Complex network

Yangxin Tan*, Junlin Wu, Qing Zhong
Central University of Finance and Economics Beijing, China
*tan6499322@foxmail.com

Abstract—With the improvement in the accuracy of solutions and the increase in the complexity of social problems, social division becomes more important, and teamwork has become an important way to solve complex problems. The efficiency of a team is affected by many factors. How to evaluate the operation of a team is a very meaningful topic. A team consists of members. The characteristics and abilities of members determine their roles in a team, and naturally affect the efficiency of the team. The cooperation between team members is also an important factor in achieving high efficiency.

1. INTRODUCTION
With the continuous improvement of science and technology, many technical instruments to observe players' performance have been added to the analysis of sports events. In previous studies, many articles have started to analyze the data of sports events and predict the results of sports events. From the perspective of the general trend, j. linger (2009) proposed the performance efficiency rating in NBA, which is an effective team evaluation system construction in the financially powerful NBA. Later this kind of research also gradually into tennis, football and other sports.

The team nature and network nature of football is also an important aspect to measure the overall strength of a team. The research in this field is also being carried out slowly. Relying on the development of various imaging technologies, many team sports with passing nature, including ball games, are also being explored continuously. In the research of Borrie (2002), the similarity of passing sequences was concerned. Subsequent scholars also analyzed the aggregation and player trajectory, and Gyarmati (2014) considered the influence of team style indicators. In the study of j.m.quets, j.quets, I.E. Chegoyen and f.sibirul. lo (2019), team network organization was compared with other teams. Determines the statistically significant differences in indexes, then focus on the characteristics of passing network computing system in a game of evolution rather than a simple average, can win the match network attributes, including, the shortest path length, clustering coefficient maximum eigenvalue, the algebraic connectivity and the adjacency matrix of the distribution center, reflect the characteristics of the Barcelona team. The study also involves the measurement of team performance at different times in a game, further making the evaluation system dynamic. In addition, in Firat's (2018) study, the ternary and quaternary structures in the passing network were analyzed in detail, which proposed a more microscopic perspective for the passing effect of football teams. In the study of Paolo Cintia et al. (2015), the passing of players was measured by multiple indicators, which were integrated into an H index and evaluated each team. The results were close to the actual results, indicating that passing is an important support for studying team performance. In our research, we try our best to incorporate the previous research into the indicator system to improve the accuracy of the final model prediction.

We have a richer indicator system, and most of the papers only consider the percentage of possession, shots, free kicks, fouls, passes and steals of the home team. These abilities all emphasize the ability of a
teammate, or simply the average of a team's abilities. In our study, we quantified indicators of team movement fluency, i.e. the number of pathways in the network that can be uninterrupted, and the frequency of pathways of different lengths. A team with a high frequency of short lanes is more aggressive, while a team with a low frequency of long lanes is more focused on square revenue. The research of network patency is the innovation of this research.

In recent years, the application of machine learning in prediction has become more and more common. A special issue of machine learning magazine shows the results of a "football prediction challenge". The study, based on a machine learning model, predicted the outcome of 206 football matches in the future (Berrar, Lopes, Davis, & Dubitzky, 2019). Some of the contestants achieved encouraging results. Hubáček, Šourek, and Železnyy (2019) using gradient boosted trees and win the game (Constantinou 2019) to complete the second using the Dolores model of combining with hybrid bayesian network dynamic rating. Others used machine learning in the past to predict football (Van Haaren & Van den Broeck, 2015). In our study, we compared the prediction accuracy of different methods and selected the best prediction model.

In this paper, we refer to the framework of Paolo Cintia et al. (2015) and decide to use some indicators to evaluate the team's ability. However, we integrated more perspectives. In the analysis, we evaluated the network characteristics from the perspective of the overall operation of the team, and evaluated the individual abilities of the players from the perspective of the individual players, and made improvement and development by referring to the indicators proposed in many studies. Next, we substituted the data into the principal component analysis and classified it into three scores with certain meanings. Then, according to the machine learning model and the whole index system, the results of the competition are predicted.

In our study, our index assessment is more comprehensive and complete, which is more complete than previous studies. At the same time, this model has a strong application, using the method of scoring to facilitate comparison, and the detailed indicators can also reflect the team's ability in different aspects. In the follow-up research, we should first continue to develop a dynamic analysis of the performance of teams in a game. Secondly, in this paper, both principal component analysis and machine learning rely on the characteristics of the data itself to carry out the analysis, which needs to be adjusted again for different data sets, and horizontal comparison is difficult.

2. INDICATOR SYSTEM

2.1. Complex Network and Network Indicators

2.1.1. Dyadic Configuration analysis in terms of Passing Ability

Through the relevant data analysis, we find that the soccer match result has a significant relationship with the difference in the number of passes made by the two teams in one game. We define it as follows: the difference in the number of passes made by the two teams, represents the difference in the passing abilities of these two teams in a certain game (Passing Factor):

$$PF = P_{Huskies} - P_{Opponent}$$

On the other hand, we use the scores difference between two teams in a certain game to measure the final result:

$$ScoreDifference = OwnScores - OpponentScores$$

Finally, the regression analysis of PF and Scores difference shows that neither the results of linear equation, nor the quadratic equation or the cubic equation are satisfactory.
Furthermore, statistical analysis of each points on the coordinate axis shows a table1 as follows. Assuming the number of passes is the only factor to determine whether a team can win or not, we predict that a team will win when PF is greater than zero, and will tend to lose when PF is less than or equal to zero. According to the form below, we can found that the accuracy of this judgment rule is only about 50%, which is not much reliable, but it still can be a bottom line to judge whether the model built later is valuable.

**TABLE I. PREDICTION ACCURACY OF PF BASED ON SIMPLE CRITERIA**

| PF  | Total Rounds | WIN | TIE | LOSE | Accuracy |
|-----|--------------|-----|-----|------|----------|
| > 0 | 18           | 9   | 4   | 5    | 0.5      |
| <= 0| 20           | 4   | 6   | 10   | 0.5      |

Although the results are not so accurate, we can still conclude that the difference in the total number of passes does have a significant impact on the final match results. Thus, we define the total number of passes as a factor $\theta$.

On the other hand, we also need to realize that some teams may have star players taking charge of passing the ball, while other teams may assign the passing tasks to everyone. Based on the above different situations, we obtained the mean and variance of passing numbers of players of each team through statistical analysis. With the exception of the Huskies who played a total of 39 games, all the other teams had only two games, therefore, we averaged the data of all teams, as shown in the table2.

**TABLE II. THE NUMBER OF PASSING OF DIFFERENT TEAMS**

| Team          | Huskies  | Opponent1 | Opponent2 | Opponent3 | Opponent4 | Opponent5 | Opponent6 |
|---------------|----------|-----------|-----------|-----------|-----------|-----------|-----------|
| Average passing | 18.33    | 21.68     | 80.81     | 60.56     | 48.12     | 48.86     | 27.67     |
| Passing standard | 15.76    | 13.80     | 60.62     | 35.08     | 28.62     | 30.08     | 17.82     |
| Goals         | 2.32     | 1.00      | 4.00      | 2.00      | 7.00      | 6.00      | 3.00      |

| Team          | Opponent7 | Opponent8 | Opponent9 | Opponent10 | Opponent11 | Opponent12 | Opponent13 |
|---------------|-----------|-----------|-----------|------------|------------|------------|------------|
| Average passing | 29.93    | 23.63     | 52.56     | 27.06      | 28.85      | 35.89      | 44.64      |
| Passing standard | 12.36    | 16.01     | 38.82     | 14.65      | 20.83      | 25.49      | 35.27      |
| Goals         | 3.00     | 1.00      | 10.00     | 3.00       | 3.00       | 3.00       | 5.00       |

| Team          | Opponent14 | Opponent15 | Opponent16 | Opponent17 | Opponent18 | Opponent19 |
|---------------|------------|------------|------------|------------|------------|------------|
| Average passing | 37.92    | 27.47     | 58.68      | 31.68      | 36.03      | 27.73      |
| Passing standard | 27.21    | 20.51     | 41.30      | 21.93      | 19.83      | 15.84      |
| Goals         | 3.00      | 0.00      | 1.00       | 0.00       | 2.00       | 1.00       |
By an regression analysis using the relevant data, we find that the score is positively correlated with the average number of passes made by the overall players. So a team may have a higher chance of scoring by having more passes in a game. On the other hand, the greater the share of passing in the team is, the lower the scoring ability of the team will be. When the task of passing is concentrated on some star players, the team has a higher chance to score. We define the average number of passes as $\mu$ while the sample standard deviation of the team passing frequencies is defined as a factor $\sigma$.

![Fig. 2. Regression between Goals and AveragePass or PassingStandard](image)

Based on the above analysis, we define Passing Ability as follows:

$$PA = \frac{1}{10} \left(6 + \mu + \sigma \right)$$  \hspace{1cm} (1)

Assuming that in any given match, the Huskies will win if the PA value is large, and draw or lose if the PA value is small. The Huskies end up with a total of 25 out of 38 matches correctly, giving the indicator a 65.8% accuracy. By comparing the accuracy of the three factors alone, FA has the highest accuracy as a judgment indicator, but it can be found that the improvement of accuracy contributes little, but we still take it as a judgment factor of our passing ability.

2.1.2. Triadic Configuration analysis

In football matches, we often encounter steals and defensive situations, so the binary structure of passing will be broken at this time. However, if there is a tight ternary configuration between the team or some players, then the team can successfully break through other teams’ defense, being more offensive and flexible. Here, we create a new indicator to measure the triadic configuration of the team, analyze the number of triangles created by any three players, and evaluate the network characteristics of the team.

We use Compactness of ternary configuration to analyze the tightness of a team's triadic configuration, which is based on the number of passes made between players in a match. The Formula is as follows:

$$C(i) = \frac{\sum_{j,k} n_{ij}n_{jk}n_{ik}}{\sum_{j,k} n_{ij}n_{jk}}$$  \hspace{1cm} (2)

The analysis results of the 20 ball pairs are shown in the table3.

| Team         | C         |
|--------------|-----------|
| Opponent2    | 646.5     |
| Opponent17   | 496.5     |
| Opponent3    | 484.5     |
| Opponent5    | 455       |
| Opponent10   | 447       |
| Opponent14   | 428.5     |
| Opponent4    | 409       |
| Opponent15   | 361       |
| Opponent13   | 323       |
| Opponent18   | 301       |
| Opponent12   | 288.5     |
We can conclude that the Huskies' situation of ternary configuration is not ideal, so the team needs to strengthen the passing coordination, improve the team's offensive flexibility, and improve the attacking path.

2.1.3. Team Formations

As shown figure3, we plotted the passing trajectories between Huskies teams throughout the season and the passing trajectories in game 14 (the game with the highest score differential of the season). In the process of drawing the following figure, we processed the number of passes, converting the absolute number of passes to a percentage of the season or game. The conversion into percentages makes the composition of our team more intuitive. In the figure, the yellow line represents the pass path which is less than 20% of the total pass for the game or the season, and the blue line represents the pass path which is greater than or equal to 20%, we define the blue lines as the critical path.

![Complex Network of the Huskies team](image)

These blue lines can better reflect the team's strategy and the teamwork situation in the whole season or a particular game.

First of all, Degree Distribution means the number \( k_i \) of other nodes connected to a certain node \( k \), representing the number of other players passing balls to a certain player. We denote it as \( \langle k \rangle \), and the average degree of the whole network is

\[
\langle k \rangle = \frac{1}{N} \sum_{i=1}^{N} k_i
\]

(3)

The larger the degree of a node is, the more important the node is in the network, that is, the player in the team is similar to the role of a "Star".

By exploiting the presence of stars in the team, we were able to develop effective defense strategies that blocked the channels for opposing star players to play a role. On the other hand, we also explored the potential relationship between the network average and the final score. The average degree of each team is shown in the figure4. It can be found that the average degree of most teams is around 9. Through simple linear regression, it can be seen that the average degree of a team has a positive correlation with the final score to some extent.
TABLE IV. AVERAGE DEGREE OF DIFFERENT TEAMS

| Team       | Average Degree | Team       | Average Degree |
|------------|----------------|------------|----------------|
| Huskies    | 8.83           | opponent10 | 9.47           |
| opponent1  | 9.05           | opponent11 | 8.65           |
| opponent2  | 11.88          | opponent12 | 8.89           |
| opponent3  | 10.19          | opponent13 | 9.56           |
| opponent4  | 10.47          | opponent14 | 9.00           |
| opponent5  | 10.17          | opponent15 | 9.05           |
| opponent6  | 9.22           | opponent16 | 10.24          |
| opponent7  | 9.21           | opponent17 | 9.32           |
| opponent8  | 8.11           | opponent18 | 9.44           |
| opponent9  | 10.35          | opponent19 | 8.53           |

Fig. 4. On the left is the histogram of degree distribution, and on the right is the regression of score and average degree.

2.2. Non-network indicators

2.2.1. Game Style

A team with good coordination is able to pass the ball several times to avoid the opponent's steals and increase the chance of a successful shot. On the other hand, the low frequency of dribbling can also reflect the characteristics of a team's quick attack, while the intermediate frequency of dribbling can reflect a robust team style. Therefore, we divide the combat style of each team into three categories according to this hypothesis. Teams are classified according to the Passing combo (PC) of the dribbling times.

In all passing events, the highest combo record is 44 consecutive passes. It is obvious that there are significant differences in the strategies of different teams.

First, we apply the max-min method to the percentages of the same strategy in different teams.

\[
P^*_i = \frac{|P_{it} - P_{i\min}|}{|P_{i\max} - P_{i\min}|}
\]

(4)

i represents the team no. i of 20 teams, and t represents one of the three strategies.

Then we add up the updated percentages as \( P^*_i = \sum_{t=1}^{3} P^*_i \) and get the final score of different teams.

We get the final scores of each team as shown table 5.

TABLE V. THE DIVERSITY OF EACH TEAM

| The Team   | P *  | The Team   | P *  | The Team   | P *  |
|------------|------|------------|------|------------|------|
| Opponent19 | 1.000| Opponent10 | 1.248| Opponent14 | 1.508|
| Opponent6  | 1.079| Opponent15 | 1.290| Opponent3  | 1.550|
| Opponent8  | 1.137| Huskies    | 1.300| Opponent5  | 1.614|
| Opponent17 | 1.141| Opponent12 | 1.302| Opponent16 | 1.743|
| Opponent11 | 1.162| Opponent18 | 1.304| Opponent9  | 1.795|
| Opponent7  | 1.174| Opponent4  | 1.421| Opponent2  | 2.000|
| Opponent1  | 1.197| Opponent13 | 1.442|            |      |
Through this method we obtained the strategic diversity of each team, according to the regression test and the initial assumptions, we can find that teams with greater diversity are more likely to get more scores. To guarantee the consistency of all indicators, with the increase of indicators, the ability of the entire team is enhanced, we take reciprocal treatment to the diversity (D) as follow:

\[ \text{Diversity}_i = \frac{1}{P_i} \]  

(5)

2.2.2. Coordination among players & distribution of contributions

Combining the previous analysis, we defined cooperation level by three factors, the distribution of passes, the distribution to dribbles and the distribution to shoots.

Generally, the larger the standard deviation of these three abilities, the more a team's abilities are determined by some members. On the opposite situation, it means that the abilities or responsibilities within the team are evenly distributed. Based on these three standard deviations, we obtain the Coordination Index (CI) to determine the cooperative level of a team.

\[ CI = \text{STD}_{\text{Passing}} + \text{STD}_{\text{Dribble}} + \text{STD}_{\text{Shoot}} \]  

(6)

The standard deviation here has completed the maximum and minimum criterion normalization processing. And the scores are as table6.

| Team    | CI   | Team    | CI   | Team    | CI   |
|---------|------|---------|------|---------|------|
| Huskies | 0.534| opponent7 | 0.785| opponent14 | 0.749|
| opponent1 | 0.389| opponent8 | 0.564| opponent15 | 0.277|
| opponent2 | 2.182| opponent9 | 1.400| opponent16 | 1.222|
| opponent3 | 1.474| opponent10 | 0.598| opponent17 | 0.608|
| opponent4 | 2.337| opponent11 | 0.817| opponent18 | 0.324|
| opponent5 | 1.150| opponent12 | 0.721| opponent19 | 1.037|
| opponent6 | 0.850| opponent13 | 1.048|

Regression analysis of CI and the final number of shots of each team found that they have a significantly positive correlation. Therefore, we can think that the clearer the division of labor of a team, the more the players can focus on their role, the team will More likely to shoot and win the game.

2.2.3. Adaptability & Flexibility

Players need to adjust to different opponents and match conditions, which is also a reflection of the adaptability of the team. Whether a player can deal with the situation well has a strong relationship with the player's own ability. Skilled players can make adjustments easily and well. Here we observe the data records of players' movement in different teams, and evaluate the adaptability of players according to the number of football technical movement they can master in the game.as shown table7.

| Team    | Mean | SD |
|---------|------|----|
| Huskies | 12.8 | 1.4|
| Opponent4 | 13.5 | 1.5|
| Opponent9 | 13.1 | 1.6|
| Opponent5 | 12.9 | 1.4|
| Opponent2 | 12.7 | 1.3|
| Opponent10 | 12.5 | 1.4|
From the table above, we can conclude that the Huskies' categories of technical movement are still at the higher level of all teams. At the same time, the standard deviation is also relatively high, indicating that there is a certain difference in the ability of the players in the team.

2.2.4. Tempo & Flow

If a team is able to pass the ball smoothly and each player's pass movements can be well connected, the team will have a greater possibility to eliminate the opponent's steals, promoting the chances of shooting, avoiding missing opportunities caused by hesitation. We calculate the number of consecutive passes completed by a team, which is the number of the events whose initial point is consistent with its last event’s end point. Then, we calculate the ratio of this number to the total events number to represent the flow factor of a team as $T_1$. The maximum number of the consecutive events for a team is the flow factor. $T_2$

$$T_1 = \frac{\text{Successive Events}}{\text{Total Event}}$$

$$\text{Successive Events} = \sum_{i=2}^{62} \text{Combo} \times \text{Occurrence}, \ k = 1, 2, 3...20$$

$$T_2 = \text{MAX} (\text{Combo}), i = 1, 2, 3...62$$

Among them, the maximum number of consecutive events recorded is 62. $i$ represents a constant between 1 and 62, and $k$ represents the NO.$k$ team in a total number of 20 teams. Since a consecutive event should contain at least 2 events, the Combo is counted from 2.

Finally, the teams' flow factor scores are obtained as in the following table8:

**TABLE VIII. THE TEMPO OF EACH TEAM**

| The Team   | T1  | T2  | NormalizedT1 | NormalizedT2 | Scores |
|------------|-----|-----|---------------|---------------|--------|
| Huskies    | 0.86| 41  | 0.38          | 0.56          | 0.94   |
| opponent1  | 0.84| 20  | 0.29          | 0.13          | 0.42   |
| opponent2  | 0.94| 62  | 1.00          | 1.00          | 2.00   |
| opponent3  | 0.92| 20  | 0.88          | 0.13          | 1.00   |
| opponent4  | 0.90| 37  | 0.75          | 0.48          | 1.22   |
| opponent5  | 0.89| 36  | 0.68          | 0.46          | 1.14   |
| opponent6  | 0.87| 20  | 0.47          | 0.13          | 0.60   |
| opponent7  | 0.81| 23  | 0.00          | 0.19          | 0.19   |
| opponent8  | 0.85| 20  | 0.31          | 0.13          | 0.43   |
opponent9 0.91 32 0.80 0.38 1.18
opponent10 0.86 32 0.39 0.38 0.77
opponent11 0.87 27 0.51 0.27 0.78
opponent12 0.89 19 0.67 0.10 0.77
opponent13 0.88 38 0.59 0.50 1.09
opponent14 0.89 26 0.61 0.25 0.86
opponent15 0.85 19 0.32 0.10 0.43
opponent16 0.92 25 0.90 0.23 1.13
opponent17 0.88 21 0.58 0.15 0.72
opponent18 0.85 24 0.36 0.21 0.57
opponent19 0.85 14 0.30 0.00 0.30

Judging from the results, Opponent2 has the best fluency, the Huskies team’s fluency level is at a moderately high level, the opponent7’s fluency is of the worst. From the perspective of the regression results, the total score factor has better results than the single use of standardized T2, which is not significantly different from the regression results of standardized T1. Therefore, we choose the total score as the measure of flow degree.

2.2.5. Shooting ability
A team's shooting percentage is crucial in the game. The team with better shooting ability has more chance to win. Therefore, we explore the correlation between this factor and the game outcome. Firstly, we set Shotting Accuracy (SA) as the ratio of the number of successful goals scored by a team to the total number of attempts at shooting in a certain game:

\[
SA = \frac{\text{Goals}}{\text{Attempts}}
\]  

Goals is the number of points that a team actually scores in each game, and Attempts is the times a player try to shoot in a given game by events as 'shot', 'free kick shot', and 'penalty'.

There are three possible outcomes of a football game: win, tie, and loss, we assign them to be 1, 0, and -1. Through regression analysis of SA and Outcome, we found that SA was positively correlated with Outcome in a linear way (figure1 A), but the fitting effect was not ideal.

Actually, an excellent shooting ability is not enough to win a game. Even if a team has a high shooting percentage and considerable goals, if it lacks defensive capabilities, the opponent will easily score and win. Therefore, we should also consider the opponent's ability to score. We define Shotting & Defence Efficacy (S&DE) as the difference between Huskies' shooting ability and that of their opponents:

\[
S&DE = \frac{\text{Goals}_h}{\text{Attempts}_h} - \frac{\text{Goals}_o}{\text{Attempts}_o}
\]  

3. Predict and Discussions

3.1. Predictions based on Machine Learning
Through the above analysis, we have given a total of 10 indicators to describe the various capabilities of a football team. First, we converted the absolute values of the indicators into relative values, and in order
to refer to the gap between the Huskies team and other teams’ capabilities, we first subtracted the Huskies
team's indicators from the other teams' values, and then normalized the gaps under the minimum and
maximum criteria. The rankings of the Huskies team among the 20 teams are as shown table 9.

### Table IX. The Ranking of Each Indicator

| Ability                          | Ranking | Ability           | Ranking |
|---------------------------------|---------|-------------------|---------|
| Passing ability                 | 1       | Diversity of play | 13      |
| Compactness of ternary          | 1       | Coordination Index| 17      |
| configuration                   |         |                   |         |
| Team formation                  | 6       | Dribbling ability | 15      |
| Mean of distance                | 9       | Adaptability      | 3       |
| STD of distance                 | 16      | Tempo             | 3       |

First of all, from the rankings (after normalization, the larger the better), we can see that the four
indicators including Coordination Index, STD of distance, Dribbling ability, and Diversity of play need
to be improved.

However, simply from the ranking, we can not completely judge the ability that really needs to be
improved, because although the ranking of some skills may be too low, the improvement of this ability
cannot bring higher scoring ability. Therefore, we use several applicable machine learning Classifiers and
combine the standardized values of 10 indicators to anticipate the outcome of each game. When the
Huskies team is higher, we give The meta output variable WinTieLoss, for short WTL, is 1. When the
two sides are the same, the WTL value is 0. When the Huskies team is lower, -1 is given.

Put the data into each model, the accuracies of Test set in different models are as shown table 10.

### Table X. The Prediction Accuracy of Different Models

| classifier     | Train set | Test set | Accuracy of Test Set |
|----------------|-----------|----------|----------------------|
| Regression     | 14        | 5        | 50.00%               |
| Random Forest  | 14        | 5        | 57.40%               |
| Adaboost       | 14        | 5        | 20.00%               |
| Bagging        | 14        | 5        | 20.00%               |
| XGBoost        | 14        | 5        | 60.00%               |

All the classifiers are implemented based on Python. It can be found that XGBoost has better
performance in 5 models, so we chose XGBoost model for the next analysis. We draw the following
Synaptic Weight diagram of the XGBoost neural network, where the gray line represents a positive
Synaptic weight and the blue line represents a negative Synaptic weight. Looking back from the WTL =
1 end point on the right, you can find that in order to improve the possibility of winning, you need to
work harder on the abilities with yellow background in the ranking diagram, as shown as figure 5.
Among them, Passing Ability and Adapatability does not have more room for improvement, so it is necessary to work on another 5 abilities. Combined with the analysis in Network, the Huskies team needs to improve the stability of their formation, expand the entire team's passing network, and increase the team's average degree. Intuitively, it is to add edges in the network and ensure that each edge is sufficiently stable.

In Performance Indicators, we should focus on improving cooperation and dribbling capabilities. From the ranking, we can also see that there is likely to be a problem of unclear division of labor in the team. The coach should clearly recognize which role members are suitable for and play in this role. Make good use of their responsibilities. On the other hand, the basic skills of dribbling ability are not enough, and it is easy for the scoring opportunities to escape. In this respect, some improvement is also needed.

3.2. Discussions-Promotion to other sports

The evaluation model for Huskies starts from the passing network and looks for network characteristics. From the characteristics of each node to the overall structure and connection of the network. Next, the model specifically analyzes the team attributes related to the game victory, such as team adaptability, flexibility, fluency, etc. Based on these indicators, we have obtained a description of team capabilities, structure, strategy and closeness. Then, three important factors were obtained through principal component analysis, namely the factor, aggressive factor, coordination factor and stability factor. These three factors are a good description of the team's situation in a football match. Finally, the machine learning model is used to make predictions for future competitions and to provide relevant suggestions for development.

This process can be used in any area that describes team capabilities and nature. Using the NNF (Notes → Network → Forecast) framework proposed in this article, starting from the two aspects of team personal and internal team relations, find a series of indicators that cover all attributes of the team as
much as possible, and then scientifically integrate the indicators. Quantitative comparison of the situation of the team in all aspects. Finally, use a reasonable prediction model to make an outlook for the team's future development. The whole process logic is rigorous, scientific, comprehensive, and has strong application capabilities. Before evaluating the efficiency of the team, it is necessary to reasonably determine the team's indicators describing the team's efficiency, and then analyze it according to the NNF model. The horizontal comparison finds the areas where the current team is insufficient, and develops them in a targeted manner.

In addition, the model can be extended from three aspects. First, make the model dynamic. Pay attention to changes in indicators during team operations. By adding the time series of parameters, the partial analysis of the entire workflow is analyzed to obtain the dynamic changes of the team. Second, consider the psychological factors and use the adjustment coefficient M to adjust the team score. In teamwork, the atmosphere within the team and the relationship between the players also affect the efficiency of the team's work. In addition, when members are in an emergency, great psychological pressure will also affect the normal work. Third, incorporate external environmental factors and introduce adjustment parameters X. Facing different scenarios, including external factors such as environmental equipment climate, will affect the team's efficiency. For example, if a player who regularly trains in the plains participates in the game at the plateau stadium, the performance of the team will also be affected.

4. RESULTS
In order to make an effective prediction of the winning and losing of football matches, we built a complex network model, combined with the experience of former scholars, built many effective indicators, and innovatively built game style, tempo and other indicators. In order to avoid the bias of the indicators, we also adopted many non-network properties, such as passing ability, shooting ability, etc., when constructing the indicators, we also fully learned from the previous experience, and built innovative indicators such as Coordination Index. Therefore, this paper puts forward many new methods in the construction of the index system. On the other hand, in order to evaluate a team's ability in a general way, we have developed a method to evaluate a football team through principal component analysis. Eventually taking the whole index system as the input variables, combining Hubáček, Šourek, and Železny’s experience (2019), we applied XGBoost as the best prediction model, and achieve the result of 60% accurate prediction in the test set. We can believe that we could have a good chance of getting a good return on real football games.

REFERENCE
[1] Liao Bin, Zhang Tao, Guo Binglei, Yu Jiong, Niu Yafeng, Zhang Xuguang, Liu Yan. Quantitative Research on Pass Quality Based on GraphX Pass Network. Computer Science, 2017, 44 (12): 175-182 + 201.
[2] Zhang Yan. Application and Research Progress of Social Network Analysis in Football Performance. Chinese Sports Science Society. Compilation of Abstracts of the Eleventh National Sports Science Conference. Chinese Sports Science Society: Chinese Sports Science Society, 2019: 5773-5775.
[3] Zhong Linfeng. Research on Mining Algorithms for Key Nodes in Complex Networks. University of Electronic Science and Technology of China, 2018.
[4] Wang Jian. Research on community structure analysis algorithm based on complex network. Dalian Maritime University, 2015.
[5] Buldú, J.M., Busquets, J., Echegoyen, I. et al. (2019). Defining a historic football team: Using Network Science to analyze Guardiola’s F.C. Barcelona. Sci Rep, 9, 13602.
[6] Cintia, P., Giannotti, F., Pappalardo, L., Pedreschi, D., & Malvaldi, M. (2015). The harsh rule of the goals: Data-driven performance indicators for football teams. 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA), 1-10, 7344823.
[7] Duch J., Waitzman J.S., Amaral L.A.N. (2010). Quantifying the performance of individual players in a team activity. PLoS ONE, 5: e10937.
[10] Zhang Yujuan. Research on the team spirit cultivation strategy of college students. Northeast Normal University, 2019.
[11] Guo Chong. Key Points Detection of Football Field Based on Deep Learning. Beijing University of Posts and Telecommunications, 2019.
[12] Xu Yi. Research on machine learning algorithm and its application [D]. Hunan University, 2014.
[13] Song Wei. Research on Basketball Sports Data Analysis Based on Spark Framework and K-means. Chinese Academy of Sports Science. Compilation of Abstracts of the Eleventh National Sports Science Conference. Chinese Sports Science Society: Chinese Sports Science Society, 2019: 5795-5797.
[14] J. Hollinger, “The player efficiency rating,” 2009. A. Borrie, G. K. Jonsson, and M. S. Magnusson, “Temporal pattern analysis and its applicability in sport: an explanation and exemplar data,” Journal of sports sciences, vol. 20, no. 10, pp. 845–852, 2002.
[15] J. Gudmunsson and T. Wolle, “Football analysis using spatio-temporal tools,” Computers, Environment and Urban Systems, vol. 47, pp. 16–27, 2014.
[16] Cintia, Paolo, et al. "The harsh rule of the goals: Data-driven performance indicators for football teams." 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA'15) IEEE, 2015.
[17] Buldu, J. M, et al. "Defining a historic football team: Using Network Science to analyze Guardiola's F.C. Barcelona." Scientific Reports (2019).