shown that for Turkish National Football Team wing players like N6 and center-midfield players like N3 are more critical. To accept it as a tactic and to call it " Turkish style" we need more evidences and more analysis.

References

Barabasi, A.L. (2002). Linked: The New Science of Networks, First Edition Perseus, Cambridge, MA.

Barabasi, A.L. (2014). Network Science, Cambridge University Press.

Clemente, F.M., Silva, F., Martins, F.M.L., Kalamaras, D., Mendes, R.S. (2015). Performance Analysis Tool for Network analysis on team sports: A case study of FIFA Soccer World Cup 2014, Journal of Sports Engineering and Technology, July, 29.

Cotta, C., Mora, A.M., Merelo-Molina, C., Merelo, J.J. (2011). "FIFA World Cup 2010: A Network Analysis of the Champion Team Play", arXiv:1108.0261v1 [cs.SI].

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.

Çobanoğlu, H.O, (2015). Organizational Change and Applicability of Six Sigma Management Model in Turkish Football Federation, Unpublished Phd. Thesis, Anatolian University, Physical Education and Sports Department, Institute of Health Science, Eskisehir, Turkey.

Dolles, H., Söderman, S., (2005). 'Ahead of the Game -The Network of Value Captures in Professional Football' Erscheinungsort: Tokyo.

Dunne, C., Shneiderman, B. (2013) 'Motif Simplification: Improving Network Visualization Readability with Fan, Connector, and Clique Glyphs', CHI 2013, April 27–May 2, Paris, France.

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

Kooij, R., Jamakovic, A., Kesteren F.V., Koning, T.D., Theisler, I., Veldhoven P., (2009). 'The Dutch Soccer Team as a Social Network", International Network for Social Network Analysis, 29 (1); 4-14.

Mossop, B. (2012), 'Basketball isn't a sport. It's a statistical network', Wired, Arizona state Universty Researcher.

Pena, J.L,, Touchette, H., (2012). 'A network theory analysis of football strategies', arXiv:1206.6904v1 [math.CO].

Valente, W.T, Coronges, K., Lakon, C., Costenbader, E., (2008). “How Correlated Are Network Centrality measures?, PMC, Jan 1; 28 (1); 16-26.

Yamamoto, Y., Yokoyama, K. (2011). Common and Unique Network Dynamics in Football Games, PLos One, 6 (12).