Research on The Offensive Characteristics of La Liga Team Based on Social Network Analysis

Wei Du[1,*], Jingwen Li[2], Feiyu Li[3]

[1] Manchester University Oxford Road, Manchester M13 9PL, UK
[2] Manchester University Oxford Road, Manchester M13 9PL, UK
[3] Manchester University Oxford Road, Manchester M13 9PL, UK

Abstract. To explore the difference of social network parameters between the network of passing before scoring and the network of passing before missing the goal, and to explore the correlation between social network parameters and team performance, this paper establishes the offensive pass network of 20 teams in the La Liga from 2017 to 2018, and 11 social network parameters are calculated. The Pearson correlation test is used to explore the linear correlation between 11 social network parameters and team performance. The results show that the linear correlation between the network parameters of passing before scoring and team performance is stronger than the network parameters of passing before missing the goal. According to the results, we can provide reliable and effective information to the football coaches to help improve the performance of football matches.

1 Introduction

Social network analysis is widely used in the performance analysis of football matches, but only a few scholars focus on the passing network in the process of attack. In football matches, the formation of every efficient attack requires multiple accurate passes, and the more effective the attacks are, the more likely the game is to win. The use of social network analysis methods in foreign countries to explore the performance of football has been very mature, and the relationship between corresponding network parameters such as network density, network centrality of individual players and team performance has been systematically sorted out [1]. There are relatively little researches on offensive patterns and social network analysis of offensive processes. Here are some related studies. Based on the passing position information, Rein et al. proposed spatial control parameters to explore the efficiency of the offensive link [2]. Mclean et al. used social network analysis (SNA) and notational analysis (NA) to study the goal-scoring passing networks (GSPN) of all UEFA Champions League matches in 2016. They found the GSPN network. Density, coherence, connectivity, and continuity are all relatively low. At the same time, it is found that there is a certain statistical relationship between network parameters and game conditions [3]. Kubayi studied the scoring tactics of all teams in the 2018 FIFA World Cup and analyzed their scoring patterns [4].

Although social network analysis has been widely used in the academic analysis of football sports, my country's research in this area is still in its infancy. Some scholars have only explored the feasibility of social network analysis in passing performance analysis. [5]; Some scholars, like those abroad, use the World Cup as the research object to do some offensive strategy classification and offensive technical and tactical analysis of football matches [6]. The most relevant study used a complex network analysis method to explore the impact of the Chinese Football Super League home and away games on sports performance [7]. However, at present, few researchers are using social network analysis to simply analyze the offensive passing network that forms shots in football matches.

2 Research objects and methods

2.1 Research object

This article takes as the research object all the passing incidents of 20 teams in the Spanish First Division Football League for a total of 380 games throughout the season and analyzes the offensive passing networks of all 20 teams in La Liga to explore different social situations. The relationship between network parameters and team record.

2.2 Research methods

2.2.1 Offensive effect evaluation standard

This study uses the conversion of the ball as a sign of the beginning of an offence, and the completion of a shot represents the end of an offence. All consecutive passes of a team during this period are defined as the pass attack

*Corresponding author: 1280538273@qq.com

network before the shot. It should be noted that in this process, whether it is through the pass organization, the process of preparing for the attack or the process of executing the offence through the pass, it is a part of the football offence [8]. Therefore, the definition of an offensive pass before shooting is reasonable.

Besides, this study adopts the shooting result as the evaluation standard of an offensive effect. Furthermore, whether a shot is hit or not can most intuitively and scientifically reflect the pros and cons of an offensive effect. Therefore, this study uses whether the shot hits or not as an evaluation criterion for an offence.

2.2.2 The method of establishing the passing network

Considering the the data sample size is too large, it is not suitable to use social network analysis software such as Ucinet or Gephithat manually enter data. Social network establishment and calculation of social network parameters. This research is based on the networkX of the Python platform, which processes a large amount of data in batches. First, perform data cleaning and preprocessing on the relational data in JSON format from figureshare, extract and identify the model objects analyzed in this research, and then divide the model objects according to 20 teams and 2 kinds of offensive effects (shooting goals and Factors such as shooting but not scoring) are classified into 40 subsets, respectively imported into networkX, and 40 directed graphs are generated (example shown in Figure 1), networkX can visualize graphs and calculate social network parameters, and data analysis The efficiency has been significantly improved.
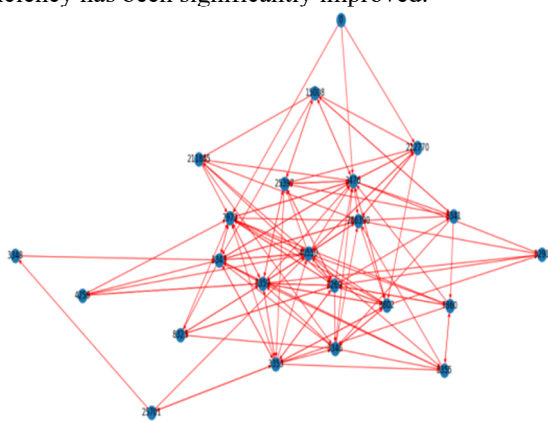

Figure 1 Schematic diagram of the Barcelona team's goal and pass network in the 2017-2018 season

3 Research results

3.1 Statistical analysis results

3.1.1 The statistical difference analysis results of the two types of shooting and passing network parameters

The average value and variance of each parameter of 11 different network parameters of all 20 teams in La Liga are calculated by the offensive network of shots and goals

and the offensive network of shots that have not been entered.

Divide the average and variance of the parameters of the goal network and the goal failure network (as shown in Table 2), and use these two ratios as the basic difference analysis index for evaluating the offensive network that causes the two shooting results.

Through the results of the ratio, it can be seen that the average of each parameter has a significant difference. Since scoring a goal is a small probability event in a football game, the number of shots and missed shots is much higher than the number of goals scored. Therefore, the size of the shot without entering the offensive network is much larger than the shot on the offensive network. Among the 11 network parameters studied in this paper, the total number of nodes, total connections, and network density are three indicators that are directly determined by the size of the network. Therefore, these 3 parameters should be eliminated when comparing network differences based on average values.

After excluding these three parameters, it is obvious that the averages of the other eight parameters have a huge difference. This result shows that there is a significant difference between the network characteristics of shooting goals and not shooting goals. This discovery lays a theoretical foundation for the subsequent correlation analysis. However, the variance of the network parameters of the two types of networks is difficult to find obvious rules. More than half of the network parameters have a lesser variance than the variance of the shot-to-goal network.

3.1.2 Correlation analysis results of the two types of passing networks before shooting and the team reco

Correlation coefficient 1 refers to the Pearson correlation coefficient between the network parameters of the forward pass network and the team record; correlation coefficient 2 refers to the Pearson correlation coefficient between the network parameters of the missed forward pass network and the team record.

Summarizing points are linearly positively correlated with team record, and the degree of correlation is moderate; the total number of connections is linearly positively correlated with team record, and the degree of correlation is strong; network density is linearly positively correlated with team record, and the degree of correlation is strong Correlation; the average clustering coefficient is linearly positively correlated with the team's record, and the degree of correlation is strong; the average reciprocity is linearly positively correlated with the team's record, and the correlation degree is strongly correlated; the average transitivity is linearly positively correlated with the team's record , The degree of correlation is a strong correlation; the global connectivity is linearly negatively correlated with the team's record, and the correlation degree is moderately correlated; the average degree centrality is linearly positively correlated with the team's record, and the correlation degree is very weakly correlated; the average neighbor degree center There is a linear positive correlation between the team's record and the team's

record, and the degree of correlation is moderate; the average close centrality is linearly positively correlated with the team's record, and the correlation degree is weakly correlated; the average betweenness centrality is linearly positively correlated with the team's record, and the degree of correlation It is very weakly correlated or not correlated.

The relationship between the network parameters of the missed front pass network and the team record:

Summarizing points are linearly negatively correlated with team record, and the degree of correlation is moderate; the total number of connections is linearly negatively correlated with team record, and the degree of correlation is moderately correlated; network density is linearly positively correlated with team record, and the degree of correlation is weak Correlation; the average clustering coefficient is linearly positively correlated with the team record, and the correlation degree is weakly correlated; the average reciprocity is linearly positively correlated with the team record, and the correlation degree is very weakly correlated; the average transitivity is linearly positive with the team record Correlation, the degree of correlation is medium correlation; the global connectability is linearly positively correlated with the team record, and the correlation degree is medium correlation; the average degree centrality is linearly negatively correlated with the team record, and the correlation degree is very weakly correlated; average neighbor degree Centrality is linearly negatively correlated with team record, and the degree of correlation is weakly correlated; average close centrality is linearly positively correlated with team record, and correlation degree is extremely weak; average betweenness centrality is linearly negatively correlated with team record. The degree of correlation is medium.

4 Analysis and discussion

4.1 Theoretical and practical significance of the research results

4.1.1 Theoretical significance

At present, more and more scholars or sports enthusiasts have come into contact with or even devote themselves to the subject of sports performance analysis for various reasons. The data processing method used in this article, social network analysis, is one of the most commonly used sports performance analysis methods. One is the theory most often used to process passing data in team sports. Although social network analysis belongs to the social sciences, it was first used to explore theories of communication modes between people, groups and groups, or organizations and organizations. But with the development of technology, all fields are inseparable from the support of data, and data analysis has become an indispensable skill in any field, and the sports field is no exception. In the papers using social network analysis to analyze sports performance, scholars from all walks of life use this method to realize their unique ideas and express their unique insights. More and more ideas collide in this

field, and the development of sports will be stimulated, and the development of the whole field will bring more scholars into contact with this discipline. This forms a virtuous circle.

4.2 Discussion of statistical analysis results

This paper uses two statistical methods to explore the differences in social network parameters between the front pass network and the missed front pass network, and the correlation between the social network parameters and the team record.

In the part of exploring the parameter differences of the two types of pre-shooting passing networks, the most basic statistics are calculated: average and variance and the two types of networks are compared. The comparison method is division. The calculation results show that the ratio of the average value of the parameters of the pass network before the shot is very small to the average value of the parameters of the pass network before the shot is missed, which means that the average value of the parameters of the pass network before the shot hits is significantly smaller than that of the missed shot. The average value of the parameters of the mid-to-forward pass network. According to the understanding of the two types of offensive networks extracted during the data processing, the reason for this phenomenon maybe since the number of missed shots in the actual game under normal circumstances is far more than the number of shots hit, that is, the shot missed the pass before the goal. The size of the network is much larger than the size of the pass network before the shot hits, resulting in the former's network complexity being much greater than the latter. Therefore, there will be a huge difference in network parameters. Based on this result, it can be boldly guessed that stable network performance will not be a key factor in winning the game. The reason for this phenomenon may be that the game situation is always changing during football games. Perhaps the unstable offensive characteristics are the offensive strategies that can help the team win.

In exploring the correlation between the parameters of the two types of pre-shooting passing networks and the team record, the Pearson correlation test is used. According to the research results of this article, two predictions can be made: first, in addition to the 11 social network parameters discussed in this article, there must be more social network parameters that are relevant to the team's record; second, use others The correlation test method must also be able to find multiple social network parameters related to the team's record. Based on these two conjectures, future research can focus on network parameters related to the team's record. First, a variety of correlation analysis methods can be used to find network parameters that are highly correlated with the team's record, and then multiple regressions can be used to establish a regression model of network parameters and the team winning percentage. It is even possible to use highly correlated network parameters like features, and build a football match result prediction model based on the passing network through machine learning algorithms.

5   Conclusion

There is a significant difference between the network parameter characteristics of the forward pass network and the network parameter characteristics of the forward pass network. The team can train according to the information feedback from the social network to improve the shooting percentage.

The linear correlation between the 11 social network parameters of the front pass network and the team record is stronger than the linear correlation between the network parameters of the missed front pass network and the team record, indicating that the pass before the shot hits The network has more research value in improving the overall sports performance of the team than the passing network of missed shots.

References

1. Clemente, F.M., Martins, F.M.L., & Mendes, R.S. (2016) Social network analysis applied to team sports analysis, Springer International Publishing, Cham.

2. Rein, R., Raabe, D., & Memmert, D. (2017) "Which pass is better?" Novel approaches to assess passing effectiveness in elite soccer. J. Human Movement Science, 55: 172–181.

3. Mclean, S., Salmon, P. M., Gorman, A. D., Stevens, N. J., & Solomon, C. A. (2018) social network analysis of the goal scoring passing networks of the 2016 European Football Championships. J. Human Movement Science, 57: 400–408.

4. Kubayi, A. (2020) Analysis of goal scoring patterns in the 2018 FIFA World Cup. J. Journal of Human Kinetics, 71, 205–210.

5. Hagberg A., Schult D., Swar P. (2021) NetworkX Reference, https://networkx.org/documentation/latest/_downloads/networkx_reference.pdf. 2021-05-19.

6. Ashenagar, B. & Hamzeh A. (2019) Forming a well-connected team of experts based on a social network graph: a novel weighting approach. J. Social Network Analysis and Mining, 9(1).

7. Fan, Y., Agyenim, B., Kim, L., & Jiang, Y. (2021) Are bonds blind? Board-CEO social networks and firm risk. J. Journal of Corporate Finance, 68.

8. Chang, C. (2016) Based on analytic hierarchy process model of big data analysis of the NBA schedule impact on team ranking. J. Journal of Computational and Theoretical Nanoscience, 13(12).