Analysis of Football Game Performance Based on Social Network

Yaoyao Qian and Xianming Wang(✉)

School of Mathematics and Information Engineering, Wenzhou Institute of Technology, Wenzhou, China
xmwung@ustc.edu

Abstract. With the advent of big data and network era, the interpersonal relationship is getting closer and closer, and the advantage of teamwork is becoming more and more prominent. By analyzing the team competition rules, combining the team members' abilities, characteristics, and interactions between team members in previous competitions, team cooperation, and coordination strategy can be further optimized effectively. Based on the data samples recorded in 38 football games, this paper takes the 14th huskies vs. O14 game with a 4-0 result as an example. After data preprocessing, DBSCAN, a clustering algorithm, is introduced to simplify the data set by presenting a scatter graph. Each player's resident coordinates in each game are obtained, which provides data for social network analysis. We used a Python program to get the overall network analysis. The D6, M1, and F2 players on the Huskies form a triad configuration, the players D3 and D4, the players D1 and M2 in the O14 team form multiple dyadic configurations. To better understand the characteristics of the passing network, the individual network analysis is carried out. Through the study of particular net research degree center degree, it can be found that the Huskies coach's tactical arrangement was 4-4-2 formation, which was changed into 4-3-3 attacking formation in an actual combat operation. The tactical arrangement of the O14 coach was the same as the essential combat operation, which was 2-3-5 formation and more defensive. The in-depth analysis found that the Huskies' core was a midfielder, while the O14's core was not a midfielder. The tempo of the game and the fluidity of the passing network of the O14 team were significantly reduced, which was also the main factor leading to the failure.

Keywords: Football Team Performance Evaluation · Social Network Analysis · Betweenness · Clustering Algorithm

1 Introduction

With the rapid development of science and technology, economy, and education, teamwork has gradually become the chief means of solving complex problems. By collecting interdisciplinary scholars with different professional knowledge and different perspectives, appropriate cooperation strategies are selected to achieve a team cooperation model. With the help of the optimal approach to build a team, it can better reflect the

N. Radojević et al. (Eds.): ICAID 2022, AHIS 7, pp. 811–819, 2023.
https://doi.org/10.2991/978-94-6463-010-7_81

team's advantages. The success of a team is not only limited to individual ability but also influenced by many factors, such as team performance, mutual understanding, the team's balance between individual and collective performance, and the team's overall ability, which can often be effectively studied to optimize cooperative coordination strategies further.

In this work, we established a mathematical model to understand the dynamics of the team further. In particular, the complicated interactions between players on the field impacted their ability to play well and win. The purpose of this study is not only to explore interactions that directly lead to scoring but also to explore team dynamics throughout a game or season to improve and optimize team cooperation strategies. The structure and dynamic characteristics of a team's success (or failure) also need to be quantified and normalized. The construction and analysis of the model in this paper are based on the Huskies' detailed competition data in a particular season. This paper solves the following problems according to the data obtained.

We built a transmission network system where each player in the network is a node and transmission between players is a link. We analytically build a model by exploring the connections between local and global players in the network at multiple scales and other information from the soccer game to determine which strategies work for the Huskies. We offered the team directions for improvement based on the network model. In the end, we provided recommendations for effective team design based on the model analysis experience.

2 A Soccer Performance Analysis Model Based on Social Network Analysis

2.1 Background

As one of the most basic modern football techniques, passing is an integral part of football matches' techniques and tactics. The planning of the passing approach between the players of a football team can affect the match's result and the level of performance.

Most studies on passing techniques are based on individual players and ignore the connections and interactions between individuals. With the advent of big data, the data sets available for analysis have become more complex and complete. Game results are recorded in increasing detail, the connections between players are becoming more robust, and the metrics are becoming more diverse. Traditional research methods clearly cannot meet the needs of current match performance research. Therefore, based on the passing networks' perspective, the proposed model utilizes social analysis to analyze the payoff characteristics in soccer matches, switching the focus of investigation from individual players to the interactions between them to explore and identify network patterns in soccer matches.

2.2 Related Elaboration of Social Network

A social network is a form of social organization based on "networks" rather than "groups". It is also an analytical perspective of Western sociology that emerged in the

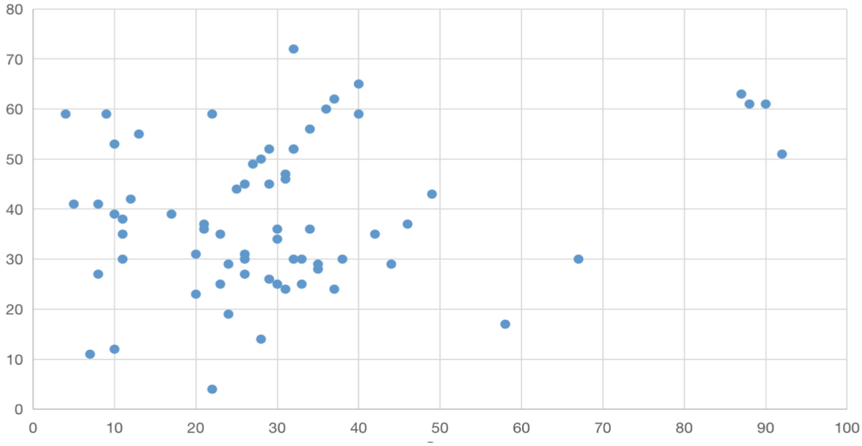

Fig. 1. Huskies_D2 scatter plot of player's initial position coordinates

1960s [1, 2]. The way people interact with each other in the social environment is observed using social networks, reflecting a stable social structure based on relationships. A quantitative analysis of this structure gives rise to social network analysis. Social network analysis builds networks by measuring or investigating the characteristics of a social system's parts and the connections between the elements and analyzing the patterns and factors of relationships. However, social network analysis is not limited to the study of social systems. It can also describe and evaluate the relationships between actors or the information and resources circulating in those relationships.

In soccer, players establish a predominantly passing relationship with each other by passing the ball. The more passes a player makes, the closer the relationship becomes. In a network composed of players and passing associations, each player represents a node in the network, and each pass forms a connection between players. This network has the same structure as the social network established by the basic social system. Thus, this model adopts the social network analysis method to quantitatively study the passing behavior in soccer games.

2.3 Data Preprocessing for Constructing Social Network

2.3.1 Construction of Scatter Plot

We performed a scatter plot based on the detailed data of players who touched the ball at each moment in the dataset to obtain each player's main formations in a given game. By analyzing the data, we found that MatchID of 14 had the most significant score difference, which was used as a model analysis case. The scatter plot drawn based on the initial position coordinates of Huskies_D2 players is given in Fig. 1.

As shown in Fig. 1, there are scattered positions. If the average coordinates are used as the players' center coordinates, a significant error will be generated. According to the situation shown in the above figure, it is more appropriate to use the processing method of clustering.

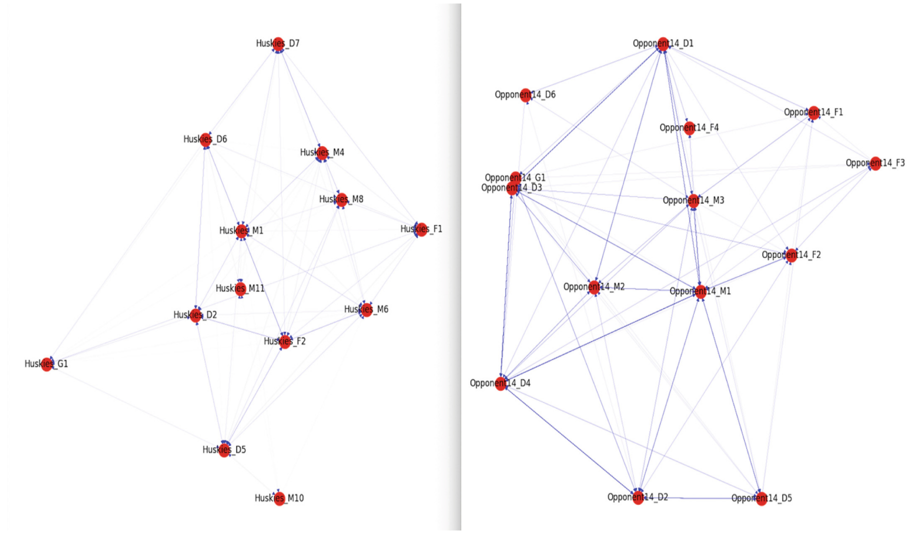

Fig. 2. Social network diagram of the Huskies and O14 team in game 14

2.3.2 The DBSCAN (Density-Based Spatial Clustering of Application with Noise)

According to related information, the DBSCAN algorithm [3, 4], a density-based spatial data clustering method, uses the R*-tree method to retrieve the neighboring points. It is very fast because the final result can be obtained by searching all data only once. Simultaneously, it can handle the clustering of arbitrary shapes and remove noise according to the threshold MinPts, which is a better algorithm for spatial data clustering. Therefore, the DBSCAN algorithm is used to calculate the coordinates of the players.

First, we perform a preprocessing operation to group the initial data based on MatchID and OriginPlayerID, and code in python to get the positions of each player's appearance in each match. When compiling the DBSCAN algorithm program, the field's total length and width are considered, and the Eps-neighborhood point is set to 25. The data after each grouping is passed into the prepared DBSCAN program to get the main appearance point of a player in the game. The field data of MatchID 1 is selected as an example to calculate the primary coordinates of Huskies_D1 player and compare them with the positions in Fig. 1.

2.4 Results and Analysis

2.4.1 Global Network Analysis

2.4.1.1 Pass Network Analysis

Based on the data obtained from the above procedure, the Huskies and the O14 team in the 14th game are taken as an example, and the main coordinates are the node coordinate positions, and the pass numbers are the connection numbers, and the images are drawn by substituting them into the social network model. The details are shown in Fig. 2.

Table 1. Index of Passing Track Network of the Huskies and O14 team

The Huskies			The O14		
Total number of passes	The number of ball holding	Goals	Total number of passes	The number of ball holding	goals
342	813	4	347	831	0

By combining the passing relationships between Huskies and O14 players in Game 14 and the information of players' positions shown in Fig. 2, we found that Huskies' thick lines were primarily concentrated in the midfield and front fields and that O14's were primarily concentrated in the midfield and backfield. Besides, it can be seen from the passing network diagram that Huskies has more vertical thick lines, while O14 has more horizontal thick lines. It indicates that although O14 had more passes, most of the passes were in the midfield and backfield, while Huskies had more lines in the midfield and front field, and that Huskies had a much greater tendency to pass forward than O14. It can be seen that the Huskies' D6, M1, and F2 players had relatively thicker connections with each other, indicating that the three teammates were closely connected, forming a three-legged configuration. The direct connections between D3 and D4, D1 and M2, D2 and D5, and M3 and M1 of the O14 team are relatively coarse, indicating that the teammates in the middle and back of the O14 team are closely related to each other, forming a dyadic configuration. The passing network diagram allows us to visualize the passing relationships and passing trends of the whole team and get a general idea of which nodes and connections within the group are more critical. However, a quantitative analysis of the passing network is necessary to understand better characteristics of the passing network.

2.4.1.2 Analysis of Network Index

By calculating the network's density and the ratio of the actual number of connections to the maximum possible number of links in the informal network, we can measure the relationships between the network members. Based on the social network model, network indicators are calculated as shown in Table 1.

All nodes are shown to be reachable in Table 1 and Fig. 2, which means that all players have touched the ball. Most of the nodes have direct contact with other nodes, which means that most of the players have direct contact with the pass. It can be seen that the overall network characteristics of the two teams are similar. In addition, as seen in Table 1, the O14 team passed the ball slightly more than the Huskies in all metrics, and the O14 team had more possession than the Huskies, but the game resulted in a 0:4 loss for the O14 team. This shows that it is not enough to analyze the passing network only at the overall network level. A deeper analysis of both teams' passing networks from the perspective of individual networks is needed to make a reasonable explanation of the match result.

Table 2. The Huskies and the O14 team degree centrality

The Huskies				The O14			
Player	The node degree	In-degree	Out-degree	Player	The node degree	In-degree	Out-degree
H_F2	8.67	4.50	4.17	O14_D2	5.08	**2.23**	**2.85**
H_D6	5.25	3.00	2.25	O14_G1	1.92	**0.69**	**1.23**
H_F1	4.00	**1.00**	**3.00**	O14_F1	1.38	**0.54**	**0.85**
H_M6	5.00	**2.33**	**2.67**	O14_M3	4.15	2.38	1.77
H_D7	4.92	2.50	2.42	O14_M2	3.54	1.85	1.69
H_D5	5.08	**2.50**	**2.58**	O14_M1	9.46	4.92	4.54
H_D2	5.42	3.17	2.25	O14_D4	8.00	4.15	3.85
H_G1	2.33	**1.08**	**1.25**	O14_D3	6.15	3.38	2.77
H_M8	3.83	**1.75**	**2.08**	O14_D1	5.92	**2.69**	**3.23**
H_M1	6.58	3.83	2.75	O14_D5	3.31	1.69	1.62
H_M4	4.58	**2.17**	**2.42**	O14_F2	2.15	**0.92**	**1.23**
H_M11	0.83	0.42	0.42	O14_F3	1.00	0.54	0.46
H_M10	0.50	0.25	0.25	O14_F4	0.62	**0.23**	**0.38**
				O14_D6	0.69	0.46	0.23

2.4.2 Individual Network Analysis

From the individual network's perspective, the analysis can be divided into the following two situations: point center degree and intermediate center degree.

In social network analysis, point degree centrality refers to the number of direct connections between a node and other nodes in the network. In directed networks, it includes both in-degree and out-degree. In social networks, the more direct connections between a node and other nodes, the higher the centrality of that node in the social network, which is usually represented as C_D. The degree centrality $C_D(V_i)$ of node V_i is defined as:

$$C_D(V_i) = \frac{k_i}{N - 1} \tag{1}$$

where N represents the number of nodes.

In this model, the degrees of the nodes represent the number of passes. By substituting the data into the model, we can calculate the point degree center of the passing network for each team player and normalize them. The statistical results are summarized in Table 2.

The value may be greater than 1 because of possible loops in the calculation process. In terms of player node degree, the Huskies have the highest node degree of F2, followed by M1 and D2; M1 has the most significant node degree in the O14 team, followed by D4, D3, and D1, all of whom are the core parts of the network. Combing with the role of

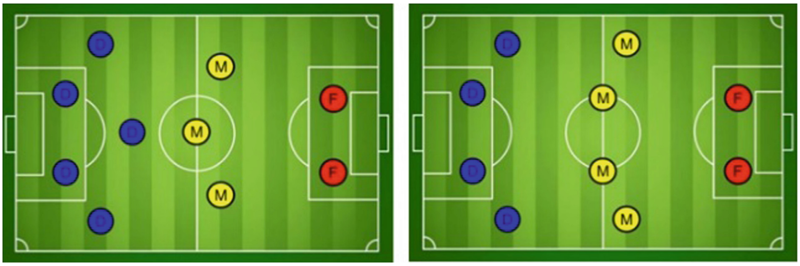

Fig. 3. Social network diagram of the O14 and Huskies team in game 14

the players, we can find that the top two players in the Huskies are forwards and centers, the top four players of the O14 team are midfielders and defenders. It indicates that the O14 team mainly adopts the defensive strategy of center forward. With the Huskies player's press, the O14 forwards could not pick up the players in the center back for the first time. The O14's passing was limited in the center back, and it wasn't easy to advance the ball to the front area.

The two teams also reflect a significant difference in in-degree and out-degree of the node. For defensive players, the node's out-degree is greater than the progress, which means that the player represented by this node has more successful tackles. For offensive players, the node's in-degree is greater than the out-degree, which means that the player designated by this node has more successful tackles.

As shown in Table 2, from an overall perspective, the Huskies' M10 and M11 players and the O14 team's F4, D6, and F3 players are considered substitutes due to their low pass rates. From the defenders' perspective, the Huskies' D6, D7, and D2 player and the O14's D3, D4, D5, and D6 player had a higher strike rate than their goal rate, indicating that they were better defenders. From the point of view of forwards, except for F2 player of Huskies and F3 player of O14, all the forwards are more offensive than defensive. From the midfield perspective, it can be found that M6, M8, and M4 are typical offensive midfielders, while M1 is the defensive barrier in the midfield and takes on the defensive duties in the midfield.

We can find that the tactical arrangement of the Huskies' coach is 4-4-2 as shown in Fig. 3, i.e., four defenders, four midfielders, and two forwards, by combining the above analysis and charts. However, in practice, the three midfielders are more active, changing the formation to a 4-3-3 attacking formation and changing the two midfielders on the sides to two wingers appearing deeper behind the defenders, making the attacking ability more powerful. In contrast, the O14 team coach adopted a 2-3-5 formation, as shown in Fig. 3, where the midfielders are defensive, focusing too much on defense at the back and neglecting the importance of offense, resulting in too much passivity and finally losing counterattack opportunities.

In Fig. 3, when ignoring the goalkeeper, the left is the 2-3-5 formation of the O14 team, and the right is the 4-4-2 formation of the H team.

2.4.3 Individual Network Analysis Based on the Betweenness

Freeman, an American sociologist, introduced the concept of closeness and betweenness in 1979. It refers to the degree to which a node is in the "middle" of other "point pairs". For example, the location of a point on a shortcut to many different pairs of nodes indicates that the node has a high degree of betweenness. The betweenness of a point represents the maximum control that a node has over other nodes' interactions.

We substitute the data into the model and calculate the betweenness of each player. Locally, players with greater betweenness can better "connect" with their teammates and act as "hubs" in the passing network, better driving the pace of play and leveraging the players' flexibility and initiative. The average fit between players is more significant, which facilitates the cooperation between players and drives the whole team to actively participate in passing and receiving, driving the pace of play of the entire team and improving the fluency of passing. The Huskies has the most prominent M1 centrality, followed by D5 and M6; the O14 team has the most prominent D1 centrality, followed by M1 and M3; this indicates that these players occupy the structural holes of the team's passing network and can effectively connect with other teammates. The O14 team cannot support the forwards in time because it is the defenders, not the midfielders, who have the closest relationship, which reduces the on-court passing network. The low level of coordination between most of the forward players in the O14 team suggests that the O14 team forwards consistently struggle to initiate attacks and make more connections with other teammates. Overall, the Huskies were able to maximize the coordination between players by building a more fluid passing network. It allowed players to pick each other up in the first place, synergize offensively or defensively and beat the O14 by a score of 4-0. It fully reflected the identity and performance of the players. Furthermore, it played an essential role in the team's pace of play and the passing network's fluidity with a higher degree of coordination.

3 Conclusion and Further Work

Through the Huskies and related football game analysis, we find that teams focus more on coordination and communication than individuals. For the management of team members, it is necessary to adapt to the development of modern information technology and meet talent training and team management under The Times' background. When building a team, it is necessary to actively forward innovative thinking and scientific thinking and supplement and improve each link and content.

Acknowledgements. This work was supported by the Humanities and Social Science Foundation of the Ministry of Education of China (Grant No. 17YJCZH178) and National Social Science Foundation of China (Grant No. 19BTJ031).

References

1. Buldú JM, Busquets J, Echegoyen I, Seirul.lo, F (2019) Defining a historic football team: using network science to analyze Guardiola's F.C. Barcelona 2019(9):13602
2. Xin J (2015) Research on the application of social network analysis method in the field of library and information. Intellectual Property Press Beijing
3. Li B, Wang L (2017) Feasibility analysis of social network analysis on passing performance in football matches. J Beijing Sport Univ 40(08):112–119
4. Li X, Li D (2005) DBSCAN spatial clustering algorithm and its application in urban planning. Sci Surveying Mapping 03:51–53