Chapter 16
A Methodology for the Analysis of Soccer Matches Based on PageRank Centrality

Julio Rojas-Mora, Felipe Chávez-Bustamante, Julio del Río-Andrade, and Nicolás Medina-Valdebenito

Abstract Data analysis in sports has adopted many different approaches given its usefulness in quantitative and objective management. Several advances have been made considering the researches and technologies that have been developed up until now. It is possible to find many complex methodologies of sport performance analysis in order to have as much as information as possible to achieve success. Therefore, a wide variety of options are available for sport managers, coaches or anyone interested, including advances on information systems, data mining, machine learning and motion analysis. However, the cost of these powerful methodologies induces the search of cheaper techniques based on basic but proper notation methodology. The aim of this chapter is to provide an observational methodology for soccer match analysis. When paired with PageRank as the main indicator of performance, it allows for a deep analysis of the data and better decision-making and performance analysis in soccer. To show some insights about the proposed model, real data from past matches are presented and discussed. Results show graph visualization that sum up the whole match in terms of the flows of a network modelled with passes and recoveries from the players as weights of its edges. One implication of our research is to be a first approach in generalizing the PageRank algorithm to soccer team's management, which could be extrapolated to other disciplines. It also points to the feasibility of making a quantitative analysis for sport managers with a reasonable

J. Rojas-Mora (✉)
School of Informatics, Universidad Católica de Temuco,
Rudecindo Ortega n°, 2950 Temuco, Chile
e-mail: jrojas@inf.uct.cl

F. Chávez-Bustamante
Faculty of Business and Economics, Universidad del Desarrollo,
Ainavillo 456, Concepción, Chile
e-mail: fchavezb@udd.cl

J. del Río-Andrade • N. Medina-Valdebenito
Business Administration Institute, Universidad Austral de Chile,
Calle Viel s/n, Valdivia, Chile
e-mail: julio.delrio.andrade@gmail.com; nico.medinav@gmail.com

M. Peris-Ortiz et al. (eds.), *Sports Management as an Emerging Economic Activity*, DOI 10.1007/978-3-319-63907-9_16

cost-benefit ratio. This analysis opens the paths to further analysis that could include spatiotemporal variables.

Keywords PageRank • Graph theory • Social network • Observational methodology • Centrality

16.1 Introduction

Knowing how to collect, access, retrieve and integrate information is critical to effective performance analysis and decision-making processes (Vincent et al. 2009). The development and research of data in sports has taken different perspectives, like data mining (Ofoghi et al. 2013; Li 2014; Leung and Joseph 2014; Haghighat et al. 2013), information systems (Shao et al. 2014; Qi and Wang 2014; Xie and Cai 2014; Luo and Deng 2014), event detection in videos (Li and Sezan 2002; Taki et al. 1996; Tong et al. 2005; Taki and Hasegawa 2000), behavioural models (Menéndez et al. 2013; Cheng et al. 2002; Hernandez Mendo and Anguera 2002), social network analysis (Lusher et al. 2010; Vaz de Melo et al. 2012; Pardalos and Zamaraev 2014; Passos et al. 2011) and also outcome prediction by different statistical applications (Baker and Scarf 2006; Groll et al. 2015; Stekler et al. 2010; Leitner et al. 2010). A characteristic feature of all of these approaches is the amount of data obtained by any method chosen or perspective taken. As Wang and Wang (2015) stated, sports data feature strong timeliness, multiple types, many specifications, large quantities and very complex storage. Rapid development of information technology allows users access to all types of information, including high-quality footage of live sport events, and has led to an expansion of market size for sport, not seen before in the history of the industry (Westerbeek 2013). Given how big sport business has become, many efforts have been made to handle as much information as possible, and even more, these advances in information technologies have allowed researchers and managers to be able to advance towards sport-specific metrics (Gyarmati and Hefeeda 2016). With these mechanisms, managers from professional to college or amateur teams are now capable to develop deeper analysis of a given match in order to take some opportune corrective measures in their team performance.

In soccer, match analysis is a fully accepted detection vehicle for any serious-minded managerial and coaching staff (Carling et al. 2005). These authors define match analysis as the objective recording and examination of behavioural events occurring during competition. It could provide objective information about the underlying causes of a determined problem, e.g. a poor delivery in the penalty area or difficulties in taking the ball off the goalkeeper's zone. The final score, in contrast to detailed information as the unique performance indicator, is insufficient to properly asses a team. As a consequence, measurement tools play an important role in this particular field because human observation and memory are not reliable enough to provide accurate and objective information from athletes in high-performance competitions (Henriques Abreu et al. 2012). Furthermore, in most team sports, the observer is unable to assimilate the entire action taking place on the field, due to its attention to the game critical areas; hence, most of the peripheral play action gets usually lost (Hughes et al. 2001).

Given this context, the aim of this chapter is to provide an observational methodology for soccer match analysis, armed with just some basic information in order to develop a non-expensive tool based on network analysis. This is achieved by using the PageRank centrality measure created by Google (Brin and Page 1998) and some insights from graph theory and social networks. Our contribution has to do with a simple and economic application of the PageRank algorithm to the performance analysis of soccer players in understanding their performance, as centres in the flow of the ball, in any given match.

A key claim of our paper has to do with the fact that network analysis, when applied to soccer, allows for the representation of teamwork, which leads to a better understanding of the team as a whole, in contrast to the analysis of individuals and their personal contributions. This possibility of reinterpreting individual statistical data based on the comprehension of the group dynamic is an example that the conciliation of the common performance indexes with the novel approaches permits to cover every single level of analysis (Maya Jariego and Bohórquez 2013).

Some of the practical applications that our model features are as follows:

- Identification of the most relevant players
- Team dominance throughout the game
- Direction of the flows
- Relevance of the substitutes (compared to the time played)
- Ball recoveries or interceptions by player
- Identification of tactics (i.e. defensive or offensive game, long passes, area where the ball most circulated, etc.)
- Analysis of the rival team

The previous features are explained and discussed in this article on the following sections. After this introduction, the next section outlines the theoretical and conceptual framework that embodies our proposition with some important background to understand the core of our methodology. Section 16.4 describes the methodology required for the match analysis and the indicators used. Section 16.5 presents the results of different matches from the group phase of the past Copa America 2015 used as example. Finally, this chapter ends with a brief conclusions section, presenting the guidelines for future researches.

16.2 Conceptual and Theoretical Framework

16.2.1 Observational Methodology

Observational methods have been explained thoroughly by Bakeman and Gottman (1987). This type of methodology is characterized because the level of participation is “nonparticipative observation”, given that the observer does not interact with the observed players and the degree of perception is complete, direct observation (Lapresa et al. 2013).

Furthermore, in the field of sports, observation is important from both the procedural and substantive points of view. In terms of procedure, it is the only scientific approach that is capable of gathering data directly from participants (athletes, coaches, trainers, etc.) in both the training and competitive contexts without eliciting a response from them (Anguera and Hernandez-Mendo 2013). In this last article, a thorough revision of the existent literature about observational methodology is applied, containing examples from basketball, handball, soccer, judo and swimming, among other several disciplines from sports.

In our specific case, we observe the gameplays from video matches that are recorded from a specific match, but any video recorded would work, considering how easy is to replay a whole match in television and on the Internet. The benefits from using video replays (Carling et al. 2005) are that video:

- Provides a permanent record of performance which can be watched as many times as desired
- Provides valuable information that may have been missed or forgotten by coaches or players during the match
- Helps to concentrate in a specific aspect or a specific player's performance
- Allows for an action sequence to be repeated as often as necessary to ensure that players have absorbed and understood the required information
- Can be used in real time for immediate analysis and evaluation or post-match for a deeper insight
- Is a familiar mean of presenting and discussing performance

The video analysis process is described as follows:

1. Match is observed/recorded.
2. Analysis is made under digital video editing and/or data coding.
3. Four core elements are identified: player, action, time and position.
4. A database is generated containing two-dimensional match reconstruction, edited match video, tables, graphs and/or spatial data.

There are some important aspects that make observation a proper and objective tool to overcome some of the weaknesses of the common analysis, e.g. the coach gets a lot of information and is not capable to exploit that data, and the observer gets baffled by the number of actions taking place simultaneously or in rapid sequence which cannot be immediately processed given that his or her attention is directed to the most critical areas of the game. Emotional factors of the observer play also an important role (Frank and Miller 1991; Carling et al. 2005; Hughes et al. 2001).

Thus, a wide variety of research has been conducted using perspective of observational studies. These have contributed to facilitate the systematic observation of sports. Some of them can be found in the following articles: Castellano et al. (2008), Leitão and Campaniço (2009), Sarmento et al. (2009), Lapresa et al. (2016), Jonsson et al. (2006) and Santos et al. (2014). To sum up, working under an observational methodology ensures to be under the most suitable methodology used in sport studies when the objective is to analyse matches in their natural context and dynamics (Anguera and Hernandez-Mendo 2014).

16.2.2 Graph Theory

It is said that Euler in 1741 founded both topology and graph theory by solving the Konigsberg bridges problem. It consisted in visiting the four land masses of the entire city, starting and finishing in the same one, while completely crossing once over each of the seven bridges. Like this one, many situations can be described by means of a diagram consisting of a set of points connected with lines. This is the basic principle of graph theory; points are called "nodes" and the lines that connect them are named "edges". A graph G is an ordered triple $V(G)$, $E(G)$, φ_G consisting of a non-empty set $V(G)$ of vertices; a set $E(G)$, disjoint from $V(G)$, of edges; and an incidence function φ_G that associates with each edge of G an unordered pair of vertices of G. If e is an edge and u and v are vertices such that $\varphi_G(e) = uv$, then e is said to join u and v; the vertices u and v are called the ends of e (Bondy and Murty 1976).

16.2.3 Network and Centrality Measures

The scientific study of networks, including computer networks, social networks and biological networks, has received an enormous amount of interest in the last few years. Much of this interest can be attributed to the appeal of social network analysis on *relationships* among social entities and on the patterns and implications of these relationships. That is, relations defined by linkages among units are a fundamental component of network theories (Wasserman and Faust 1994). The structure of networks has been of interest of many branches of science: methods for analysing network data, including methods developed in physics, statistics and sociology; the fundamentals of graph theory, computer algorithms and spectral methods; mathematical models of networks, including random graph models and generative models; and theories of dynamical processes taking place on networks (Newman 2010).

Centrality has been widely studied in the context of social network analysis (Clemente et al. 2016; Lusher et al. 2010). Thus, several measures have been developed, like "betweenness" (Freeman 1979), "eigenvector centrality" (Bonacich 1972) and "closeness" (Freeman 1979), among others. Even though many measures and different approaches about the concept of centrality in a network exist, Freeman (1979) offers three intuitive conceptions:

(a) The most intuitive conception is that point centrality is some function of the degree of a point. The degree of a point p_i is the count of number of other points (nodes) $p_j (j \neq i)$ that are adjacent [1] to it.

[1] Two incident vertices with a common edge are *adjacent*.

(b) The second view is based upon the frequency with which a point falls between pairs of other points on the geodesic paths connecting them.
(c) The third conception is based upon the degree to which a point is close to all other points in the graph.

The main idea of these different approaches of centrality is to define a measure that determines the relative importance of a node within a graph. The discussion of these different researches focuses on what the most appropriate measurement should be. A complete summary and revision of the concept of centrality in networks and the different existing measures and interpretations of the concepts can be found in Borgatti (2005). The key claim of that paper is that centrality measures can be regarded as generating expected values for certain kinds of node outcomes (such as speed and frequency of reception) given implicit models of how traffic flows.

Regarding network analysis and sports, intra-group relationships are important for sport teams and include aspects such as cohesiveness and hierarchies among players (Lusher et al. 2010). Social network analysis (SNA) methods allow for the exploration of "social[2]" relations between team members and their individual-level qualities simultaneously. Its usefulness has to do with addressing the issue of interdependencies in the data inherent in team structures. The most basic concept of relationship in network analysis is defined by the existence of a link between two players (e.g. i and j). It is binarily defined, where $e_{i,j} = 1$ models the existence of a relation, and $e_{i,j} = 0$ represents its absence. More complex networks might consider valued edges, depending on the importance or strength of the bond.

16.2.3.1 PageRank Centrality

PageRank, a registered trademark of Google, is an algorithm introduced by Brin and Page (1998) and used to determine the relative importance of a node, i.e. its centrality in the graph. Its most intuitive definition is that a node has high rank if the sum of the ranks of its backlinks[3] is high.

The definition of a simple ranking is as follows: Let u be a node, F_u be the set of nodes u points to, B_u be the set of nodes that point to u, $N_u = |F_u|$ be the number of edges to/from u and c be the factor of normalization; in order to keep constant the total rank of all the nodes, then the definition of R, a simplified version of PageRank, would be:

$$R(u) = c \sum_{v \in B_u} \frac{R(v)}{N(v)} \tag{16.1}$$

Even though the equation is recursive, it may be computed by starting with any set of ranks and iterating the computation until it converges. To overcome the problem

[2] "Social", in this case, refers to how frequently a player passes the ball to another.

[3] For a given node in a graph, its "backlinks" are the nodes linking to it.

that, during the iteration, the loop will not distribute rank because two nodes are linked only between each other and, therefore, there are no out edges, Page et al. (1999) introduced the following rank source.

Let $E(u)$ be some vector over the nodes that corresponds to a source of rank, then the PageRank of a set of nodes is an assignment R', to the nodes which satisfies:

$$R'(u) = c\sum_{v \in B_u} \frac{R'(v)}{N_v} + cE(u) \tag{16.2}$$

Such that cc is maximized and $\|R\|_1 = 1$, where $\|R\|_1$ denotes the L_1 norm of R'. As presented from the creators of the method, the PageRank algorithm may be computed as:

$$\begin{aligned}
&R_0 \leftarrow S \\
&\text{loop:} \\
&R_{i+1} \leftarrow AR_i \\
&d \leftarrow \|R_i\|_1 - \|R_{i+1}\|_1 \\
&R_{i+1} \leftarrow R_{i+1} + dE \\
&\delta \leftarrow \|R_{i+1} - R_i\|_1 \\
&\text{while } \delta > \varepsilon
\end{aligned} \tag{16.3}$$

The complete axiomatization of this page ranking algorithm can be found in Altman and Tennenholtz (2005). Some known applications of the algorithm presented are citation networks (Ding et al. 2009; Ma et al. 2008), and like ours, many other applications have been made in chemistry, biology, bioinformatics, neuroscience and engineered systems, among others (Gleich 2014). Several approximations of the PageRank algorithms have been developed, and quantitative analyses have been provided to illustrate the effectiveness of the PageRank computation (Chung 2014).

In simple words, our approach generalizes the PageRank algorithm, commonly used to rank the importance of websites, to value each node given the frequency of the flows in the network, which in this case refers to the passing of the ball. This, in order, allows for a complete visualization of the game that shows important information. This representation of the match in one map of relationships between the players as a team allows for the identification of game patterns (Lago and Anguera 2003). Given that goals are infrequent events, it is necessary to have a different independant variable, i.e. passes, in order to better understand the performance of a player in a match. Supporting this, Lago and Martín (2007) found that the variance of the goals scored in soccer matches is not large enough to identify statistically significant determinants.

16.3 Methodology

16.3.1 Data Collection

To provide a structure for the application of the methodology, we propose a simple graph where players are nodes and the weights of the edges represent the passes between any two players. In order to do this, there are some considerations in the notation to bear in mind.

Notational analysis is defined as a means of recording events so that there is an accurate and objective record of what actually took place. It provides a factual record that does not lie (Carling et al. 2005). The first part of our proposal is pretty straightforward. It begins with an almost standard notational procedure, like those presented in Sarmento et al. (2009), Anguera and Hernandez-Mendo (2013) and Lapresa et al. (2013). It consists on dividing the area of the field into 12 same-sized zones: three rows (left, centre and right) and four columns (e.g. I, II, III and IV).

With this starting point, the aim is to collect information of each play of the game considering the following information:

- Time of the play
- Sender of the ball
- Zone of the field that the ball is sent from
- Defending player (if any)
- Receiver of the ball
- Zone of the field where the ball is being received
- Defending player of the receiver player (if any)

It is important to bear in mind that the analysis is going to be made under the basic assumption that the relative importance of a player is given by the importance of the flows received, which we model as passes. Therefore, the whole match is going to be understood as a network where flows are given by the ball passing through the players. The collection of all of these data can be made with a simple spreadsheet. With the information of each play obtained in this simple way, it is possible to model an adjacency matrix of 28 columns and 28 rows (11 starting players plus the 3 substitute players for each team).

In the second part, the collected data is processed through an R script (a GNU software[4]) using the "*igraph*" package (Csardi and Nepusz 2006), which contributes with routines for graph and network visualization and analysis. This package includes an implementation of the PageRank algorithm that we used for our analysis.

[4] For further inquiries about GNU R (R Core Team 2016), check the website of the R Foundation for Statistical Computing, URL https://www.R-project.org/

16.4 Results

In order to visualize the results of our work, the methodology was applied to three soccer matches from the group phase of the Copa America 2015. In specific, those matches analysed were Chile-Ecuador, Chile-Mexico and Chile-Bolivia. They were broadcasted by a public television channel, so the recording was made with digital video tools.[5]

The nodes represent the players, the notation "C" is for the Chilean players, while "E" is used for the Ecuadorian players, and the number next to each letter corresponds to the shirt number worn by any given player in that match. The size of the node is proportional to its PageRank centrality measure, and the boldness of the edges is proportional to the number of passes from one player to another. Nodes that seem to be outliers on the bottom of the figures are the substitute players. The same goes for the next figures, where "M" denotes Mexican players and "B" denotes Bolivian players. We present a brief discussion of the graphs showing some basic examples of the types of analysis that can be made with this tool.

Figure 16.1 shows the Chilean dominance on the match against the Ecuadorians with a higher volume of passes and variety of options to deliver; it meant a higher possession of the ball through the entire match. Boldness of the Chilean arrows indicates an offensive tendency towards the middle field to get to the centre forward C7, Alexis Sánchez. Another conclusion has to do with the principal nodes of each team. For the Chilean case, the bigger node corresponds to C8, Arturo Vidal, the most relevant player given the flows that passed through him and the importance of the players who passed the ball to him, i.e. his PageRank value. Therefore, Vidal became the most important player for the Chilean team which won for 2 goals against 0, with the first one scored by Vidal.

For sport managers, it is useful to analyse other aspects too, like the flows between two nodes of different teams. For the case of E13, Enner Valencia, Fig. 16.1 shows that from the several flows that are directed to him, many correspond to Chilean nodes. This means that he interfered with many passes intended for Chilean players and, accordingly, recovered the ball many times, playing a defensive role even though his position was that of a winger.

In the Chile-Mexico match, Fig. 16.2 shows that the ball circulated significantly through every player, with a higher density on the Chilean side of the field. In time, Chilean nodes are bigger, have a higher centrality or are more important, with the most relevant players being C8, Arturo Vidal, and C10, Jorge Valdivia, with Vidal scoring 2 goals throughout the match.

For the Mexican team, the most relevant players are their three forwarders, showing a considerable amount of passes received from the central and lateral defenders, which means that the Mexican game was characterized by long passes (e.g. M3 to M7 edge). The most relevant players for Mexico were M9, Raul Jimenez, and M19, Matías Vuoso, both managed to score goals for their team.

[5] The dataset can be requested to the corresponding author.

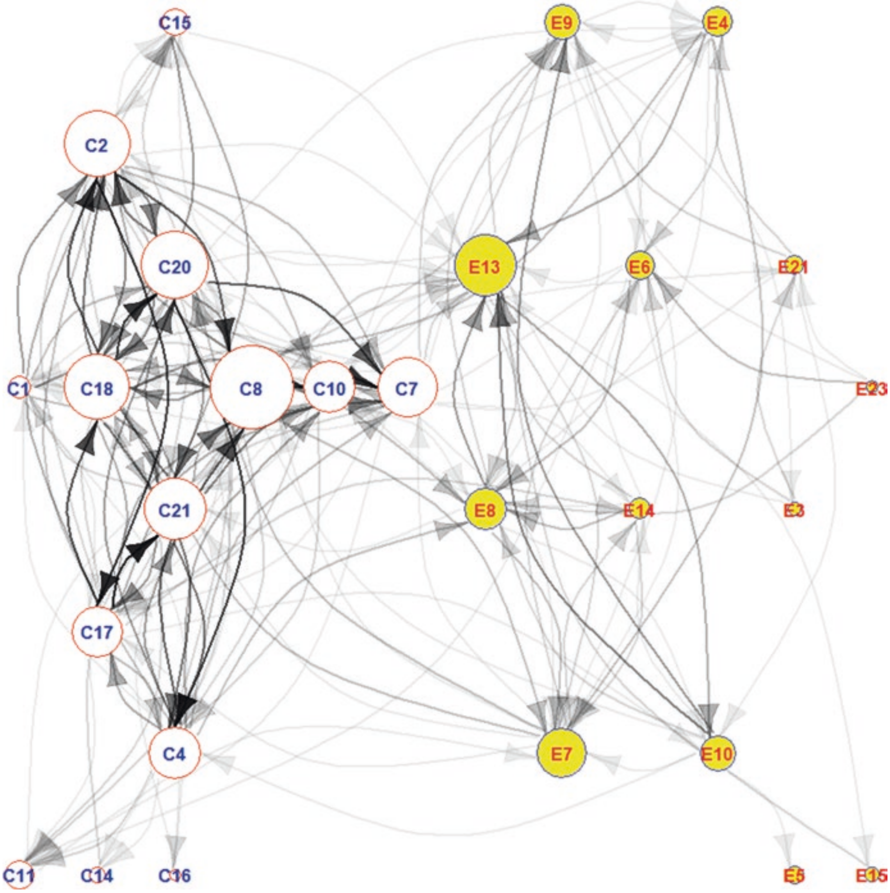

Fig. 16.1 Network of the Chile-Ecuador match in the group phase of the Copa America 2015

Finally, the Chile-Bolivia game ended with a lopsided score of 5-0. In this match, Chile played with two more "open" forwarders than the previous games, which allowed them to dominate more positions of the field with a considerable flow of the ball going to C11, Eduardo Vargas. There is clear relation between the volume of the flows for the Chilean team against the Bolivian team and the final score of the match, i.e. Chile owned the ball and, thus, the opportunities to score. In this aspect it is important to consider: in the discipline, one of the most important findings is the correlation between the ability to retain possession of the ball for prolonged periods of time and success (Bate 1988; Gómez and Álvaro 2002; James et al. 2004; In: Lago and Martín 2007).

Another analysis to be made is that in this match, the substitute players had a bigger importance in the development of the plays for both of the teams, which could not be seen, e.g. in the Ecuadorian team on Fig. 16.1. This adds another type of analysis

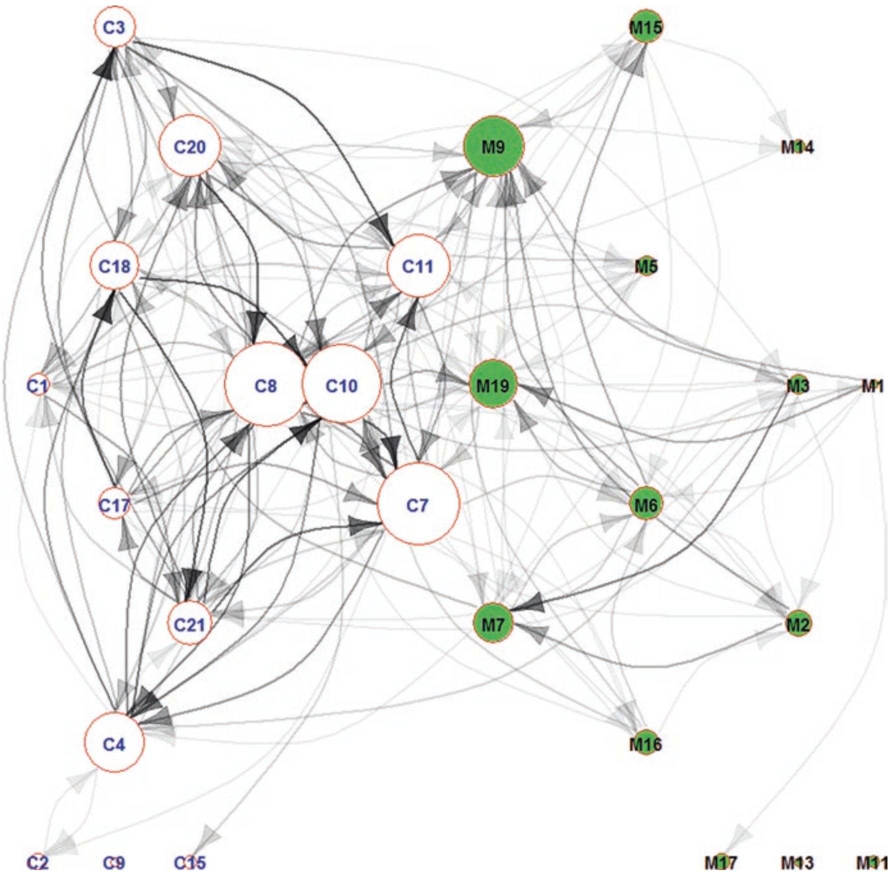

Fig. 16.2 Network of the Chile-Mexico match in the group phase of the Copa America 2015

to our approach: relevance of the substitute players during the time they played. In Fig. 16.3, the difference of the relevance of the substitute nodes for both teams is remarkable. Some are even similar to the nodes of players that started the game from the minute one. Hence, sport managers can take decisions and evaluate performance, given the relevance that substitutes manage to acquire during a shorter period than the regular first team players, and maybe evaluate if they are making the most of their time on field. For example, a coach might decide to turn a substitute into a regular starter because of the relevance of his node and flows directed from and to him.

A last thing to keep in mind is that the analysis can also have another focus. Sport managers can use this methodology not only to analyse their team's weaknesses and strengths but also to analyse the adversaries through the data collection of other previous matches in order to set proper tactics and, in this way, develop that competitive advantage that is so much important in such a competitive environment (i.e. see which are the opponents that do not have so much relevance and then focus the direction of the ball through the less explored areas of those nodes).

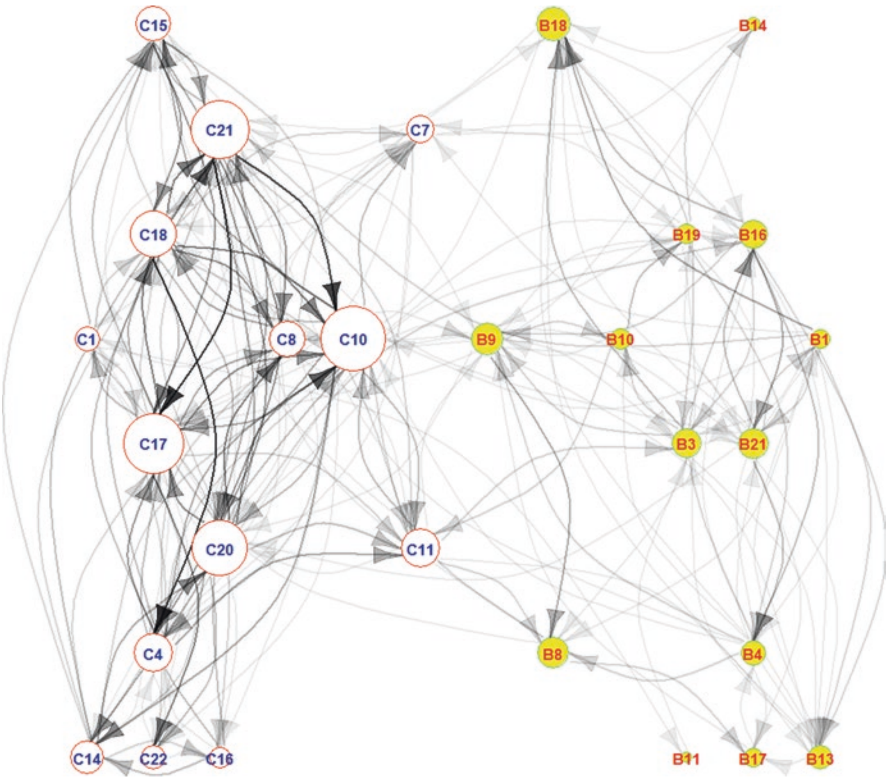

Fig. 16.3 Network of the Chile-Bolivia match in the group phase of the Copa America 2015

16.5 Conclusion

Graph-based algorithms have been proved to be relevant to a wide variety of applications. Even though, there is no such thing as a "perfect" algorithm to rank in sports, there is strong evidence to believe that the Google PageRank algorithm provides reliable insights (London et al. 2014; Govan et al. 2008). In our case, it aids to provide answers to the "*who, when and where*" of the plays of the match, gathering the data from point to point and transforming it into a single graph that becomes a powerful visualization, in order to aid coaches or managers to make deeper analysis. This information gets often lost at the time being played, given the huge amount of interactions that occur between the players, which disables the observers to take in-game decisions based on this data.

Our approach is within context of social networks and graph theory, by using the PageRank centrality measure as a key element to rank the nodes. That algorithm has demonstrated a wide variety of applications and a resourceful way to measure the relevance of the vertices of a given graph in many different situations. It is also embodied within the observational methodologies and notational analysis, which ensures to be under the most suitable design methodology for sport analysis as stated before, given the importance of studying sports with its natural dynamics and that contributes to keep an objective record. Other benefits of this type of methodologies are related to overcoming the lack of attention that is paid during a match due to the amount of information for the observers and the emotional or personal factors of the observer that could potentially affect the analysis carried out.

We applied the methodology proposed to three real matches for the Copa America 2015 (Chile-Mexico, Chile-Bolivia and Chile-Ecuador), which illustrated some of the ideas that can be concluded with this observational methodology. Some of the insights that this method could provide for a match analysis are most relevant players, direction of the ball through the entire game, ball recoveries, effective passes, zone of the field that was less covered by seeing the participation of the node in its assigned section of the field, relevance of the substitutes and analysis of the rival team, among others.

The application of social network analysis (SNA) methodologies to sports has been of wide interest (Lusher et al. 2010). Its application to soccer allows modelling a team as a micro-system, whose components are linked by stable and ordered interactions that represent the collective work (Lago and Anguera 2003). At the same time, it allows to analyse the role that different nodes in the graph play, having a wider perspective on how this so-called system works (Maya Jariego and Bohórquez 2013).

Implications for sport managers are significant. Considering the low cost of this methodology, it could generate many benefits for the right performance analysis and decision-making process through a simple but powerful visualization. Although many big teams are using different methods specially developed for sport analysis (e.g. video detection analysis, specialized software or appropriate databases for large amounts of data storage) which have a considerable cost, it is important to bear in mind the budget that smaller teams might have. Our proposal contributes by adding a simple, but scientific, methodology to analyse the data, with a widely used algorithm as PageRank, which will contribute with more angles than just the final score as a performance index. The only costly side of our proposal has to do with the time needed to write down all the plays of the match. However, with just a basic spreadsheet and some simple automation, this burden can be somewhat eased.

For future work, we intend to study the indirect flow of passes given our initial adjacency matrix, to see which player of a team can indirectly allow for the flow of the ball between two almost unlinked players. We would also like to analyse the network flows from a spatiotemporal perspective, to obtain information like the position of the field which is more exploited, or the dynamics in the flow of passes of a given game. This would lead to the need of adding more detailed information at the data collection stage, and therefore, a larger database should be required. However, it would not change the core of the proposed methodology. The context for this future research has to do with the main idea of being able to find that “little bit extra” which could potentially make the difference between success and failure in sport team management (Carling et al. 2005).

References

Altman, A., & Tennenholtz, M. (2005). Ranking systems: The PageRank Axioms. In: *EC '05 Proceedings of the 6th ACM conference on electronic commerce, ACM, 1–8*.

Anguera, M. T., & Hernandez-Mendo, A. (2013). Observational methodology in sport sciences. *Revista de Ciencias del Deporte, 9*(3), 135–160.

Anguera, M. T., & Hernandez-Mendo, A. (2014). Técnicas de análisis en estudios observacionales en ciencias del deporte. *Cuadernos de Psicología del Deporte, 15*(1), 13–30.

Bakeman, R., & Gottman, J. M. (1987). Applying observational methods: A systematic view. In J. D. Osofsky (Ed.), *Handbook of infant development* (pp. 818–853). New York: Wiley.

Baker, R., & Scarf, P. (2006). Predicting the outcomes of annual sporting contests. *Journal of the Royal Statistical Society Series C (Applied Statistics), 55*(2), 225–239.

Bate, R. (1988). Football chance: Tactics and strategy. In T. Reilly, A. Lees, K. Davids, & W. Murphy (Eds.), *Science and football* (pp. 293–301). London: E & FN Spon.

Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. *The Journal of Mathematical Sociology, 2*(1), 113–120.

Bondy, J. A., & Murty, U. S. R. (1976). *Graph theory with applications*. Great Britain: The Macrmillan Press Ltd.

Borgatti, S. P. (2005). Centrality and network flow. *Social Networks, 27*, 55–71.

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems, 30*(1), 107–117.

Carling, C., Williams, A. M., & Reilly, T. (2005). *Handbook of soccer match analysis*. New York: Taylor & Francis Group.

Castellano, J., Perea, A., Alday, L., & Hernandez-Mendo, A. (2008). The measuring and observation tool in sports. *Behavior Research Methods, 40*(3), 898–905.

Cheng, F., Christmas, W. J., & Kittler, J. (2002). Recognising human running behaviour in sports video sequences. *Proceedings - International Conference on Pattern Recognition, 16*(2), 1017–1020.

Chung, F. (2014). A brief survey of PageRank algorithms. *IEEE Transactions on Network Science and Engineering, 1*, 38–42.

Clemente, F. M., Lourenço, F. M., & Sousa, R. (2016). *Social network analysis applied to team sports analysis*. Cham: Springer International Publishing.

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *Interjournal, Complex Systems, 1695*(5), 1–9.

Ding, Y., Yan, E., Frazho, A., & Caverlee, J. (2009). PageRank for ranking authors in co-citation networks. *Journal of the American Society for Information Science and Technology, 60*(11), 2229–2243.

Frank, I. M., & Miller, G. (1991). Training coaches to observe and remember. *Journal of Sports Sciences, 9*, 285–297.

Freeman, L. C. (1979). Centrality in social networks conceptual clarification. *Social Networks, 1*, 215–239.

Gleich, D. F. (2014). PageRank beyond the web. *SIAM Review, 57*(3), 321–363.

Gómez, M., & Álvaro, J. (2002). El tiempo de posesión como variable no determinante del resultado en los partidos de fútbol. *El Entrenador Español, 97*, 39–47.

Govan, A. Y., Meyer, C. D., & Albright, R. (2008). Generalizing Google's PageRank to National Football League Teams. In: *Proceedings of the SAS Global Forum*, San Antonio, paper 151.

Groll, A., Schauberger, G., & Tutz, G. (2015). Prediction of major international soccer tournaments based on team-specific regularized Poisson regression: An application to the FIFA world cup 2014. *Journal of Quantitative Analysis in Sports, 11*(2), 97–115.

Gyarmati, L., & Hefeeda, M. (2016). Analyzing in-game movements of soccer players at scale. MIT Sloan Sports Analytics Conference 2016.

Haghighat, M., Rastegari, H., & Nourafza, N. (2013). A review of data mining techniques for result prediction in sports. *Advances in Computer Science: an International Journal, 2*(5), 7–12.

Henriques Abreu, P., Moura, J., Castro-Silva, D., Reis, L. P., & Garganta, J. (2012). Performance analysis in soccer: A Cartesian coordinates based approach using RoboCup data. *Soft Computing, 16*(1), 47–61.

Hernandez Mendo, A., & Anguera, M. T. (2002). Behavioral structure in sociomotor sports: Roller-hockey. *Quality and Quantity, 36*(4), 347–378.

Hughes, M., Evans, S., & Wells, J. (2001). Establishing normative profiles in performance analysis. *Journal of Performance Analysis in Sports, 1*, 4–27.

James, N., Jones, P. D., & Mellalieu, S. D. (2004). Possession as a performance indicator in soccer. *International Journal of Performance Analysis in Sport, 4*, 98–102.

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

Lago, C., & Anguera, M. T. (2003). Utilización del análisis secuencial en el estudio de las interacciones entre jugadores en el fútbol de rendimiento. *Revista de Psicología del Deporte, 12*(1), 27–37.

Lago, C., & Martín, R. (2007). Determinants of possession of the ball in soccer. *Journal of Sport Sciences, 25*(9), 969–974.

Lapresa, D., Álvarez, L., Arana, J., Garzón, B., & Caballero, V. (2013). Observational analysis of the offensive sequences that ended in a shot by the winning team of the 2010 UEFA futsal championship. *Journal of Sports Sciences, 31*(15), 1731–1739.

Lapresa, D., Arana, J., Anguera, M. T., Pérez-Castellano, J. I., & Amatria, M. (2016). Application of logistic regression models in observational methodology: Game formats in grassroots football in initiation into football. *Anales de Psicología, 32*(1), 288–294.

Leitão, J. C., & Campaniço, J. (2009). Research methods support in observation sports laboratory. *Motricidade, 5*(3), 84.

Leitner, C., Zeileis, A., & Hornik, K. (2010). Forecasting sports tournaments by ratings of (prob) abilities: A comparison for the EURO 2008. *International Journal of Forecasting, 26*(3), 471–481.

Leung, C., & Joseph, K. (2014). Sports data mining: Predicting results for the college football games. *Procedia Computer Science, 35*, 710–719.

Li, B., & Sezan, M. (2002). *Event detection and summarization in American football broadcast video. Proceedings of SPIE*, 4676.

Li, H. (2014). Strategy and analysis of sport events based on data mining technology. *Applied Mechanics and Materials, 687-691*, 1137–1140.

London, A., Németh, J., & Németh, T. (2014). Time-dependent network algorithm for ranking in sports. *Acta Cybernet., 21*(3), 495–506.

Luo, S., & Deng, G. (2014). Research on the development of sports information system based on the sports information management theory. *Advanced Materials Research, 998-999*, 1327–1330.

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.

Ma, N., Guan, J., & Zhao, Y. (2008). Bringing PageRank to the citation analysis. *Information Processing & Management, 44*(2), 800–810.

Maya Jariego, I., & Bohórquez, M. R. (2013). Análisis de las redes de distribución de balón en fútbol: pases de juego y pases de adaptación. *Revista Hispana para el Análisis de Redes Sociales, 24*(2), 135–155.

Menéndez, H., Bello-Orgaz, G., & Camacho, D. (2013). Extracting behavioral models from 2010 FIFA world cup. *Journal of Systems Science and Complexity, 26*(1), 43–61.

Newman, M. (2010). *Networks*. Oxford: Oxford University Press.

Ofoghi, B., Zeleznikow, J., MacMahon, C., & Raab, M. (2013). Data Mining in Elite Sports: A review and a framework. *Measurement in Physical Education and Exercise Science, 17*(3), 171–186.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). The pagerank citation ranking: Bringing order to the web. Technical Report, Stanford University, Stanford.

Pardalos, P., & Zamaraev, V. (2014). The impact of social networks on sports. In P. Pardalos & V. Zamaraev (Eds.), *Social networks and the economics of sports* (pp. 1–8). Cham: Springer.

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.

Qi, N., & Wang, L. (2014). Constructing of sports information management theories and developing research of application system. In: *Proceedings of the IEEE Workshop on Advanced Research and Technology in Industry Applications (WARTIA)*, 2014, Ottawa.

R Core Team. (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org

Santos, S., Sarmento, H., Alves, J., & Campaniço, J. (2014). Construcción de un instrumento para la observación y el análisis de las interacciones en el waterpolo. *Revista de Psicología del Deporte, 23*(1), 191–200.

Sarmento, H., Leitao, J., Anguera, T., & Campaniço, J. (2009). Observational methodology in football: Development of an instrument to study the offensive game in football. *Motricidade, 5*(3), 19–24.

Shao, L., Luo, L., & Qi, H. (2014). Theoretical construction of sports information management and research on application system development. *Advanced Materials Research, 926-930*, 2759–2762.

Stekler, H., Sendor, D., & Verlander, R. (2010). Issues in sports forecasting. *International Journal of Forecasting, 26*(3), 606–621.

Taki, T., Hasegawa, J., & Fukumura, T. (1996). Development of motion analysis system for quantitative evaluation of teamwork in soccer games. In: *Proceedings of 3rd IEEE International Conference on Image Processing*, 1996, Laussane.

Taki, T., & Hasegawa, J. (2000). Quantitative measurement of teamwork in ball games using dominant region. *International Archives of Photogrammetry and Remote Sensing, XXXIII*(Suplement B5), 125–131.

Tong, X., Duan, L., Lu, H., Xu, C., Tian, Q., & Jin, J. (2005). A mid-level visual concept generation framework for sports analysis. In: *Proceedings of the IEEE International Conference on Multimedia and Expo*, 2005, Amsterdam.

Vaz de Melo, P., Almeida, V., Loureiro, A., & Faloutsos, C. (2012). Forecasting in the NBA and other team sports. *ACM Transactions on Knowledge Discovery from Data, 6*(3), 1–27.

Vincent, J., Stergiou, P., & Katz, L. (2009). The role of databases in sport science: Current practice and future potential. *International Journal of Computer Science in Sport, 8*(2), 50–66.

Wang, Q., & Wang, Y. (2015). Standarized storage of sports data based on XML. *The Open Cybernetics & Systemics Journal, 9*, 2312–2316.

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

Westerbeek, H. (2013). *Global sport business: Community impacts of commercial sport*. London: Routledge.

Xie, P., & Cai, X. (2014). Research on sports information integrated management application development. *Advanced Materials Research, 926-930*, 4182–4185.