Journal of Physical Education and Sport ® (JPES), 16(3), Art 148, pp. 940 - 944, 2016
online ISSN: 2247 - 806X; p-ISSN: 2247 – 8051; ISSN - L = 2247 - 8051 © JPES

Original Article

Development of sports network analysis: Methodological considerations

FILIPE MANUEL CLEMENTE[1,2], FERNANDO MANUEL LOURENÇO MARTINS[2,3], FRUTUOSO SILVA[2,4]
[1]Instituto Politécnico de Viana do Castelo, Escola Superior de Desporto e Lazer, Portugal
[2]Instituto de Telecomunicações, Delegação da Covilhã, Portugal
[3]Polytechnic Institute of Coimbra, ESEC, Department of Education, IIA, RoboCorp, ASSERT, Portugal
[4]University of Beira Interior, Portugal

Published online: September 26, 2016
(Accepted for publication September 10 2016)
DOI:10.7752/jpes.2016.03148

Abstract:
The understanding of dynamic and complex systems requires a multi-directional approach toward the whole system. New mathematical approaches have been proposing new tools and techniques to understand the collective dynamics in team sports. Nevertheless, to ensure the quality of the techniques it should be considered the data collecting procedures. For that reason, the aim of this article is to suggest a set of methodological considerations to optimize the match analysis based on network.
Keywords: Graph theory; Network analysis; Team sports; Match Analysis.

Introduction

Collective behaviour in team sports and the possibilities of match analysis

Networks emerges from the interactions between teammates in the match (Peña & Touchette, 2012). By using some analysis techniques, it is possible to determine the style of play (Clemente et al., 2013a). The information available from the reality of the game may help to characterize the collective processes and also to predict next events (Bartlett, Button, Robins, Dutt-Mazumder, & Kennedy, 2012). Moreover, it is also possible to re-organize the strategy of the team and tactical behaviour of players (Clemente, Couceiro, Martins, & Mendes, 2015), and even optimize the sports training based on the knowledge of strengths and weaknesses of the team and an opponent (Clemente, Couceiro, Martins, Mendes, & Figueiredo, 2013b).

Several methods of and different approaches to match analysis can be used to provide information for coaches (Carling, Williams, & Reilly, 2005; Hughes & Franks, 2005). From that data, it is possible to classify such information into categories (Clemente, Couceiro, Martins, Mendes, & Figueiredo, 2014): (i) notational analysis based on individual actions; (ii) tactical analysis based on observational methods; (iii) computational metrics to characterize the spatio-temporal relationship within and between teams; and (iv) observational methods to classify the level of cooperation and patterns within a team. Three main types of observational methods have been used to classify the level of cooperation: (i) temporal patterns (Jonsson et al., 2006); ii) neural network (Grunz, Memmert, & Perl, 2012); and iii) social network analysis based on graph theory (Lusher et al., 2010). The last example, social network analysis, is a very useful and user-friendly method of applying social network metrics to any game and competition level. The method involves observation and codifying of team interaction and processing the data using software. For a better understanding about the concept of *social network analysis*, the following section presents the global approach and its use in sport sciences.

Social Network Analysis and Graph Theory: Concepts and Definitions

The Social Network Analysis (SNA) is based on Graph Theory (Barnes & Harary, 1983), a mathematical study of sets of nodes connected by lines. The techniques model pairwise relations between the vertices. Some understood fundamentals of graph theory are directed graphs, undirected graphs, and weighted graphs (digraphs) (Pavlopoulos et al., 2011).

In the following presented elementary concepts on Graphs Teory

Definition 1. (Gross & Yellen, 2004)

A graph G=(V,E) can be characterized as two sets V and E, where V is a set of vertices (or nodes), E is a set of edges and each edge has a set of one or two vertices associated to it, which are called its endpoints (or neighbors) and an edge is said to joint its endpoints.

Definition 2. (Gross & Yellen, 2004)

A graph that has no loops and includes no more than one edge between a pair of nodes is called a simple graph.

Corresponding Author. FILIPE MANUEL CLEMENTE, E-mail: filipe.clemente5@gmail.com

The Definition 1 represent the class of graphs called of undirected graph.

The other broad class are the directed graphs that we show the following in Definition 3.

Definition 3. (Gross & Yellen, 2004)

A directed graph (or digraph) is a graph each of whose edges is directed shut that a direct edge (or arc) is an edge that linked in the initial node to the terminal node.

Definition 4. (Gross & Yellen, 2004).

A digraph that has no loops and includes no more than one arc that linked in the initial node to the terminal node is called a simple digraph.

Definition 5. (Gross & Yellen, 2004)

In the case of a any graph (digraph) with all connections measured between any two nodes is called weighted graph (digraph).

Definition 6. The $n \times n$ matrix $A = [a_{ij}]$ is the adjacency matrix of a graph or digraph such that $a_{ij} = 1$ if $(n_i, n_j) \in V$ or $a_{ij} = 0$ if $(n_i, n_j) \notin V$, $i, j = 1, \dots, |V|$. In the case of weighted graphs or digraphs $a_{ij} = w_{ij}$ if $(n_i, n_j) \in V$ or $a_{ij} = 0$ otherwise.

A social network can be understood as a finite set or sets of actors that relate to each other (Wasserman & Faust, 1994). Network data are defined by actors (nodes) and by relations (edges or arcs) between them (Scott, 2000). Generally, network analysis focuses on the overall relationship between the group of nodes and not only on one node in particular (Hanneman & Riddle, 2005). The relations between nodes can be defined by the edges or arcs (social ties) that link them. A defining feature of a social tie is an established linkage between a pair of nodes (Wasserman & Faust, 1994).

Some additional network concepts has been explained by Wasserman and Faust (1994). In network analysis, a dyad is the lowest level of interaction that represents a tie between two nodes. A triad represents a relationship between three nodes and the possible ties between them. A subgroup is any subset of nodes and all possible ties between them. The group represents the full system of nodes and their possible ties. A relation is the collection of ties of a specific kind between nodes of a group.

For a global perception about the strategy and tactical behaviour of players, SNA can be a very useful approach, giving coaches appropriate information about the specific characteristics of players' interactions (Passos et al., 2011). Social network analysis allows coaches to apply the techniques in official matches or even in daily training sessions. In fact, one of the main advantages of network approaches is that players need not wear monitoring devices, thus SNA can be useful during the official matches.

Collecting and processing the data: Building the adjacency matrix based in the observation

For this survey, we used a dataset observed from an official soccer match, the FIFA World Cup 2014 final. We collected data on the national teams of Germany and Argentina and then codified it, utilizing an expert in match analysis with more than 5 years of professional practice.

To ensure high standards of quality and reliability of the data, we tested the observers using a test-retest process (Costa, Garganta, Greco, Mesquita, & Seabra, 2010). Usually, the test-retest must have an interval between two and three weeks to minimize the observer's familiarity with the task (Altman, 1991). The data tested must have more than 10% of overall dataset that will be used (Tabachnick & Fidell, 2007). After data collection, the reliability can be tested using the statistical technique of Kappa of Cohen coefficient (Robinson & O'Donoghue, 2007). When more than one observer is working, the test-retest process should comply with the inter-observer and intra-observer agreement coefficients. If the coefficient results are higher than 0.61, the reliability indicates a conventional level of acceptance (Landis & Koch, 1977). If minimum values are not achieved, another round of observers' training should be done.

Observation and Codification

After considering the important issue of data collection reliability, the first task to complete is the observation and codification of interactions. The following step defines the code for each node. In this application, a node represents each player. For the majority of team sports, the players occupy specific strategic positions and have a tactical role in the team. Nevertheless, the possibilities to replace the player for another during the match may increase the number of nodes, making it confusing to understand the tactical role. One possible solution is to classify the players by the name so when a new player enters the match, a new node is added to the graph. This alternative can be a problem for teams that put new players in permanently and have no limit on substitutions. To avoid this issue we can codify players by their strategic position on the field. This solution ensures that only the tactical roles matters independently of the player in question. This can be very useful to identify a team's style of play and also some specific tactical characteristics.

Both solutions have pros and cons. When each player is identified independently of their tactical role, it is easier to collect data and it is possible to increase the accuracy of centrality metrics. Nevertheless, because the tactical role is missing, the coach's main information may not be enough. However, when the player is classified by their tactical role and replaced, the new player assumes the same classification. When the team is observed in a strategic and tactical fashion, it is possible to identify some properties of the team that are not exclusively dependent from the individual player. Thus, team patterns of play and team strategic plans can be better identified. Nevertheless, tactical role is hardest to observe in practice because teams may change their strategy and even players change their strategies permanently, thus the observer must be much trained for this. Moreover, the information about the influence of each individual player may decrease because the 'names' are 'lost' in the tactical role. An example can be observed in Figure 1.

For Figure 1a, all the players were inserted in weighted digraph. In the FIFA World Cup, only three players can be replaced, although there are team sports in which the entire team can be replaced, thus making hard to understand the graph and the following network metrics results. Figure 1b best illustrates the tactical behaviour and play style of the team, especially when a great number of replacements occur during a match. For our survey, we used the second approach, making use of the method showed in Figure 1b.

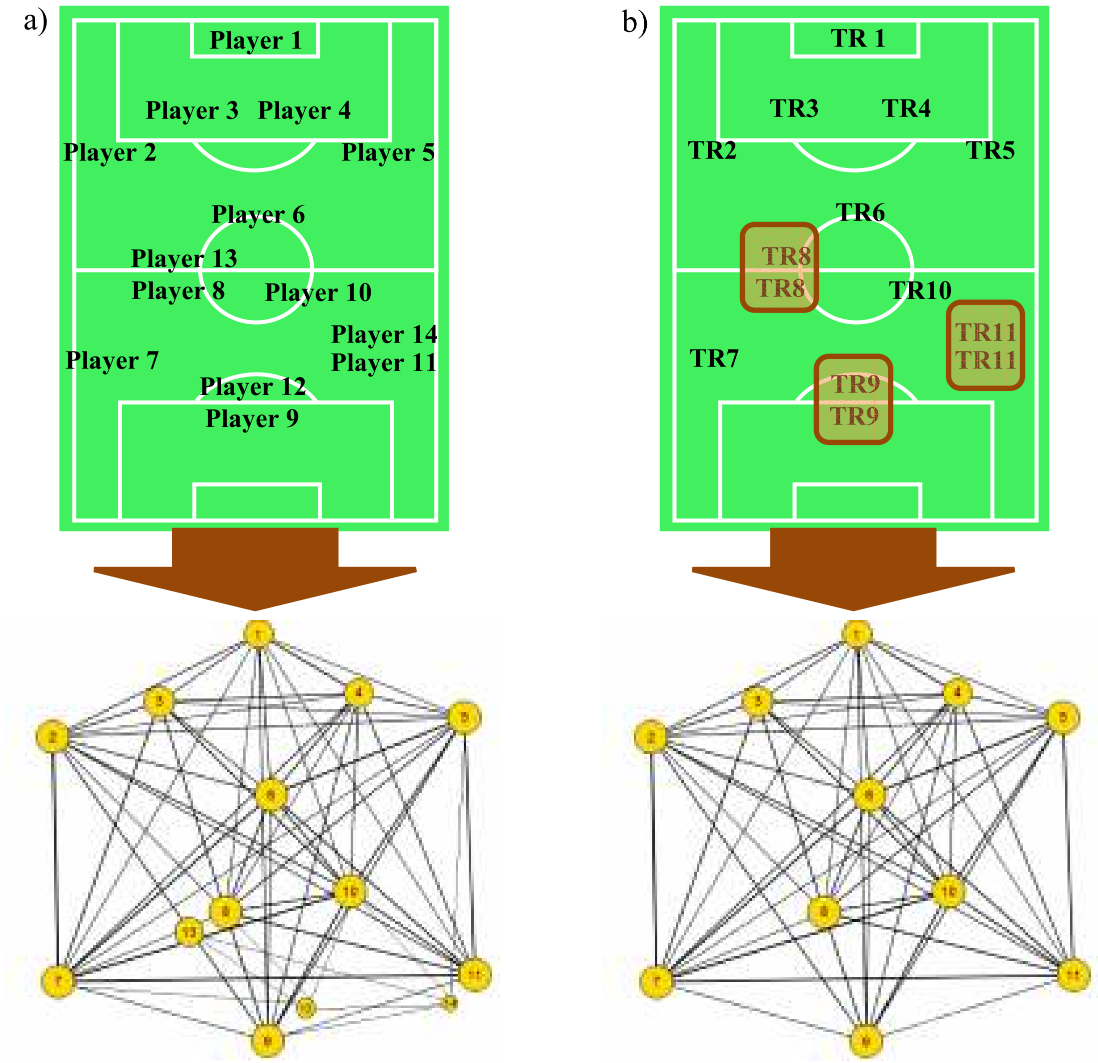
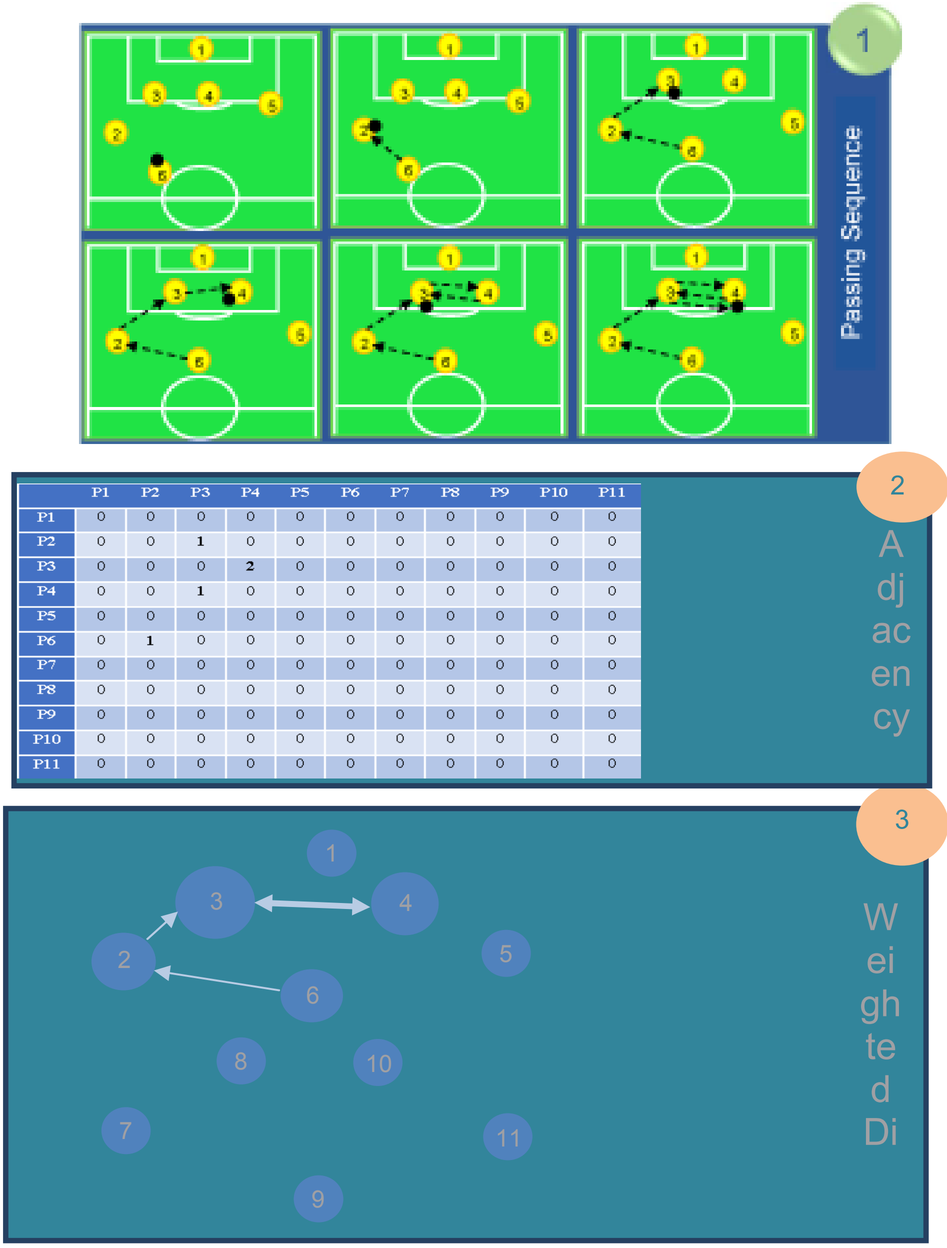


Fig. 1. Examples of players coding and weighted digraphs: a) collecting the analysis of individual players; b) collecting the analysis by tactical role.

4.2. Building the Adjacency Matrix

After observing and codifying the players and their tactical roles, we began to graph the data using nodes (in our case soccer players) and edges. The edges assume many technical indicators such as positional changes, defensive marking or ball recovery. The easiest indicators to observe were the passes. In soccer, the most common individual action performed between teammates is the pass, which keeps possession of the ball while moving it forward to score (Clemente, Couceiro, Martins, Ivanova, & Mendes, 2013). We selected the pass as a linkage indicator, which completed the cycle of data collection, i.e., nodes, and links between nodes, were selected.

Before building the adjacency matrix, we determined how to record the data. During the soccer match, more than 300 passes can be performed by the same team (Lago & Martín, 2007). The easiest way is to split the data recording by attacking units (Passos et al., 2011). An attacking unit can be classified by noting the moment of ball recovery, followed by a set of passes without losing the ball. In the case of ball lost, this attacking unit ends and we generated a specific adjacency matrix of the attacking unit. The criteria for building the adjacency matrix included codifying the passes between teammates as 1 (one) and no passes as 0 (zero). If more than 1 pass was performed in the same direction (for instance, from player 1 to player 2) the code must represent the volume of passes. Passes may also represent the direction. In the majority of the cases network lead to weighted digraphs. In other cases of interactions (other linkage indicators) can be used other approach where the direction or the weight it is not relevant, thus can be represented by graphs and not digraphs. An illustrative representation can be observed in the following Figure 2.



	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11
P1	0	0	0	0	0	0	0	0	0	0	0
P2	0	0	1	0	0	0	0	0	0	0	0
P3	0	0	0	2	0	0	0	0	0	0	0
P4	0	0	1	0	0	0	0	0	0	0	0
P5	0	0	0	0	0	0	0	0	0	0	0
P6	0	1	0	0	0	0	0	0	0	0	0
P7	0	0	0	0	0	0	0	0	0	0	0
P8	0	0	0	0	0	0	0	0	0	0	0
P9	0	0	0	0	0	0	0	0	0	0	0
P10	0	0	0	0	0	0	0	0	0	0	0
P11	0	0	0	0	0	0	0	0	0	0	0

Fig. 2. Process of data collecting: a) observation of passing sequence; b) building the adjacency matrix of passing sequence; c) weighted digraph based on adjacency matrix.

Fig. 2c represents the weighted digraph of the adjacency matrix. Bigger nodes meant that the player performed more passes. Larger arcs meant that the intensity of linkage is greater between the nodes. A larger arc was found between nodes 3 and 4 (Figure 2c). Moreover, player 3 was the player that performed and received more passes. No arcs between nodes represent no recorded connections. This graphical representation helps coaches to better understand the overall connections between teammates.

Practical Applications

Each adjacency matrix represents only one attacking unit, even though per each match, more than 100 attacking units can be recorded. Thus, researchers and coaches can apply the adjacency matrix in many different ways. If a more specific analysis is necessary, a temporal analysis may be performed by adding 5 minutes to the adjacency matrix. If only a global analysis is needed, a representation per each half or per the complete match can be made. In summary, the adjacency matrices must be added to a final adjacency matrix once analysts decide which the kind of analysis they will perform. Despite the usefulness of graphical representation, it is not enough to guarantee a full understanding about either the weighted digraph characteristics, or about how each player contributes to the team and their prominence on it. To do that, it is necessary apply a set of social network metrics that have been developed and used in other scientific areas.

Acknowledgements

This work was supported by the FCT project PEst-OE/EEI/LA0008/2013. This article was made in the scope of uPATO project granted by Instituto de Telecomunicações.

References

Barnes, J. A., & Harary, F. (1983). Graph theory in network analysis. *Social Networks*, *5*, 235–244.

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., Couceiro, M. S., Martins, F. M. L., Ivanova, M. O., & Mendes, R. (2013). Activity profiles of soccer players during the 2010 world cup. *Journal of Human Kinetics*, *38*(September), 201–11. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3827759&tool=pmcentrez&rendertype=abstract

Clemente, F. M., Couceiro, M. S., Martins, F. M. L., & Mendes, R. S. (2015). Using Network Metrics in Soccer: A Macro-Analysis. *Journal of Human Kinetics*, *45*, 123–134. doi:10.1515/hukin-2015-0013

Clemente, F. M., Couceiro, M. S., Martins, F. M. L., Mendes, R. S., & Figueiredo, A. J. (2014). Practical Implementation of Computational Tactical Metrics for the Football Game: Towards an Augmenting Perception of Coaches and Sport Analysts. In Murgante, Misra, Rocha, Torre, Falcão, Taniar, … Gervasi (Eds.), *Computational Science and Its Applications* (pp. 712–727). Springer.

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.

Gréhaigne, J. F., & Godbout, P. (1995). Tactical Knowledge in Team Sports From a Constructivist and Cognitivist Perspective. *Quest*, *47*, 490–505.

Gréhaigne, J. F., Godbout, P., & Zerai, Z. (2011). How the "rapport de forces" evolves in a football match: the dynamics of collective decisions in a complex system. *Revista de Psicología Del Deporte*, *20*(2), 747–765.

Gréhaigne, J. F., Richard, J. F., & Griffin, L. (2005). *Teaching and learning team sports and games*. New York, USA: Routledge Falmar.

Gross, L., & Yellen, J. (2004). *Handbook of Graph Theory*. New York, USA: CRC press.

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

Hughes, M., & Franks, I. (2005). Analysis of passing sequences, shots and goals in soccer. *Journal of Sports Sciences*, *23*(5), 509–514.

Jonsson, G. K., Anguera, M. T., Blanco-Villaseñor, Á., Losada, J. L., Hernández-Mendo, A., Ardá, T., … Castellano, J. (2006). Hidden patterns of play interaction in soccer using SOF-CODER. *Behavior Research Methods*, *38*(3), 372–381.

McGarry, T. (2005). Soccer as a Dynamical System: Some Theoretical Considerations. In T. Reilly, J. Cabri, & D. Araújo (Eds.), *Science and Football V* (pp. 570–579). London and New York: Routledge, Taylor & Francis Group.

Passos, P., Davids, K., Araújo, D., Paz, N., Minguéns, J., & Mendes, J. (2011). Networks as a novel tool for studying team ball sports as complex social systems. *Journal of Science and Medicine in Sport*, *14*(2), 170–176.

Peña, J. L., & Touchette, H. (2012). A network theory analysis of football strategies. In *arXiv preprint arXiv* (p. 1206.6904).