Problem Chosen	2020 MCM/ICM Summary Sheet	Team Control Number
D		2012356

Teamwork capacity evaluation: using Topological passing network and Fuzzy estimation to evaluate huskies

Summary

This paper is based on the ability to evaluate the team's ability to cooperate with the team, the feature extraction of tactics and strategies, and finally put forward the reasons for winning, and analyze the team's various game data throughout the season.

For purpose of describe the characteristics of tactics and strategies, capture the team's scheduling during the match, an average passing network was constructed, and the network was constructed based on data such as the number of player passes and player distance. The value that can effectively reflect the player's personal level: player attractiveness. When the graph is visualized, we reflect it on the edge color, node distance, and node size. We then used the constructed passing network model to make a tactical analysis of the Huskies team and its opponents.

In order to evaluate the teamwork capacity to cooperate with the team, we have adopted a fuzzy estimation method and innovatively proposed to consider the complex situation in team cooperation with two levels of indicators. The first level of indicators is mainly in football sports competition technology. Level to analyze. We mainly select some indicators that can directly reflect the team cooperation ability, for example, ball control rate, number of shots, etc. Through the analysis of these indicators, the team's team cooperation ability is generally obtained. The second level of indicators, from the perspective of data, is based on the previously defined passing network and combined with relevant knowledge of topology for analysis. Then through correlation analysis, the correlation between these indicators at the first level and the indicators at the first level is calculated to obtain the correlation coefficient. The correlation coefficient is used to judge the degree of correlation with the low-level indicators, and then the team cooperation ability is tapped A deeper level of information. The indicators at this level are mainly the centroid, weighted average passing path, clustering coefficient, and binary clustering coefficient.

Based on the evaluation indicators of the passing network and teamwork capacity, we made recommendations for Huskies' training methods for next season, and measured how to build a more effective team from the perspective of network structure.

Keywords: Pass network, Fuzzy evaluation, Clustering coefficient, Topological analysis

Content

1 Introductions

1.1 Background

Humans are social animals, and it is impossible for everyone to exist independently without others. Nowadays, with the continuous development of an increasingly connected society, teamwork has become an indispensable part of people's survival and development in society, and success through effective teamwork is undoubtedly the topic that people are most concerned about. Although we are facing a series of complex and highly challenging teamwork success issues, we can rely on scientific, creative expertise and mathematical models to solve these problems.

The most typical example of teamwork is competitive team sports, such as football games. Because of the strict rules of team sports, cooperation between players is particularly important. Through research on teamwork in football games, we find that the success of the team includes many factors, such as team formation strategy, player interaction, coaching leadership style, personal contribution and collective performance.

Good team coordination will contribute to the success of the team's game, and analyzing team coordination problems from the failed game will also accumulate experience and provide suggestions for the team's game.

Analyzing and solving the cooperative problems of competitive team sports can not only provide tactics for the sport, optimize team performance and achieve success, but also provide new ideas and innovation points for exploring the team cooperation process in society. For this reason, we urgently need a complete scientific model to analyze the interaction between players and team strategies in football. On this basis, we have established networks and models for teamwork to analyze the factors and indicators that affect their success.

1.2 Restatement of the Problem

According to the requirements of the Husky team coach, we need to clean the data, analyze and process the data provided by the team, quantify the success of the team's cooperation, quantify the success of the team's cooperation, and use the various dynamic action behaviors of the team members and the team. The result of the interaction under cooperation (not only considering the victory or not of the game and the team's score), evaluate the team's teamwork throughout the season, so as to help the team adjust teamwork and improve next season. Cooperative efficiency provides specific strategies.

Task1： We need to create a network between players that reflects the position of team players, the relationship between various actions and football transmission, and use nodes and links to reflect the dynamics between teams and on the court. The network needs to reflect the structural indicators of teamwork, such as interactions between players and individual contributions of players, on multiple scales from micro to macro.

Task2： We need to find and quantify performance indicators that reflect teamwork. Performance indicators should not be limited to scores and victories. We should also conduct qualitative analysis of other team-level processes and strategies. We should also analyze the various behaviors and structural strategies of the team. We use multiple performance indicators to create an evaluation model to analyze successful teamwork and the structure, configuration, and dynamics of the team during cooperation, and draw conclusions.

Task3: We will start from the various action behaviors and structural strategies obtained from the data network and their results to evaluate and optimize effective structural strategies. Using the team cooperation model that can bring benefits to the team in the current season, combined with the structural strategies of the competing teams, it provides optimization suggestions for the team's plan for the next season.

Task4: Finally, we will summarize all the analysis to answer how to design more effective teamwork strategies and the aspects that need to be understood to develop performance index models. Use real analysis and evaluation data to explain the problem, make suggestions, and analyze Possible benefits after implementation. The dynamic and complex factors related to teamwork in football should also be combined with society to provide innovative suggestions for the development of social solutions to team problems.

1.3 Our Work

We define the attractiveness of the players according to the number of passes of the players and the coordination of the passes between the players to establish a passing network.

Then, the fuzzy analysis method was used to define 8 original indicators and 4 second-level indicators, and the teamwork capacity of the Huskies team was evaluated from two levels.

Combining the most commonly used tactical coordination methods of the team extracted by the passing network, we found the strengths and weaknesses of the Huskies team lineup and made recommendations to the coach.

2 Assumptions and Symbols

When constructing a passing network, ignore the rest of the fullevents except for passing. Some tactics in football involve deliberately creating fouls to get the ball out of bounds, simplifying this part of the ball-handling skills here.

The trajectory of the ball during the passing is a straight line from the passing point to the receiving point.

In team cooperation, only three or less players who have recently passed are considered. The position of football players is constantly changing. A micro-discussion of a small number of players can better reflect the teamwork capacity to cooperate.

In team cooperation, the interval from one pass to the next pass is less than five seconds. In team cooperation, the interval from one pass to the next pass is less than five seconds.

For brevity, we define a series of symbols for the formulas in Table 1.

Table 1: Symbol

Symbol	Definition
N	Total number of passes by the team in a game
N_e	Total number of edges of the team's passing network graph in a game
N_{team}	The total number of players on the team who have played in a game
N_{class}	Pass Types
M_2, M_3	The total number of times a team has completed a 2 and 3 team match
R_i	The total number of passes received by the i player in a game
O_i	Total number of j-th passes made by the i-th player in a game
R_{ij}	Total number of j-th passes received by the i-th player in a game
O_{ij}	Total number of j-th passes made by the i-th player in a game
T_i	Cumulative possession time of the i-th player in a game (unit: minute)
G_i	The i-th player 's average attractive to other players
w_j	Weight of the j-th outgoing pass type
h_j	Weight of pass type received by j
Co_j	Weight of the j-th outgoing pass type
Ci_j	Weight of pass type received by j
e_j	Number of passes for the j-th edge of the passing network in a game
$(I_j\ ,I_j)$	The starting coordinate of the j-th edge of the passing network in a game
$(O_j\ ,O_j)$	The coordinates of the end point of the j-th edge of the passing network in a game
$(x_i\ ,\ y_i)$	The average position of the i-th player in a game
$(x_c\ ,y_c)$	The centroid coordinates of the team's passing network in a match
$d(x_1\ ,y_1\ ,x_2\ ,y_2)$	Euclidean distance from coordinate (x1, y1) to coordinate (x2, y2)

3 Model design

3.1 Passing Network Model[1]

We use a directed node graph in network science to build a passing network.

First, the player is regarded as a node in the network graph, and the individual attraction of the player is defined according to the gravitational formula in physics, which is used as a basis for measuring the size of the node. The pass between players is regarded as an edge, and the edge is given a certain weight according to the number of passes. Finally, the average coordinate of the player's entire game is used as the coordinate in the node graph.

For the team's overall passing ability, judge the quality of the team by the total number of passes. The number of passes N of the entire team has been given in the attachment. This indicator evaluates a team's passing ability from a macro perspective. From the perspective of the individual player, the number of passes between the player and the player is counted as the edge weight in the passing network, and the gravity model is used to describe the importance of the player, and it is also the size of the passing network node Important reference.

For the player's personal attractiveness, it can be intuitively understood as the player's personal ability on the field. We use a game as a unit to analyze the personal attractiveness of each player who played in the game. Firstly, through the attachment, I directly obtained the number of passes received by the i-th player of the Huskies team in this game as R_i, the number of passes passed as O_i and the cumulative effective ball possession time T_i. The number of received and outgoing passes $R_i$and O_i is large, and the ball control time $T_i$is long, so it can certainly be attractive. Because there are different types of passes, the total number of passes is composed of different types of passes:

$$R_i = \sum_j^{N_{class}} r_{ij} \tag{1}$$

$$O_i = \sum_j^{N_{class}} o_{ij} \tag{2}$$

N_{class}represents the total number of passing types, i represents the i-th player, and j represents the j-th pass. For different types of passing balls, the difficulty is different. Naturally, the degree of influence on the attraction is different. Therefore, for different types of passing balls o_{ij} , p_{ij}should be given different weights:

$$weight(o_{ij}) = w_j$$

$$weight(p_{ij}) = h_j$$

Due to the calculation of weight, if only a certain game is random, in order to reduce this randomness, we count the various types of passes throughout the season:

Table 2: Huskies' pass out and catch statistics this season

Type of passing	Symbol	Number of passing	Symbol	Number of catching
Head pass	Co_1	572	Ci_1	571
Simple pass	Co_2	8735	Ci_2	8730
Launch	Co_3	212	Ci_3	211
High pass	Co_4	586	Ci_4	583
Hand pass	Co_5	127	Ci_5	127
Smart pass	Co_6	73	Ci_6	73
Cross	Co_7	130	Ci_7	130

First calculate the frequency of various types of passing balls in the game, and then take the countdown and normalize to get:

$$\mathrm{w_j} = \frac{\frac{1}{\mathrm{Co_j}}}{\sum_{\mathrm{j}}^{\mathrm{N_{class}}} \frac{1}{\mathrm{Co_j}}} \tag{3}$$

$$\mathrm{h_j} = \frac{\frac{1}{\mathrm{Ci_j}}}{\sum_{\mathrm{j}}^{\mathrm{N_{class}}} \frac{1}{\mathrm{Ci_j}}} \tag{4}$$

The final calculated weight is:

Table 3: Huskies team passing weight statistics

Type of passing	Received symbol	weight	Passed symbol	weight
Head pass	h_1	0.047	w_1	0.047
Simple pass	h_2	0.003	w_2	0.003
Launch	h_3	0.126	w_3	0.126
High pass	h_4	0.046	w_4	0.045
Hand pass	h_5	0.209	w_5	0.210
Smart pass	h_6	0.364	w_6	0.365
Cross	h_7	0.205	w_7	0.205

Through weights, we can infer that this is a reasonable allocation through experience combined with the statistics of various types of passes previously counted. First, observe the distribution chart of the number of passes:

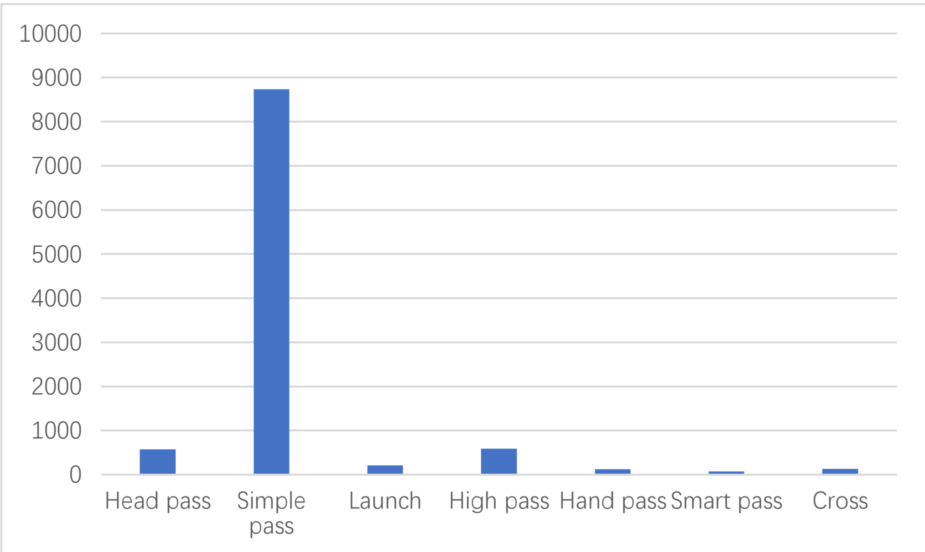


Fig. 1: Statistics of the number of passes made by Huskies during the season

For example, Smart pass occurs relatively rarely. We can think that Smart pass requires players to have strong personal ability. Both the receiver and the passer have a high degree of ability. Therefore, such a ball is more correct. Give higher weight, and Simple pass is the most common pass. This type of pass obviously does not require players to have strong personal ability, so naturally give lower weight.

According to the above analysis, in order to comprehensively consider the above-mentioned player's holding time T_i the number of passes O_i, R_i and the number of passing passes divided by the type of passing o_{ij}, r_{ij}, the player's attractiveness is defined as:

$$G_i = T_i + w_j \cdot o_{ij} + h_j \cdot p_{ij} \qquad (5)$$

Then the player's position is obtained by averaging (x , y) when receiving the ball throughout the game. Here, (x, y) is explained. According to the description of the establishment of the football stadium in the annex and based on it We divide the football field into 100 equal parts, that is, the value range of x and y is: [0,100].

Taking each player as a node, the position of the node in the picture is (x_i,y_i), and we use the size of the attraction as the radius of the node, and then count the passes between players in each game , Draw the connection between the nodes, that is, the edge of the passing network, the color depth of the edge is linearly related to the number of passes between the nodes or the players, and the color depth of the edge also reflects the edge with weights. As a result, a picture of the passing network was obtained:

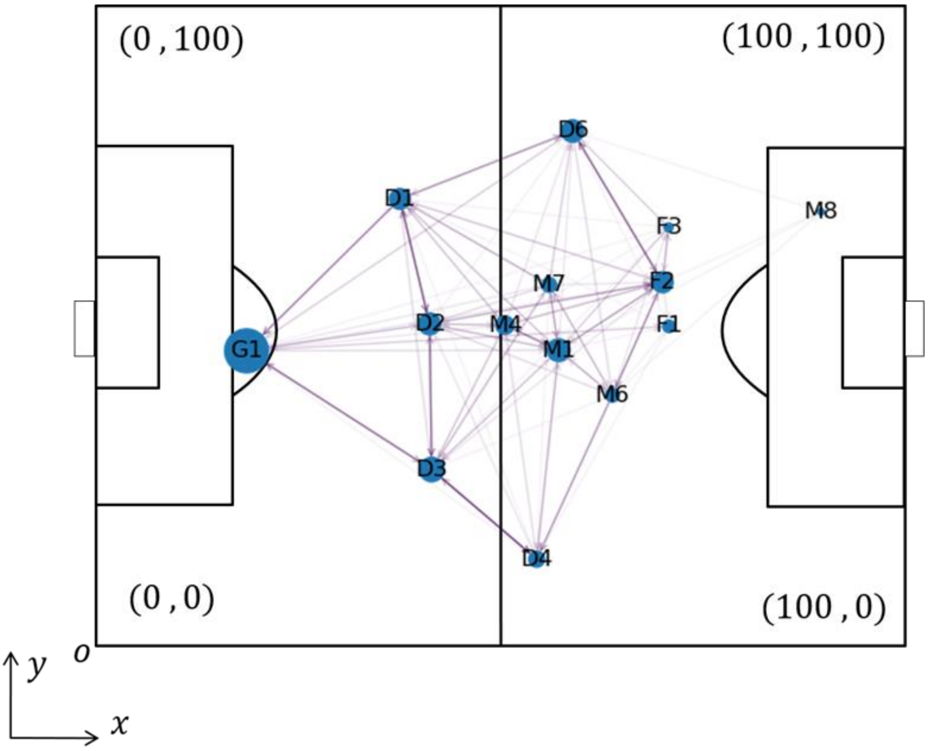


Fig.2: Passing Network of Huskies in Game 3

Through the statistics of all teams throughout the season, we obtained a total of 38 Huskies' passing networks and 38 opponents passing networks. The passing network diagram can intuitively see the interaction between players and the importance of a player in the entire game, such as the comprehensive game results and goals.

3.2 Teamwork Evaluation Model

Football is a multiplayer sports event, and even top players cannot win alone without teamwork. As a result, indicators that measure team capabilities have received much attention.

Teamwork capacity is a very complicated concept, and it is a very complicated thing to evaluate it directly, because it is affected by many factors, each factor is often not independent, and mixed with a large number of subjective judgments. Therefore we decided to adopt a fuzzy comprehensive evaluation method[2] to evaluate the concept of teamwork. Fuzzy comprehensive evaluation method is a comprehensive evaluation method based on fuzzy mathematics. This comprehensive evaluation method converts qualitative evaluation to quantitative evaluation according to the membership theory of fuzzy mathematics, that is, uses fuzzy mathematics to make an overall evaluation of things or objects subject to multiple factors.

It has the characteristics of clear results and strong systemicity. It can solve fuzzy and difficult to quantify problems, and is suitable for solving various non-deterministic problems.

The general steps of the fuzzy comprehensive evaluation method are:

[1] Determine the factor set, which is the evaluation index

[2] Determine the comment set

[3] Determine the weight of each factor

[4] Determine the fuzzy comprehensive evaluation matrix and evaluate the factors

First of all, the teamwork capacity is used as the final evaluation purpose and as a set of reviews. The eight original indicators: the number of passes, the success rate of passes, **the number of shots, the number of goals, the goal rate, the rate of possession,** and **the winning and losing rate** are used as the set of determining factors. After determining the weight of each factor, we obtained a fuzzy comprehensive evaluation matrix to reflect the teamwork capacity.

On this basis, in combination with the passing network established by the previous model, we also proposed five indicators as two pole indicators of the eight original indicators: **the passing network centroid, the passing network node dispersion, the passing weighted average path, and cluster Coefficient, a binary clustering coefficient,** is a new dimension to measure teamwork capabilities.

We simply named the Huskies team Husk. Since there is only data on the Huskies team playing against other teams, it does not make sense to model each specific opponent separately.

So we combined all the opposing teams to get the average. For the analysis, we combined the opposing teams into a team. We named Opponent x (x = 1 ~ 19), that is, the other 19 teams that played against Huskies, Oppo.

(1) Original indicator

There are eight original evaluation indicators, which are the winning and losing situation, the number of goals, the number of passes, the number of shots, the goal rate, the percentage of possession time, and the pass success rate. Among them, the winning and losing situation, the number of goals, the number of passes, and the number of shots are often used as the classic performance indicators of reaction teams [1]. These four indicators can be obtained by filtering directly from the attachment.

The goal rate can be used as a basis for measuring whether the team's offense is effective.

The goal rate R is defined as:

$$R = Total\ points\ /\ Total\ shots$$

The percentage of possession time is calculated by counting the time occupied by both sides in a game.

In football games, the ball transfer often occurs. Obviously, the team cooperates with the team with a high degree of understanding and will also show a high success rate when passing the ball, thereby assisting the team's offense and scoring. Therefore, a team will tend to maintain a high pass success rate during the game.

Pass success rate P is defined as:

$$P = \frac{\text{(Passed by the own received by the own)}}{All\ passed}$$

Count the winning and losing situation of the Husk team and all the teams of Oppo in 38 games, that is, the situation of wins, draws and losses, the number of goals and shots, and the goal rate. Due to the comparison of the influence of home and away on these indicators.

Large, so these indicators are divided into home and away to discuss and analyze:

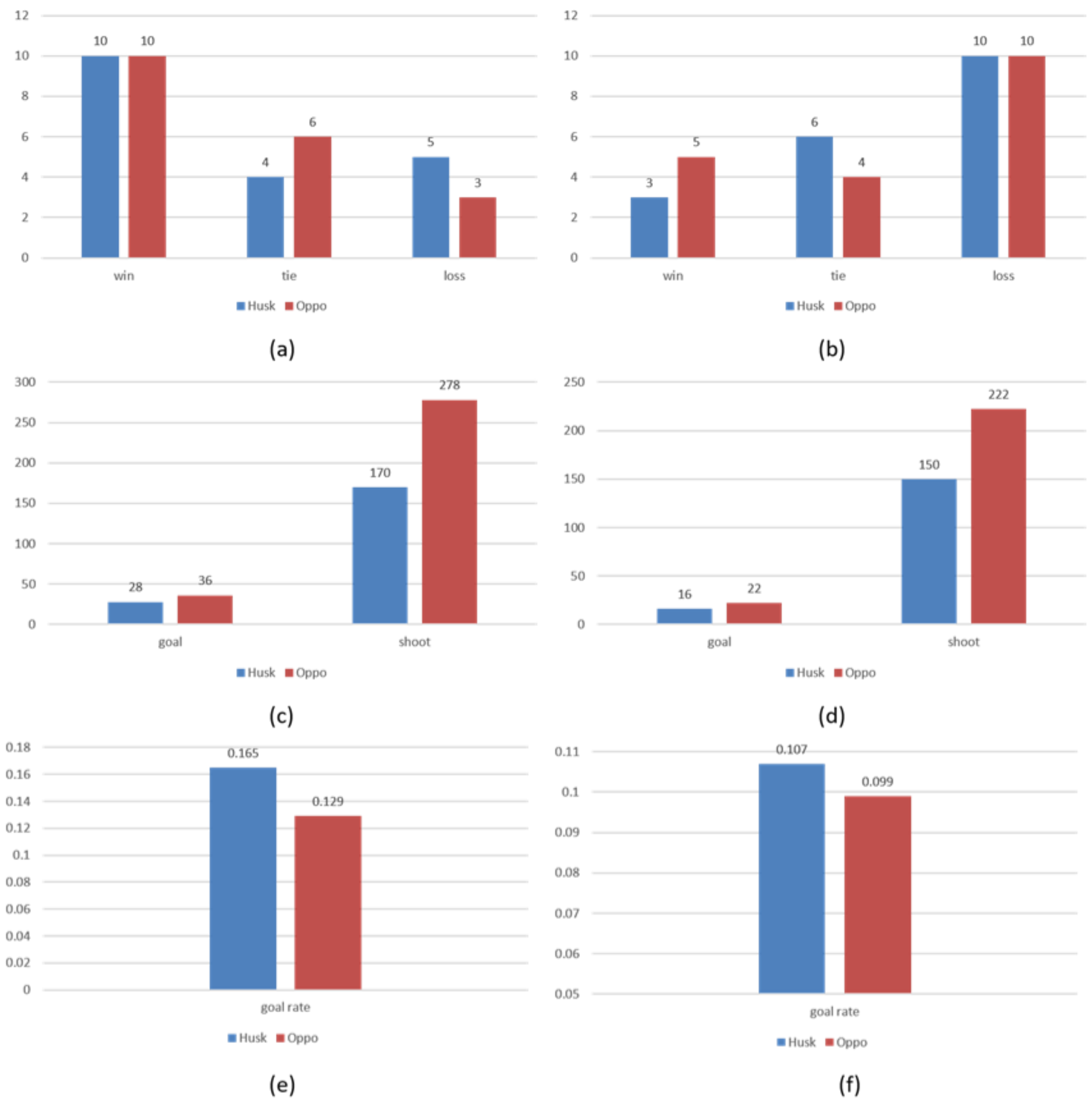
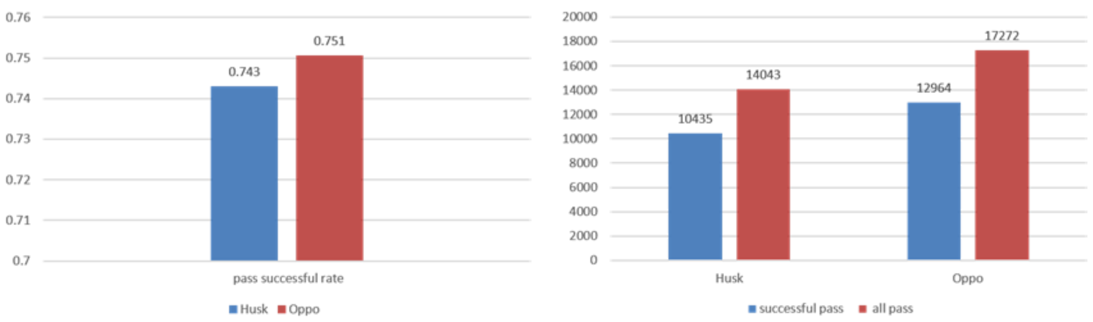


Fig. 3: (a) Wins and losses of Husk and Oppo at home; (b) Wins and losses of Husk and Oppo at home; (c) Husk and Oppo's shots and goals at home; (d) Husk and Oppo's shots and passes athome; (e) Husk and Oppo's goals at home; (f) Away Husk vs. Oppo's goal rate at the time;

Then we analyze from the perspective of the number of passes, the success rate of passing, and the rate of possession. At this time, the impact of the home and away on the indicator is small, so you can ignore the home and away at this time:



(a) (b)

Fig.4: (a) Comparison of pass success rate between Husk team and Oppo team; (b) Successful pass and total number of passes between Husk team and Oppo team;

(a) (b)

Fig. 5: (a) Comparison of ball control time between Husk team and Oppo; (b) Comparison of ball control rate between Husk team and Oppo

From this, we obtained seven original indicators, and obviously found that the data scale span was relatively large. In order to make it a factor that determines the concentration of factors, it is relatively normalized. The results are as follows:

Table 4: Set of normalized determinants

	Win	**Shot**	**Goal**	**Goal rate**	**Pass Success Rate**	**Passes**	**Possession time**	**Possession**
Husk	1	1	1	1	1	1	1	1
Home	1	1.28	1.63	0.78	1.01	1.22	1.06	1.08
Away	1.66	1.38	1.48	0.92	1.01	1.22	1.06	1.08

Winning and losing situations are uniformly evaluated by winning games. Since the influence weight of each factor is not specified, according to the usual processing method of analytic hierarchy process, the weight is uniformly set to 1.

Finally, the fuzzy evaluation results of Husk and Oppo were obtained:

Table 5: Fuzzy evaluation results

	Teamwork capacity
Husk	1
Home	1.13
Away	1.22

The results show that the teamwork ability of Huskies, whether at home or away, is less than the teamwork ability of the opponent.

(2) The second indicators

First, the centroid of the passing network, the dispersion of the nodes of the passing network, the weighted average path of the passing, the clustering coefficient, and the binary clustering coefficient are defined respectively. Then calculate them separately, and finally make a qualitative analysis on the results of correlation analysis.

In order to judge the management level of the second-level evaluation index and the original index, a correlation analysis is attached after the calculation results of each index. We have introduced a method of correlation test, that is, calculating the correlation coefficient[3] and correlation The formula for the coefficient is given as follows:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2} \tag{6}$$

The x and y appearing here are not the meaning of the coordinates, but refer to the two groups of samples that need to be analyzed for correlation, where x¯ and y¯ are the average values of the samples, and the value obtained from the correlation coefficient must be -1 ~ .

Between 1, the closer the absolute value of the correlation coefficient is to 1, the stronger the correlation between the two groups of samples, and the closer the absolute value is to 0, the weaker the correlation between the two groups of samples.

The value of the correlation coefficient and the degree of correlation are explained here:

When | r | ≤0.3, it can be considered that there is no linear correlation between the two samples.

When 0.3 < | r | ≤ 0.5, it can be considered that there is a weak linear correlation between the two groups of samples.

When 0.5 < | r | ≤ 0.8, it can be considered that there is a significant linear correlation between the two groups of samples.

When 0.8 < | r | ≤ 1, it can be considered that the two samples are highly linearly related.

The value of r can reflect the reliability of the second indicators.

Passing Network Centroid

In the passing network, the player's position is directly obtained by taking the arithmetic average of the distance when receiving the ball. From this, it can be guessed that the player's receiving position in the entire game can reflect the team's performance in that game, so We define a parameter called the network centroid, which is used to measure the relative position of the entire team on the court.

Formula for calculating the center of mass in analog physics:

$$r_\delta = \frac{\sum_i^n m_i r_i}{M} \tag{7}$$

In the formula, m_i is the mass of the M-th object, and M is the total mass of all nodes considered. Replace the concept of mass with player attraction:

$$G = \sum_{i=1}^{N_{net}} G_i \tag{8}$$

$$x_c = \frac{\sum_{i=1}^{N_{net}} x_i \cdot G_i}{G} \tag{9}$$

$$y_c = \frac{\sum_{i=1}^{N_{net}} y_i \cdot G_i}{G} \tag{10}$$

In the formula, $N_{team}$is the number of players in the game, and G is the sum of player appeal.

Considering that the y coordinate indicates the left and right hand on the field, this has no direct impact on the analysis of the offensive and defensive trends of the two teams, so it is not considered here. According to the definition of the x coordinate, the closer the coordinate is to 0, the closer it is to the goal of the own side, and the closer the coordinate is to 100, the closer is the goal of the opponent. The larger the x_c coordinate of the center of mass, it can reflect the team's tendency to appear in the opponent's half to a certain extent, which indicates that the team tends to be more aggressive and this should explain the team's offensive organization ability Offensive organizational ability is of course part of teamwork capacity.

Based on this, we calculated the average network centroid x_c coordinates of the Husk team and Oppo team in 38 games throughout the season:

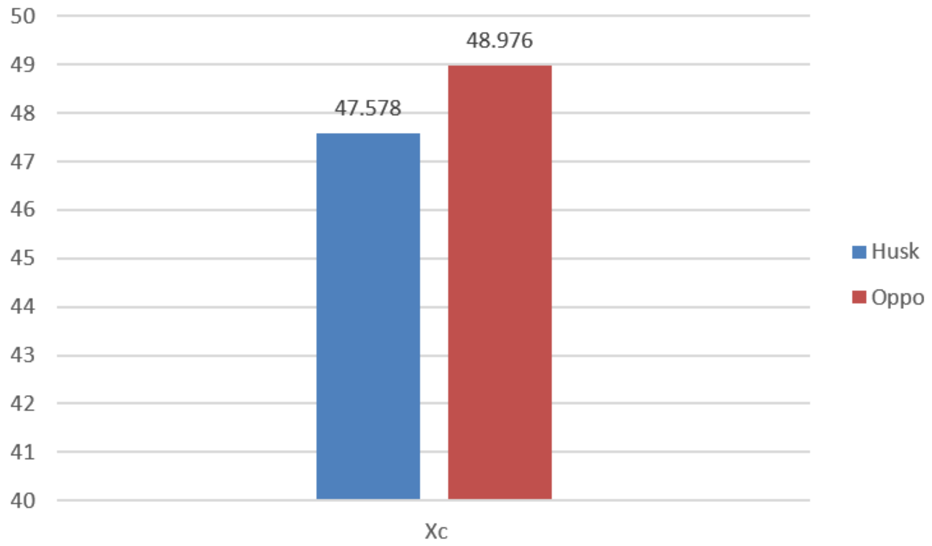


Fig.6: $x_c$axis coordinate average of network centroid

Intuitively, the coordinates of the center of massx_c of the Husk team are smaller than the center of mass x_c of the Oppo team. It is defined by the network centroid Shots.

By calculating the correlation coefficient between the center of mass and the number of shots of the passing network, the calculation results are r_{Husk} = 0.423 and r_{Oppo} = 0.378. According to the rules of correlation analysis introduced earlier, it can be considered that thex_c coordinate of the center of mass has a weak correlation directly with the number of shots. Then it can reflect the offensive organization ability to a certain extent, thereby further reflecting the team cooperation ability.

Weighted average node distance

In the process of passing, the passing distance between two players can affect the team's range of activities on the field. Long-distance passing is conducive to the organization and movement of the attack. Therefore, we believe that there needs to be an indicator to measure the overall passing distance of the passing network, and we hope to reflect the team's offensive organization ability through this distance. The calculation of the weighted average node distance is based on the construction method of the passing network, so it will not be repeated here, and then consider a certain game, after obtaining the passing network of the game, we can get the coordinates of each node , You can get the weight of each side, that is, the number of passes between each node divided by the total number of passes, and then traverse all sides, or take the weight and the length of the side to perform a weighted sum:

$$weighted\ average\ length_j = \sum_{k=1}^{N_e} \frac{e_j}{\sum_{i=1}^{N_e} e_i} \cdot\ d(I_j\ , I_j\ , O_j\ , O_j) \tag{11}$$

We counted 38 games and averaged them to get the data and draw the graph as follows:

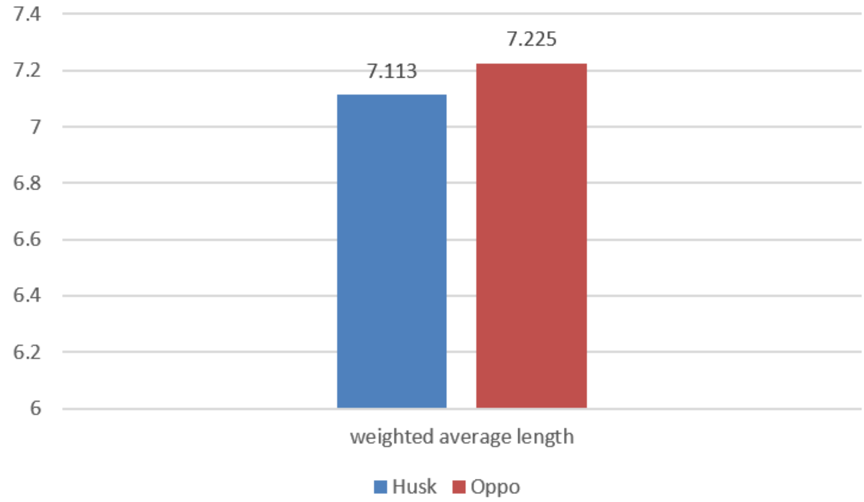


Fig.7: Weighted average node distance of both sides

Intuitively, there is a certain difference between the weighted average node distance of Husk and Oppo teams, which is similar to the comparison of the number of shots of the second team, so it can be considered that it has a contribution to the number of shots.

Clustering coefficient[4]

In a network, the connection between vertices and vertices can easily occur from vertex 1 to vertex 2 and vertex 3, and whether vertex 3 is connected to vertex 1 constitutes closed triples.

And open triples, which is a topic often discussed in topology. By capturing closed triples and open triples, computing the Clustering coefficient (that is, the clustering coefficient) can well capture the connection between network nodes. Tightness, which is defined as:

$$C = \frac{3n_\Delta}{3n_\Delta + n_\Lambda} \tag{12}$$

Where $n_\Delta{}^i$ is the number of closed triples in the network, and $n_\Lambda{}^i$is the number of open triples in the network, and here we use another definition for easier calculation, proposed by Watts[5], for the passing network.

A team member i has a local clustering coefficient C_i:

$$C = \sum_{i=1}^{N_{net}} C_i \tag{13}$$

We calculated the Clustering coefficient of the Husk team and the Clustering coefficient of the Oppo team in 38 games:

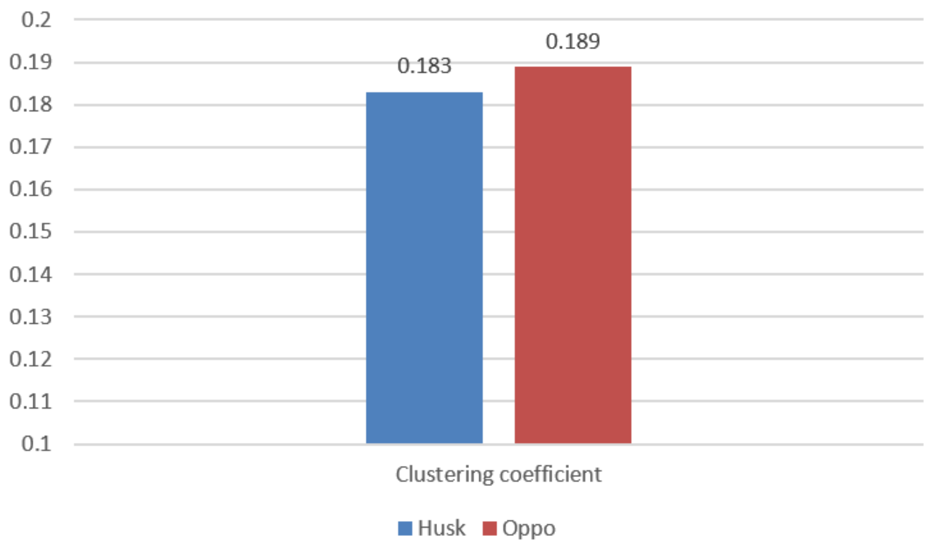


Fig. 8: Clustering coefficient of both sides

From the results, starting from the passing network itself, the Oppo team has a higher Clustering coefficient than the Husk team. From a topology perspective, the Oppo team's passing network is between the nodes of the Husk team's passing network. The connection is closer, and then we calculate its correlation coefficients with the number of passes and shots and the rate of possession:

Table 6: relative coefficient between Clustering coefficient and pass count, shot count, possession rate

	Husk	**Oppo**
r_1	0.437	0.351
r_2	0.314	0.289
r_3	0.413	0.387

Among them, r_1 refers to the correlation coefficient between the Clustering coefficient and the number of passes, r_2 refers to the correlation coefficient between the Clustering coefficient and the number of shots, and r_3 refers to the correlation coefficient between the Clustering coefficient and the possession rate.

Comprehensive r_1、r_2、r_3 can be analyzed to find that the Clustering coefficient has a weak correlation with the ball control rate. From an intuitive point of view, the closer the clustering coefficient represented by the larger Clustering coefficient, the closer the player.

In this case, the player is less likely to make a mistake, and the possession rate should be higher.

Pairwise Clustering coefficient

In order to further evaluate the dynamic performance of the team, we have expanded the Clustering coefficient to analyze from the perspective of pair. Similarly, we connect node 1 to node 2 and node 2 to node 1, which is defined as closed second. Tuples, while node 1 is connected to node 2 and node 2 is connected to other nodes. It is defined as an open two-tuple. The Pairwise Clustering coefficient Cp is calculated by capturing closed and open two-tuples in the network.

$$Cp = \frac{2n_=}{2n_= + n_\neq} \tag{14}$$

Where $n_=$ is the number of closed doubles in the network, $n_\neq$ is the number of open doubles in the network, and in order to simplify the calculation, we also use the concept of local pairwise clustering coefficient.:

$$Cp_i = \frac{2n_=^{\ i}}{2n_=^{\ i} + n_\neq^{\ i}} \tag{15}$$

Where $n_=^{\ i}$ is the number of closed doubles starting from the i-th player in the passing network, and $n_\neq^{\ i}$ is the number of opening doubles from the i-th player in the passing network.

Then average all Cp_i to get the clustering coefficient of the entire network:

$$Cp = \sum_{i=1}^{N_{net}} Cp_i \tag{16}$$

We calculated the Pairwise Clustering coefficient of the Husk team and the Pairwise Clustering coefficient of the Oppo team in 38 games:

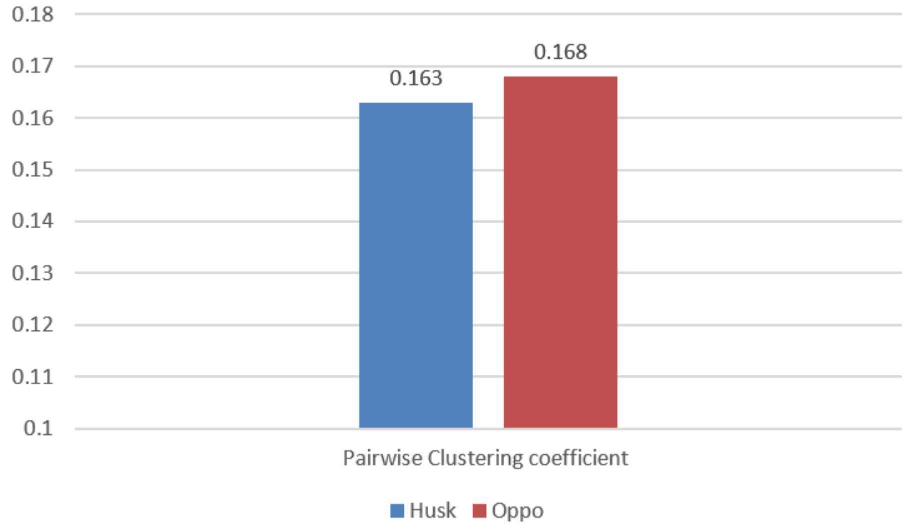


Fig. 9: Pairwise Clustering coefficient of both sides

From the perspective of the results, from the perspective of the passing network, or from the perspective of topology, similar to the triples, the binary can also reflect the connection between the nodes. From a point of view, it is a successful interaction, and interaction is a way to connect nodes more closely. So in this way, Pairwise Clustering coefficient can measure the degree of connection between nodes in the entire network. It should be able to reflect the team's ability to cooperate, and then we use correlation analysis to calculate the correlation coefficient with the number of passes, shots, and ball control rate:

Table 7: relative coefficient between Pairwise Clustering coefficient and pass count, shot count, possession rate

	Husk	**Oppo**
r_1	0.476	0.519
r_2	0.174	0.211
r_3	0.395	0.362

r_1 refers to the correlation coefficient between Pairwise Clustering coefficient and the number of passes, r_2 refers to the correlation coefficient between Pairwise Clustering coefficient and the number of shots, and r_3 refers to the correlation coefficient between Pairwise Clustering coefficient and the rate of possession.

Comprehensive r_1、r_2、r_3 can be analyzed to find that the Pairwise Clustering coefficient has a weak correlation with the ball control rate. It can be considered that it has a contribution to the ball control rate. It is observed that the correlation coefficient decreases compared to the Clustering coefficient. Because the pass between the two is easy to target, after all, the two-man offense is easier to defend than the three-man offense. From this perspective, it should be reasonable. Overall, the Pairwise Clustering coefficient has a certain teamwork capacity contribution.

4 Strategies based on Models

In a game, both the opponent and the opponent often have specific passing motifs.

The special equipment is football. The interaction between players is considered a complex network. By extracting the passing network, the network can be further subdivided into sub-graphs, which appear more frequently in the network than in random networks.

Higher[6], If we want to disrupt the opponent's game format, we must first be aware of the pass motifs that the team often uses,[7] In the pass network we have built, Position-dependent passing motifs to identify common fits in football games. Simplify the network by identifying commonly-used fits. Use the following algorithm to traverse the passing network:

① **find the ten edges with the greatest weight.**

② **find the nodes adjacent to the edge and delete the nodes that have no edges adjacent to it.**

Get the simplified passing network:

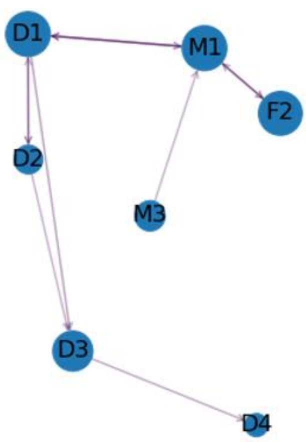


Fig. 10: Simplified passing network

In this figure, the lines that appear are the closest match for this game. The statistics of the number of appearances of the players in each position in the simplified passing network and the eight players with the most appearances were obtained:

Table 8: Appearance statistics of Husk team

Position	G	D	M	F
Time	19	140	99	50

Player	**time**
M1	31
F2	27
D1	24
D3	21
D5	21
M6	20
D2	19
D4	19

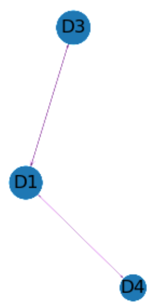
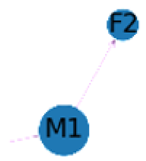
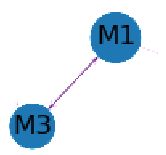
Subsequently, the number of fits is further filtered based on the data on the premise that assumptions ③ and ④ are satisfied. The screening method is:

For two-person cooperation, it is considered that after passing the ball from one player to another, passing the ball back is regarded as completing a two-person cooperation.

As for the three-person match, it is considered that the ball is passed from one player to the second player, and finally reaches the third player as a complete three-person match.

In simplifying the passing network, we again consider the coordination situation, that is, the connection between the two nodes, and counted out the most commonly used cooperation of the Huskies team.

Table 9: different type of teamwork statistics

Type of matching	**time**
	11
	15
	11

Combining the winning and losing situation of the game, we analyzed the Huskies team's game strategy:

① **From the perspective of the players: The players in the team have different influences on the court. In the midfield, Huskies likes to take advantage of the M1 player's ball control advantage, which is reflected in many games where he is the central node of cooperation.**

In the frontcourt position, the player F2 is often the main offensive firepower. He often receives passes from the Quartet teammates on the rear line. Huskies likes to use D1, D3, D5, and D4 players to create a tight rear line.

② **From the perspective of team cooperation: In the 38 games, the most commonly used cooperation of the Huskies team is the above three types, that is, to control the game through a large number of midfield ball possessions, and to cooperate with forward F2 to organize attacks when possible. Or pass the ball back to the guard and wait for it.**

At the same time, it was found that in the game won by Huskies, the team's midfield and frontcourt are often far away from the network centroid, which means that they can give full play to the advantage of controlling the ball and approach the ball to the other half.

In Huskies' tie or loss games, the rear line often bears a lot of pressure, which is reflected in the fact that most of the passes appear on the rear.

At this time, the distance between the midfield is relatively short, and in the opponent's passing network, there is often a certain attractive forward, which can put a lot of pressure on our back defense and lead to losing the game.

For example, in match number 3, compare the passing networks of Huskies and their opponents:

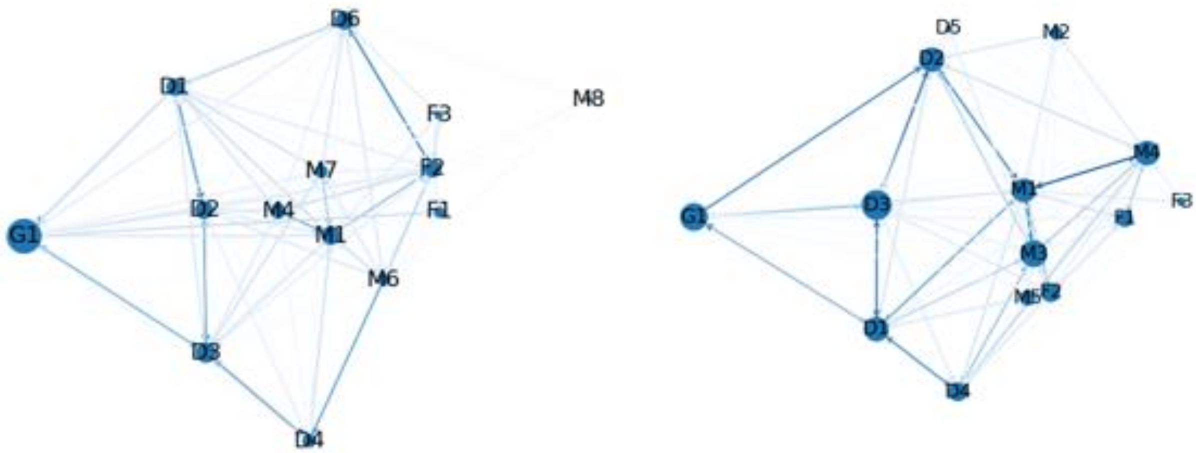


Fig. 11: In the third match, Huskies' passing network (left), opponent's passing network (right)

In this game, Huskies lost to their opponents. As mentioned in the analysis above, Huskies' defense line often bears a lot of pressure when heading against the wind, which is reflected in the passing network as dark lines basically appear in the backcourt. There is no great attraction. The cooperation of M1-M4 makes the dispersion of the entire team network large, and the distance of M4 is very close to our backcourt. As a result, Huskies has lost the initiative.

Based on the analysis of the cooperation between the passing network and the players, make recommendations for the Huskies team next season:

① **Strengthen the control of the midfield to take the initiative. It is learned from the simplified passing network analysis that Huskies has a large force on the rear defense line, which has given up control of the midfield to some extent.**

② **When the opponent has a prominent front peak, the organization of a more effective midfield and frontcourt cooperation can effectively improve Huskies' win rate. On the contrary, the performance of giving the key forces to the backcourt is not good.**

③ **Pay attention to the three long-standing cooperation mentioned above. Team training for these three cooperation may help effectively improve the team's coordination ability, but it also requires whether the main opponent already has the ability to target these coordination.**

From the perspective of teamwork indicators, Huskies has lower than average opponents to a certain extent in both the original and secondary indicators, which means that the coaching staff should strengthen the player's basic skills training to improve the overall team strength.

In the subsequent analysis, if it can provide player-based information such as **physical fitness, mental state, and skill level**, it will further help the establishment of reliable models.

5 Strengths and Weaknesses

- **Strength:**

① The teamwork capacity model has many evaluation indicators. It analyzes the data from multiple dimensions, and finally provides correlation analysis, which is reasonable enough.

② Using the pass network model, the team lineup is displayed intuitively, which is convenient for the coaching team to analyze specific matches.

- **Weakness:**

① Simplification takes into account many events that can occur on the football field, such as fouls, corner kicks, etc. The model established does not consider the rules of complete football.

② Failure to analyze the individual abilities of the players, which led to some players' personal factors not being captured.

6 Conclusions

According to our model, the thinking of a node network to measure an effective team should have network configuration and node weights. In simple terms, an excellent team depends on whether there are excellent team members and a team structure that can bring out the characteristics of members.

For football matches, there are still many aspects worth digging. Such as many factors based on the player's own (psychological state, competitive level) and factors based on objective conditions (weather, wind direction). Combining the above data, the mathematical model provided will be more accurate and reliable.

References

[1] Buldu J M, Busquets J, Echegoyen I. Defining a historic football team: Using Network Science to analyze Guardiola's FC Barcelona[J]. Scientific reports, 2019, 9(1): 1-14.

[2] Cheng J, Tao J. Fuzzy comprehensive evaluation of drought vulnerability based on the analytic hierarchy process:—an empirical study from Xiaogan City in Hubei Province[J]. Agriculture and Agricultural Science Procedia, 2010, 1: 126-135.

[3] Taylor R. Interpretation of the correlation coefficient: a basic review[J]. Journal of diagnostic medical sonography, 1990, 6(1): 35-39.

[4] Saramaki J, Kivelä M, Onnela J P, et al. Generalizations of the clustering coefficient to weighted complex networks[J]. Physical Review E, 2007, 75(2): 027105.

[5] Watts D J, Strogatz S H. Collective dynamics of 'small-world'networks[J]. nature, 1998, 393(6684): 440.

[6] Itzkovitz S, Alon U. Subgraphs and network motifs in geometric networks[J]. Physical Review E, 2005, 71(2): 026117.

[7] Zhang L L, Thomas W, Ashiabar A. Comparing predictive powers of Network Motif Distribution and structure of Overlapping Communities[J].

Appendix

The material url is https://github.com/sennnnn/competetion, All of our data analysis and code has uploaded.

Because the code is so long that we can't put all of them in the appendix, so we just put part of the code.

Function: cooperation_valid_detect:

Usage: to detect if two events compose a cooperation.

```
 def cooperation_detect(info_1, info_2):
    """
    detect if info_1 and info_2 construct a cooperation.
    Args:
        info_1, info_2:info item
    Return:
        players:cooperation player name list
        flag:if constructing a cooperation
    """
    players = []
    time_accu = 0
    Origin_1 = info_1['OriginPlayerID']
    Origin_2 = info_2['OriginPlayerID']
    Dest_1 = info_1['DestinationPlayerID']
    Dest_2 = info_2['DestinationPlayerID']
    time_stamp_1 = float(info_1['EventTime'])
    time_stamp_2 = float(info_2['EventTime'])
    # judge between the first pass and the second pass
    if((time_stamp_2-time_stamp_1) > 5 \
        or (time_stamp_2-time_stamp_1) < 0):
        return players, False
    else:
        if(Origin_1[:4] != Origin_2[:4]):
            return players, False
        else:
            if(Dest_1 != Origin_2 or \
               Origin_1 == Dest_1 or \
               Origin_2 == Dest_2):
                return players, False
            else:
                players.append(Origin_1)
                players.append(Dest_1)
                players.append(Origin_2)
```

```
            players.append(Dest_2)
            temp = list(set(players))
            players = sorted(temp, key=players.index)
            return players, True
```

Function: catch_ball_time_calculate

Usage: to calculate the ball-catching time.

```
def catch_ball_time_calculate(info_1, info_2, member):
    """
    count single catch ball time.
    Args:
       info_1,info_2:info item
       member:player name
    Return:
       bool:fail to catch ball
       time_period:catch ball time.
    """
    Origin_1 = info_1['OriginPlayerID']
    Origin_2 = info_2['OriginPlayerID']
    Dest_1 = info_1['DestinationPlayerID']
    Dest_2 = info_2['DestinationPlayerID']
    time_stamp_1 = float(info_1['EventTime'])
    time_stamp_2 = float(info_2['EventTime'])
    if(Dest_1 != member or Origin_2 != member):
       return False,0

    return True,time_stamp_2-time_stamp_1
```

Code block

Usage: to calculate centroid

```
x_y = [members_info[member]['coordinate'] for member in members_ID]
weight = [members_info[member]['attrc'] for member in members_ID]
M = sum(weight)
mixi = sum([xy[0]*m for xy,m in zip(x_y, weight)])
miyi = sum([xy[1]*m for xy,m in zip(x_y, weight)])
weight_point = (mixi/M*100, miyi/M*100)
f_.write('{} {} {}\n'.format(MatchID, *weight_point))
```

Code block

Usage: to calculate desperation

```
node_distance_list = []
for i in range(len(members_ID)):
    x = members_info[members_ID[i]]['coordinate'][0]*100
    y = members_info[members_ID[i]]['coordinate'][1]*100
    temp = distance((x,y), weight_point)
    node_distance_list.append(temp)
avg = average(node_distance_list)
S = [(x - avg)**2 for x in node_distance_list]
S = math.sqrt(sum(S)/len(node_distance_list))
```

Function: attractive_force_calculate

Usage: to calculate weight of all edges.

```
def attractive_force_info_extract(file_object, MatchID, \
    info_dict, catch_kind_weight, pass_kind_weight):
    line = file_object.readline()
    while(1):
        line = file_object.readline()
        line = line.strip()
        if(line == '[%d]'%(MatchID)):
            line = file_object.readline().strip()
            length = int(line)
            for i in range(length):
                line = file_object.readline().strip()
                infos = line.split(' ')
                catch_time = float(infos[1])
                catch_count = int(infos[2])
                pass_count = int(infos[3])
                catch_all_type_list = \
                [int(x)*weight  for  x,weight  in  zip(infos[4:11],
catch_kind_weight)]
                pass_all_type_list = \
                [int(x)*weight  for  x,weight  in  zip(infos[11:],
pass_kind_weight)]
                info_dict[infos[0]]['attrc'] = \
                catch_time/60  +  sum(catch_all_type_list)  +
sum(pass_all_type_list)
            return info_dict
```

Code block

Usage: to calculate weighted average length

```
edge_weight = [y[0] for x,y in edge_dict.items()]
edge_point = [(y[1][0],y[1][1]) for x,y in edge_dict.items()]
sum_weight = sum(edge_weight)
weighted_average_length = []
for w,p in zip(edge_weight, edge_point):
    member_1 = members_ID[p[0]]
    member_2 = members_ID[p[1]]
    point_1 = members_info[member_1]['coordinate']
    point_2 = members_info[member_2]['coordinate']
    dis_point_1_2 = distance(point_1, point_2)
    weighted_average_length.append(2/sum_weight*dis_point_1_2)
     weighted_average_length = sum(weighted_average_length)
```

Function: clustering_coef_cal

Usage: to calculate clustering coefficient

```
     def clustering_coff_cal(member_index_list, edge_list):
    avg_c_f = 0
    for member_index in member_index_list:
        open_single = 0
        close_single = 0
        first_connect_node = []
        for first_edge in edge_list:
            if(first_edge[0] == member_index):
                first_connect_node.append(first_edge[1])
        for first in first_connect_node:
            second_connect_node = []
            for second_edge in edge_list:
                if(second_edge[0] == first):
                    second_connect_node.append(second_edge[1])
            for second in second_connect_node:
                for third_edge in edge_list:
                    if(third_edge[0] == second):
                        if(third_edge[1] == member_index):
                            close_single += 1
                        else:
                            open_single += 1
        avg_c_f += 3*close_single/(3*close_single + open_single)
    return avg_c_f/len(member_index_list)
```

Function: r_coeff

Usage: to calculate relative coefficient

```
def r_coeff(x, y):
    avg_x = average(x)
    avg_y = average(y)
    under_x = 0
    under_y = 0
    above = 0
    for i,j in zip(x,y):
        above += (i - avg_x) * (j - avg_y)
        under_x += (i - avg_x)**2
        under_y += (j - avg_y)**2

    ret = above/(math.sqrt(under_x) * math.sqrt(under_y))
    return ret
```

Function: pair_clustering_coff_cal

Usage: to calculate pairwise clustering coefficient

```
def pair_clustering_coff_cal(member_index_list, edge_list):
    avg_c_f = 0
    for member_index in member_index_list:
        open_single = 0
        close_single = 0
        first_connect_node = []
        for first_edge in edge_list:
            if(first_edge[0] == member_index):
                first_connect_node.append(first_edge[1])
        for first in first_connect_node:
            for second_edge in edge_list:
                if(second_edge[0] == first):
                    if(second_edge[1] == member_index):
                        close_single += 1
                    else:
                        open_single += 1
        avg_c_f += 2*close_single/(2*close_single + open_single)

    return avg_c_f/len(member_index_list)
```