Network Motifs in Football

4 authors, including:

Firat Melih Yilmaz
Dokuz Eylul University

5 PUBLICATIONS   **1** CITATION

Halil Orbay Çobanoğlu
Alanya Alaaddin Keykubat University

37 PUBLICATIONS   **16** CITATIONS

Sandy Cagliyor
Halic University

6 PUBLICATIONS   **1** CITATION

Some of the authors of this publication are also working on these related projects:

Network Science in Sports View project

Sexuel Harassment and Abuse in Sports View project

Network motifs in football

Necmi GÜRSAKAL[1,] Fırat Melih YILMAZ[2],Halil Orbay ÇOBANOĞLU[3], Sandy ÇAĞLIYOR[4]

[1]Fenerbahçe University, Faculty of Economics and Administrative Sciences, İstanbul, Turkey.
[2]Bursa Uludağ University, Faculty of Economics and Administrative Sciences, Bursa, Turkey.
[3]Eskişehir Technical University, Faculty of Sports Sciences, Eskişehir, Turkey
[4]Haliç University, Faculty of Economics and Administrative Sciences, İstanbul, Turkey.
Address Correspondence to HO, Çobanoğlu e-mail: orbay_79@hotmail.com

Abstract

Complex networks often display network motifs and these can be described as subgraphs. Methods for analyzing complex networks promise to be of great benefit to almost all scientific disciplines including sports. In football if we want to disrupt the opponent's game format, we must first be aware of the pass motifs that the team often uses. Determining how to break these motifs will make an important contribution to the success of a team. In this study, 3-nodes and 4-nodes pass motifs of the teams were examined within the frame of a data set of ten games and the most frequent repetitions of these motifs were determined. In addition, we suggest that in a match, the balance can be measured by the correlation between the frequencies of the motif types and there may be an inverse relationship between this correlation and the difference in the goals of the match.

Keywords: Network Science, Network Motifs, Football

INTRODUCTION

Today, sport is now considered as a Complex networks often display network motifs and these can be described as subgraphs that recur in the network much more often than in randomized networks (5). Motif analysis counts the number of the motifs in a network. Network motif detection is mainly consists of enumerating the number of occurrences of each subnetwork of size k, and assessing statistical significance of subnetwork occurrences. Methods for analyzing complex networks promise to be of great benefit to almost all scientific disciplines (13). Therefore motif detection network science tools are valuable for brain, medicine, genetics, sports researches.

Network motifs are circuits of 3 to 6 nodes in directed networks and graphlets are similar to motifs but are defined as undirected networks and motifs proved to be efficient in predicting protein function, protein-protein interactions and development of drug screening techniques (4). It was found that there are several significantly recurring motifs within the brain network, including the two-node reciprocal motif and five classes of three-node motifs and these recurring motifs were distributed in distinct patterns to support intra- and

inter-module functional connectivity (12). For example, also it was found that number 8 type 3-node motif in large scale cortical networks are recurring (9).

Motifs are like bricks in a wall, they are small building blocks of networks. "Milo et al. (7) suggests that network motifs inform on the function of networks (7). In complex networks modules are formed by motifs: They are expected to play proper functions and their disconnections leads to functional decay. From this point of view, also motifs can be thought as disconnection locations. A complex system cannot function properly unless it's motifs and modules function properly.

"Network motifs, overrepresented small local connection patterns, are assumed to act as functional meaningful building blocks of a network and, therefore, received considerable attention for being useful for understanding design principles and functioning of networks (8)." Taking Multiplex network motifs as building blocks of corporate networks, it was found that the motifs reflect known corporate governance practices related to the monitoring of investments and the concentration of ownership and motifs involving financial firms are over-represented amongst the larger and more

complex motifs (10). In other words, corporate governance practices are functioning within the motifs of this system.

Critical points that can be used to control the entire system are called driver nodes (11). It was found that the driver nodes tend to avoid the high-degree nodes (6), the set of driver nodes that can control and move the networks into specific states has been identified and shown to be of low degree (13). The number of driver nodes is determined mainly by the network's degree distribution: Sparse inhomogeneous networks, which emerge in many real complex systems, are the most difficult to control, but that dense and homogeneous networks can be controlled using a few driver nodes (6).

In soccer, "passing is the backbone of the game and forms the basis of important decisions made by managers and owners; such as buying players, picking offensive or defensive strategies or even defining a style of play" (1). It was mentioned that the analysis of the motifs in the pass networks allows us to compare and differentiate the styles of different teams (2). Who are the driver nodes that can control and move the networks into specific states? What are the types of main pass motifs formed by them?

Standard approaches in evaluating and predicting team performance are based on history-related factors such as past victories or defeats (3).

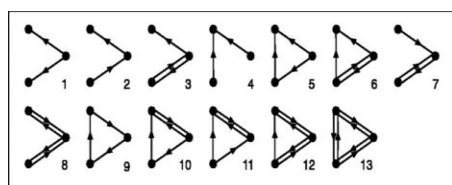**Figure 1.** 2 undirected 3-nodes motifs

Figure 2. 13 directed 3-nodes motifs

Also 13 directed 3-nodes motifs can be seen in Figure 2 above.

Nowadays analyzing data is everything and data is the backbone of monitoring players and preparing team improvement strategies. In contrast with traditional models, in this paper we propose a pass network model to determine the main type of motifs of the games.

Network motifs can be found in a process like this (7):

- Find n-node subgraphs in a real graph.

- Find all n-node subgraphs in a set of randomized graphs with the same in and out degree distributions.

- For each subgraph find a Z score: $Z = \dfrac{N_{real} - N_{rand}}{\sigma_{rand}}$

Subgraphs with high Z scores can be denoted as network motifs.

Number of Motifs and Network Types

Table 1. Number of motifs according to network types			
Network type	3 nodes	4 nodes	5 nodes
Undirected	2	6	21
Directed	13	199	9364

As it can be seen from Table 1, for undirected networks, number of 3-nodes motifs are 2 (Figure 1):
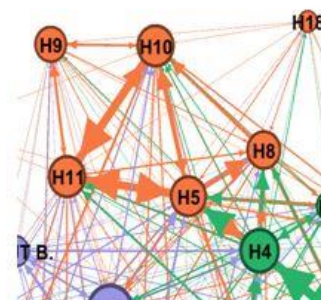

Figure 3. Motifs from Holland-Turkey Match (02-09-2012)

Figure 3 shows us that 3-node motifs like H10-H11-H5 (This is 13th motif in Figure 2) and H5-H8-H10 (This is number 5th motif in Figure 2).
Football match analysis like other fields use large data sets. In a paper data was obtained using a custom Python web crawler from www.squawka.com, and this data set has the last 4 seasons of 6 big European leagues with 8219 matches (1). In another paper it was found that, "The concentration of motif 238 (3-node complete subgraph) in the real Facebook network is 4 times higher than that of the real Twitter network.

Similarly, the concentration of motif 31710 (4-node complete graph) is 33 times more abundant in the real Facebook network than in the real Twitter network" (15). In Figure 4, 3-node and 4-node subgraphs and their motif ids can be seen.
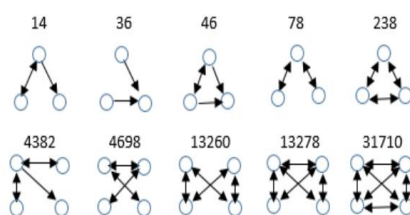

Figure 4. 3-nodes and 4 nodes subgraphs an their motif ids (15)

For directed and undirected 18 networks it was found that (*Italics* are from us):

"We applied this method to identify 3-node and 4-node motifs from the datasets of 18 networks (4 directed and 14 undirected) covering social interactions, co-authorships, web document hyperlinks, neuronal circuitry, protein-protein interactions (PPI), trophic relations in a food web, and others. Presence and absence of enriched motifs provide rich information regarding each type of network relations. *In undirected networks, triangles are enriched in almost all datasets*, suggesting the prevalence of transitivity in diverse networks. However, 4-node structures lacking transitivity – *diamonds and stars – are also enriched in the majority of undirected networks. In directed networks, variations of feed-forward loops are over-represented in the networks* of web document and political weblog hyperlinks as well as neuronal connections. In contrast, the food web is enriched with unidirectional motifs with distinct trophic levels (14)".

MATERIALS & METHOD

The study sample consists of 10 matches of Turkish National Football Team in the 2014 World Cup Qualifiers. 10 matches were analyzed with e-analysis soccer program, the pass actions obtained from the analysis were transferred into Excel files. The Excel pass files were processed with the open-source program Gephi to obtain the network of the matches and the measurements of networks' motifs.

RESULTS

Football Motifs

In our analysis we will use as our data set ten national teams matches and these matches can be seen in Table 2. Using R and R package igraph 3-nodes and 4-nodes pass motifs are found. As an example let's take a match between Turkey and

Estonia and apply the codes below fist for the Turkish team and then for the Estonian team:

```
> data<- read.csv("turkiye_estonya1.csv", header=T, as.is=T)
> data1 <- subset(data, select = -c( Type :Weight ))
> library('igraph')
>network<- graph_from_data_frame(d=data1, directed=T)
> plot(network)
```
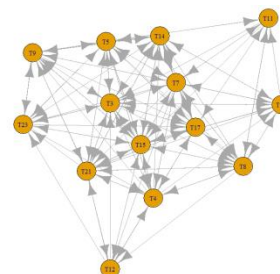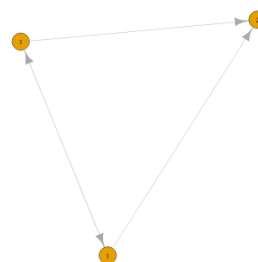

Figure 5. Pass network of the Turkish team (131 passes)

```
> motif=graph.motifs(network,size=3)
> motif
```

[1] NA NA 2 NA 4 34 3 1 7 20 47 1 16 4 83 103

```
> plot(graph.isocreate(size=3,number = 13))
```


Figure 6. The mode motif of The Turkish team

Now for the Estonian team:

```
> data<- read.csv("türkiye_estonya1.csv", header=T, as.is=T)
> data1 <- subset(data, select = -c( Type :Weight ))
> library('igraph')
> network <- graph_from_data_frame(d=data1, directed=T)
> plot(network)
```
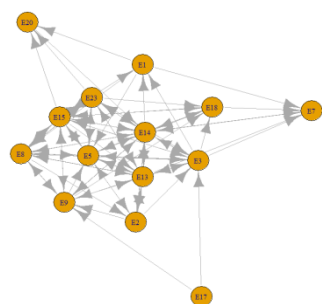
Figure 7. Pass network of the Estonian team (84 passes)

> motif=graph.motifs(network,size=3)

> motif

 [1] NA NA 9 NA 17 30 12 4 1 38 25 2 8 7 38 34

> plot(graph.isocreate(size=3,number = 7))
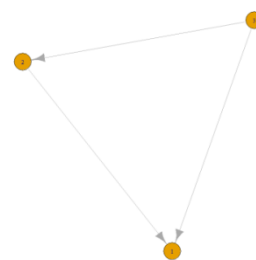

Figure 8. The first mode motif of The Estonian team
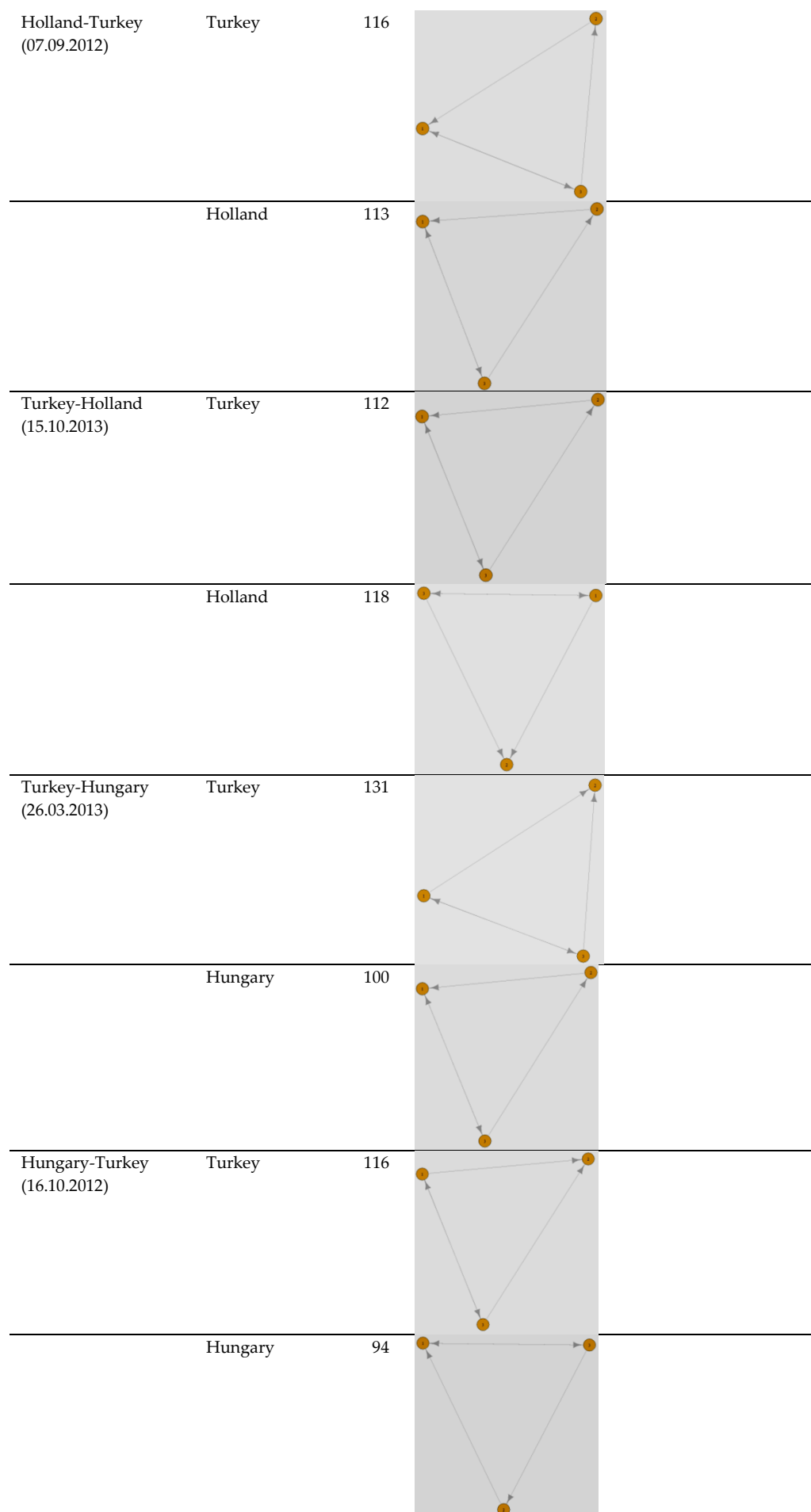
> plot(graph.isocreate(size=3,number = 12))

Figure 9. The second mode motif of The Estonian team

Table 2. 3-nodes motifs results
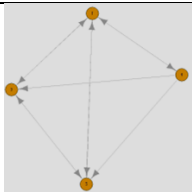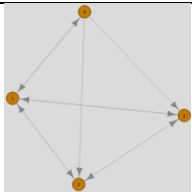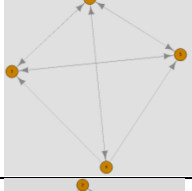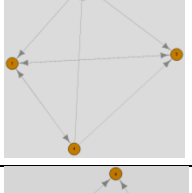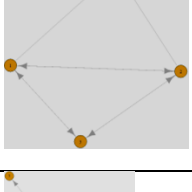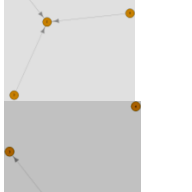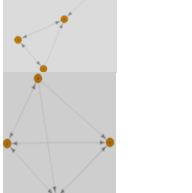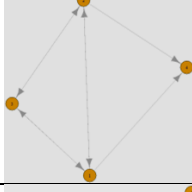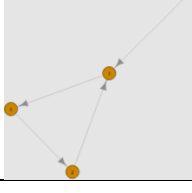
Matches	Team	Number of passes	Mode motif 1	Mode motif 2
Turkey-Estonia (11.09.2012)	Turkey	131		
	Estonia	84		
Estonia-Turkey (11.10.2013)	Turkey	110		
	Estonia	111		

Turkey-Andorra (06.09.2013)	Turkey	132	
	Andorra	68	
Andorra–Turkey (22.03.2013)	Turkey	120	
	Andorra	65	
Romania-Turkey (10.09.2013)	Turkey	108	
	Romania	122	
Turkey-Romania (12.10.2012)	Turkey	118	
	Romania	116	

Holland-Turkey (07.09.2012)	Turkey	116	
	Holland	113	
Turkey-Holland (15.10.2013)	Turkey	112	
	Holland	118	
Turkey-Hungary (26.03.2013)	Turkey	131	
	Hungary	100	
Hungary-Turkey (16.10.2012)	Turkey	116	
	Hungary	94	

As a result, in Table 2 for 3-nodes considering Figure 2, type 10 has been seen as mode 11 times, type 6 has been seen as mode 8 times.

Table 3. 4-nodes motifs results

Matches	Team	Number of passes	Mode motif 1	Mode motif 2
Turkey-Estonia (11.09.2012)	Turkey	131		
	Estonia	84		
Estonia-Turkey (11.10.2013)	Turkey			
	Estonia			
Turkey-Andorra (06.09.2013)	Turkey			
	Andorra			
Andorra–Turkey (22.03.2013)	Turkey			
	Andorra			

Romania-Turkey (10.09.2013)	Turkey	
	Romania	
Turkey-Romania (12.10.2012)	Turkey	
	Romania	
Holland-Turkey (07.09.2012)	Turkey	
	Holland	
Turkey-Holland (15.10.2013)	Turkey	
	Holland	
Turkey-Hungary (26.03.2013)	Turkey	

Hungary

Hungary-Turkey   Turkey
(16.10.2012)

Hungary
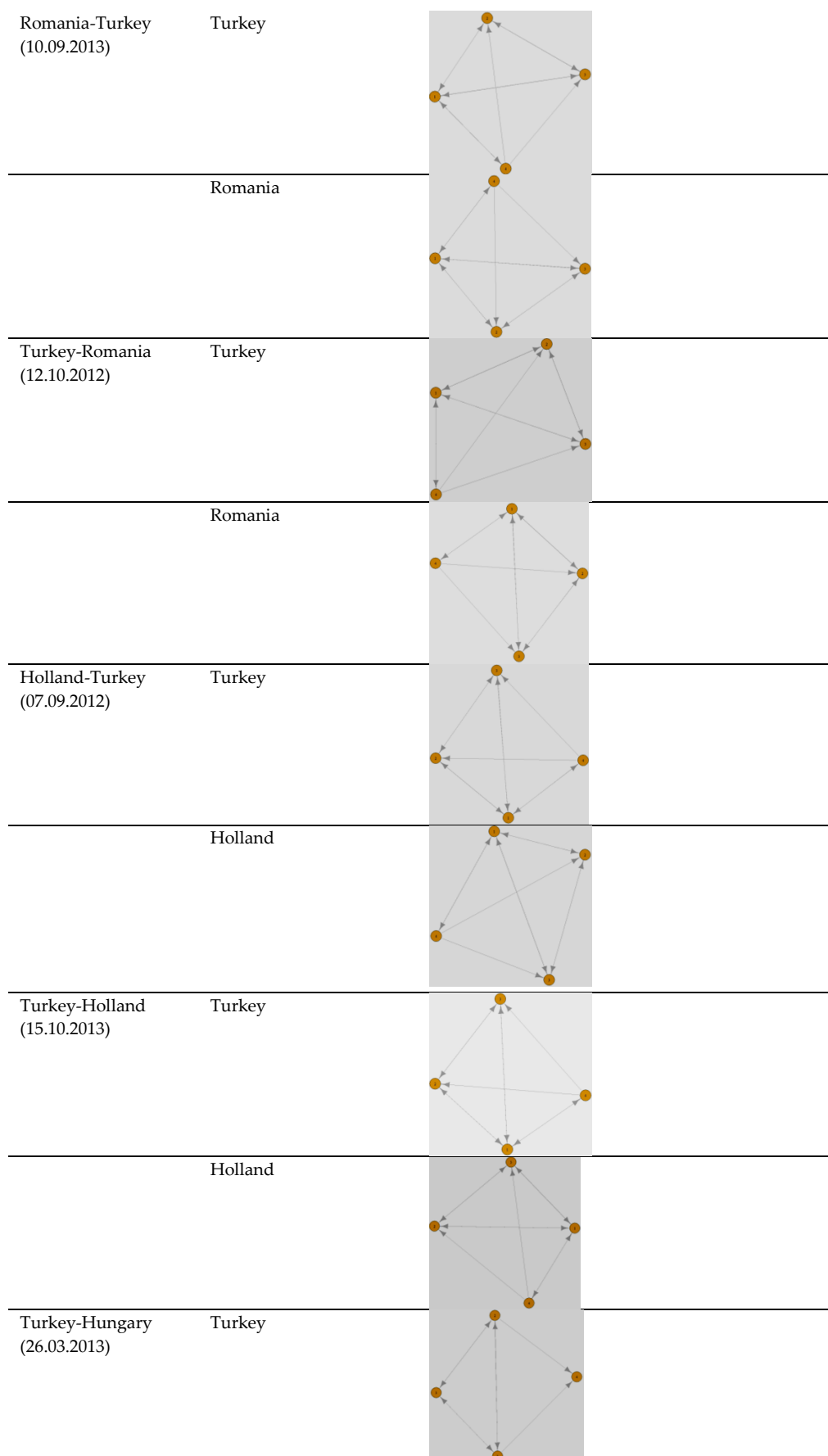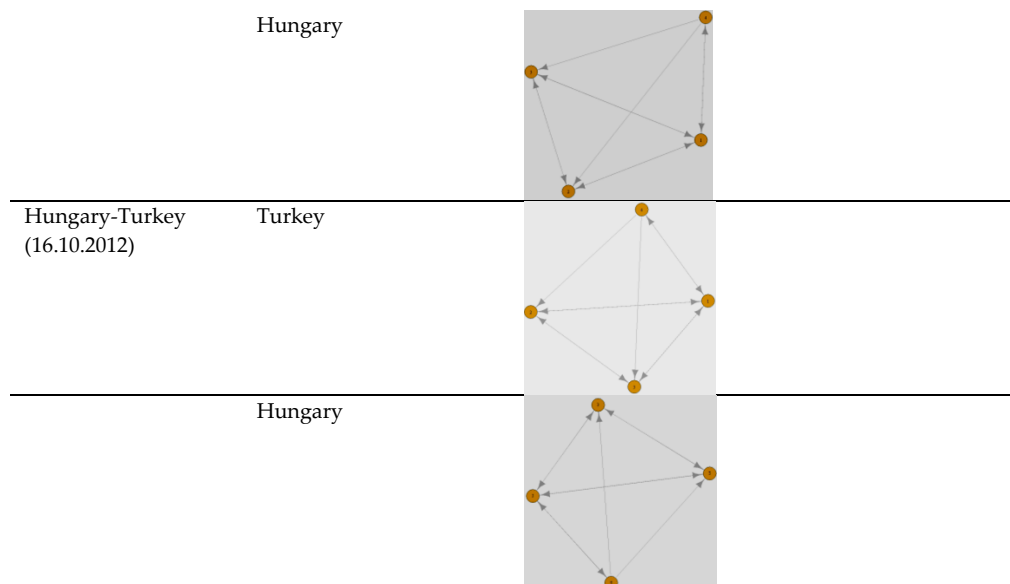
4-nodes motifs results can be seen in Table 3 and Figure 10 shows us that mode motif type in 4-nodes. The motif in Figure 10 has been repeated 14 times between 20 motifs

The Balance Concept of a Match

In a match the "balance" concept corresponds to the draw, the win or the defeat is the deterioration of this balance. The distribution of n-nodes motif types of a team must be different from the corresponding distribution of the competing team in order to disrupt the balance. In other words, to the extent that the harmony of the motifs of the two teams playing the match is impaired, the balance is impaired to that extent. Undoubtedly, the balance will be disrupted and the balance can be disrupted by a single shot kicked from the centre, regardless of the pass and pass motifs.

However, in general the deterioration of the balance with goals is not independent of the pass network and its motifs. We propose to measure this conceptual balance concept operationally with the correlation coefficient between the frequencies of the motifs of the two teams in a match. For this relation a high correlation coefficient shows that the harmony between the frequencies of the motifs of the two teams is high and there is a balance; on the contrary, a low correlation coefficient tells us that there is no balance.

In accordance with our small data set, correlations between the frequencies of the tripartite motifs, except for Turkey-Andorra match are significantly high correlation coefficients. 5-0 Turkey-Andorra match resulting correlation

Figure 10. Mode motif type in 4-nodes motifs

coefficient of -0.081. The correlation coefficient for 2-0 Turkey-Andorra match was 0.205. For the 4-nodes motifs, the correlation coefficients for Turkey-Andorra matches were 0.03 and 9.25; the correlation coefficient between the frequencies of 4-nodes motifs of Turkey-Holland match was 0.83. In Estonia matches, these coefficients were 0,735 and 0,679.

As a result for our data set: The correlation coefficient between the frequencies of the 3-nodes motif types of the two teams in a match and the goal difference was -0,556 (p-value = 0.095) and this coefficient of correlation was significant at the level of significance of 0.10. In other words, there is an indistincly reversed relationship between the harmony of 3-nodes motifs and the goal difference of the match. As the compliance increases, the goal difference is decreasing.

As another result: The correlation coefficient between the frequencies of the 4-nodes motif types of the two teams in a match and the goal difference was -0,579 (p-value = 0.08) and this coefficient of correlation was significant at the level of significance of 0.10. If we exclude the Turkey-Hungary match form our data set, this coefficient increases at -0.762 (p-value = 0.017) and in this case

this relation will be statistically significant at the 0.05 significance level.

We actually know that n = 25 is required for correlation analysis. Therefore, while the inverse relationship hypothesis between the correlation coefficients and the difference between the goals using 0.05 significance level cannot be confirmed by the small sample we have, we maintain the expectation that this hypothesis can be confirmed if the sample size grows.

DISCUSSION

In our study, it was determined that which 3-nodes and 4-nodes motif types were the mod motif types within the framework of a 10-match data set. As a result, in Table 2 for 3-nodes considering Figure 2, type 10 has been seen as mode 11 times and type 6 has been seen as mode 8 times.

Also, within the same dataset, the inverse relationship hypothesis between the correlation coefficients of the motif types for the two teams and the difference between the goals using 0.05 significance level cannot be confirmed by the small sample we have, but we maintain the expectation that this hypothesis can be confirmed if the sample size grows.

If we want to disrupt the opponent's game format, we must first be aware of the pass motifs that the team often uses. Determining how to break these motifs will make an important contribution to the success of a team. Examination of the mode pass types of the teams that are very successful and consistently winning is important for the other teams who want to achieve this success. The most important deficiency of the pass network analysis method is that this analysis does not take into account the position of the players in the field in any way.

REFERENCES

1. Bekkers J, Dabadghao S. Flow Motifs in Soccer: What can passing behavior tell us? MIT Sloan Sports Analytics Conference, 3-4 March 2017. Hynes Convention Center, from http://www.sloansportsconference.com/wp-content/uploads/2017/02/1563.pdf

2. Buldu JM, Busquets J, Martinez JH, Herrera-Diestra JL, Echegoyen I, Galeano J, Luque J. Using network science to analyse football passing networks: dynamics, space, time and the multilayer nature of the game. *Frontiers in Psychology*. 2018; 9: 1900.

3. Cintia P, Rinzivillo S, Pappalardo L. A network-based approach to evaluate the performance of football teams. Machine Learning and Data Mining for Sports Analytics Workshop (MLSA'15): ECML/PKDD Conference 2015; September.

4. Csermely P., Korcsmáros T., Kiss HJM., London, G., Nussinov, R. Structure and dynamics of molecular networks: A novel paradigm of drug discovery a comprehensive review. *Pharmacology & Therapeutics*. 2013; 138(3): 333-408.

5. Itzkovitz S, Alon U. Subgraphs and network motifs in geometric networks. *Physical Review*. 2005; E 71(2 Pt 2): 026117.

6. Liu YY, Slotine JJ, Barabasi AL. Controllability of complex networks. *Nature*. 2011; 473: 167-173.

7. Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U. Network Motifs: Simple Building Blocks of Complex Networks. *Science*. 2002; 298(5564): 824-827.

8. Schmidt C, Weiss T, Komusiewicz C, Witte H, Leistritz L. An Analytical Approach to Network Motif Detection in Samples of Networks with Pairwise Different Vertex Labels. *Computational and Mathematical Methods in Medicine*. 2012; Article ID 910380.

9. Sporns O, Kotter R. Motifs in Brain Networks, *Plos Biol*. 2004; 2(11): e369.

10. Takes FW, Kosters WA, Witte B, Heemskerk EM. Multiplex network motifs as building blocks of corporate Networks. *Applied Network Science*. 2018; **3**(39): 1-22.

11. Trafton A. (2011, May 12). *How to control complex networks: New algorithm offers ability to influence systems such as living cells or social networks*. Retrieved September 20, 2018, from http://news.mit.edu/2011/network-control-0512.

12. Wei Y, Liao X, Yan C, He Y, Xia M. Identifying topological motif patterns of human brain functional Networks. *Hum Brain Mapp*. 2017; 38(5): 2734-2750.

13. Yaveroğlu ÖN. Malod-Dognin N, Davis D, Levnajic Z, Janjic V, Karapandza R, Stojmirovic A, Przulj N. Revealing the Hidden Language of Complex Networks. *Scientific Reports*. 2014; 4: 4547.

14. Yeang CH, Huang LC, Liu WC. Recurrent structural motifs reflect characteristics of distinct Networks, Proceedings, The 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Istanbul, Turkey, from http://www.stat.sinica.edu.tw/chyeang/Yeang_network_motifs_final.pdf

15. Zhang LL, Thomas W, Ashiabar A. Comparing predictive powers of Network Motif Distribution and structure of Overlapping Communities. Class Projects: 2014; CS224W: Social and Information Network Analysis. Stanford University, from http://snap.stanford.edu/class/cs224w-2014/projects2014/cs224w-69-final.pdf