Problem Chosen	2020 MCM/ICM Summary Sheet	Team Control Number
D		2000732

Riding on the green: On perfecting Huskies' teamwork

Summary

Right now, we are in a world where the societies are highly interconnected as never before, and the vital importance of teamwork has helped to tackling knotty problems across all sectors and domains. Specifically, effective interactions between teammates during a soccer game contribute to the winning of a match. However, the jury is still out on the evaluating standard of team performance. Just as Everton's motto: "Nil satis nisi optimum", which means nothing but the best is good enough, soccer is not about goals or wins, but about the pursuit of perfection. Thus, instead of just evaluating team performance based on the outcome of the match, we consider the structural and dynamical features which are conducive to the overall improvement of the team.

To solve task 1, we propose the ball-passing network which consists of the average passing network and the N-pass Network. In the average passing network, we use Python and Gephi to visualize the average range of each players motion, the interactions between two players, the passing pattern of each game, and the importance of each player suggested by eigenvector and clustering coefficient. Besides, we discuss three topological scales including the microscale, the mesoscale and the macroscale by making social network analysis, condensed subgroup analysis, and core-periphery analysis in Ucinet. While in the N-pass Network, we track the time-varying metrics of the passing network, which contain the centroid coordination, the centrality dispersion and the ratio of advance, to monitor the dynamic properties of the network.

As for task 2, we establish the BPNN-teamwork evaluation model, using Back Propagation Neural Network (BPNN) algorithm. First, we take a data processing to get some indicators that represent the accuracy, distance and types of passes, coach and side (which reflect the adaptability of the team), and game tempo. Then, we construct the BPNN with five input layers and one output layer, while the number of hidden layers is defined by the empirical formula. After that, we train the network with the objective to minimize the mean squared error of the model. Finally, the validation result shows that BPNN can help to build a reliable evaluating system.

When it comes to sincere advice to the coach on structural strategies according to our model, we first select several representative average passing networks of Huskies and its opponent respectively and carry out a detailed analysis on the topological organization of the network, the pass-based features, the role of different soccer positions and the clustering coefficient to explain which structure the coach should apply for teamwork improvement. Plus, the performance of Huskies during the whole season reflected in the properties of the network is also interpreted to provide more comprehensive suggestions. Second, according to the N-pass Network, our suggestions are more about striking a balance between attack and defense and how to seize the moment. At last, we adopt control variate technique to determine the best condition for perfect teamwork based on our BPNN-teamwork evaluation model.

Eventually, with regard to generalizing our models of team performance, we draw an analogy between soccer and other human activities and summarize some factors that should be captured from broad consensus on what qualifies as preeminent team.

Contents

1 Introduction

1.1 Background

Teamwork is of crucial importance to all human activities, ranging from business to science and from art to sport. However, little is known about the evaluation criteria of good teamwork. Thankfully, the application of Network Science to social systems makes for novel methodologies to unveil the organization of complex human activities, using metrics of the network for further analysis.

Among all of the human activities, sport competitions best demonstrate the significant role that teamwork plays. Soccer, in particular, is widely viewed as the most popular sport world-wide. Yet, due to the nearly uninterrupted flow of ball passing and to the complexity of the play, it is challenging to analyze soccer quantitatively.

Recently, thanks to digital image processing technology, soccer statistics have evolved at an amazing speed, which help to produce critical analysis, insights and scoring patterns of different soccer teams. However, despite the increasing wealth of data, reference indicators for various facets of team performance has not yet been defined objectively. Simply considering statistics such as number of assists, number of shots or number of goals rarely provide a reliable measure of teamwork because subtle changes in the dynamics of the interactions between players may have huge impacts on the matchs outcome. Instead, measures of teamwork are hidden in the dynamical and structural features that have been conducive to the outcome of the game or season and to the overall improvement of the team. As a result, it is of necessity to visualize the interactions between teammates and define an objective and convincing indicator of team performance for team improvement.

1.2 Restatement of the Problem

Competitive team sports are one of the most informative settings to explore team processes. Thus, in this paper, we will analyze how the complex interactions among soccer players on the field impacts their success of the game or the whole season. We will create a network for the ball passing between players and establish several models to evaluate team performances, with the aim of putting forward some sincere strategies to help perfect a team work. The following specific tasks are to be completed:

•Determine the average position of the players in the game according to the data, and then the node location in the network is determined. By determining the number of passing and the value of each pass, we can get the edges between nodes in the network. All these works done, the average passing network is easy to established. And then we can make good use of this network to analyze interactions among players from three topological scales (i.e., the microscalethe mesoscalethe macroscale).

•Construct N-pass network which evolves in time and thus is dynamic in order to focus on the topological organization of the network in stead of the number of passes.

•Identify performance indicators that reflect successful teamwork, which could be used to establish the teamwork evaluation model.

•Use two models to analyze the Huskies' performance in the whole season, and make some sincere suggestions for their improvement in the next season from different aspects.

•Explore the universal influencing factors to evaluate team performanceand give some insight on how to design effective teams, aiming at extending the teamwork model to other social areas.

1.3 Literature Review

The application of Network Science to social systems allows for novel methodologies where it is possible to analyze and quantify the behavior of a soccer team, together with the role of a single player (Gudmundsson and Horton, 2017). In the network, each node denotes the player in the team, while each arc represents the passing events between two teammates. By calculating the number of passes performed in each possession phase, we could determine the brute connectivity degree (number of passes/receptions), and the weight or importance of a player in the oensive phases (Arriaza-Ardiles et al., 2018).

Once the network is constructed, several topological scales (i.e., scales at which the network organization is analyzed) can be identified inside the passing network of a soccer team, including the microscale, the mesoscale and the macroscale (Buldú et al., 2019). And then, we can draw some conclusions about the level of cooperation in terms of different scales

In addition to the average passing network which presents the impact of passing numbers, passing types and passing positions, and thus is static, new methods which involving changes in time should be adopted to analyze dynamic features of the soccer team. A l-pass network was thus proposed by Buldú et al. to probe into the temporal evolution of the network metrics. (Buldú et al., 2019)

While the interaction between team members is visualized by networks, other performance indicators should also be taken into account. Duch et al. combined the flow network with the passing and shooting accuracy of the players, which suggests a natural measure of performance of a player and a team(Duch et al., 2010). More comprehensively, from a list of events (passes and goal attempts) occurred in the game, Cintia et al. defined for every team five pass-based performance indicators, each capturing a different aspect of the passing behavior of the team, and then combined the ve indicators by their harmonic mean to get the H indicator. (Cintia et al., 2015). The results of their study showed that the H indicator of a team is better correlated with its success (goals, attempts, points) than the mere amount of passes.

1.4 Our work

We construct two distinctive models: The ball-passing network model, and The BPNN-teamwork evaluation model.

In the ball-passing network model, we develop the Average Passing Network and N-pass Networkwhich help us to make an analysis of the interaction among players according to ball-passing number and weighted passing number. The specific process is as follows:

Firstly,we choose the social relation model and make an analogy between the set of relationships between social actors to the set of players passing the ball in a football match. Therefore, the model simplifies the multivariate relationship and lays a foundation for the establishment and solution of the ball-passing network model. Moreover, we assign different weights to different types of passes and value each pass to make the connection between players more meaningful. Meanwhile, Gephi is used to visualize the network and observe the interaction of players in the whole game. We also use Ucinet6 to analyze the network not only as a whole, but locally and individually.

More importantly, we develop N-pass Networkwhich makes the interaction between players more than the number of passing. From three aspects(X coordinate of the networks centroid, ratio of advance δ_y/δ_x and the centrality dispersion)we develop dynamic network, facilitating time - dependent analysis of the whole game.

In the BPNN-teamwork evaluation model, we make use of BP-neural network to establish our model and make a solution. Firstly we analyze what makes a perfect football team, and find it that pass completion rate, weighted passes, game tempo, coach, home or away are all the indispensable influence factor. After processing the data of these five influencing factors, the BP-neural network is built up.

Then using the two models we have built, we analyze how the Huskies could perfect themselves and become better next season. We make some graphic analysis from the network graph itself and various kinds of coefficients we get in the network. And the factors influencing the team interaction are explored from the BP-neural network by utilizing the control variable method. Finally, we give Mr. Coach some pertinent advice.

All in all, in the process of establishing the models, we consider various factors and apply various data analysis methods, and finally we obtain quite good results.

2 Preparation of the Models

2.1 Notation

Table 1: Notation

Symbol	Definition
a_{ij}	from player i to player j
α	n by n adjacency matrix which represents the number of passes between soccer players(n≤30)
α^T	the transpose of the adjacency matrix
A	adjacency matrix without direction
L	the distance of passing
d	the coefficient of ball-passing difficulty
v	the value of passing
FMX	adjacency matrix which represents passing during the whole game
n	the number of consecutive passes in N-pass Network
IN	the number of inaccurate passes
TN	the total number of the pass event
PCR	pass completion rate
TT	total time of a game
ST	The time point of each shot
TST	total shooting time
M	the total number of shooting
β_1, β_2	weight factor

2.2 Assumptions

•The sample data of players in the table can represent the real level of the whole game and the whole season. And some players' subjective factors, which are irrelevant to the data, would not influence individual performance and game results.

•There exist differences between each player, since each player has different physical qualities due to their talent and training level. These differences are reflected in the players' on-court positioning, passing ability, offensive/defensive ability and other aspects.

•Ignore the impact of the referee's movements during the game.

•Ignore the influence of weather, wind speed and other objective factors on players(home and away factors need to be considered of course).

•The x and y coordinates of the players' positions are non-absolute coordinates. When the network graph is drawn, the unit length of the x and y coordinates is projected according to the actual size of the court.

3 The ball-passing network model

3.1 Concepts introduction

Since we are building a network for the ball passing between players we should think about both players and their passing actions. Each player has his or her own personality. Their performance, position and value created for the team are different. Since each pass connects different players, different types of passes create "multiple relationships" between players. We know the social network refers to the set of social actors and their relationships, where players are social actors and passing is the relationship between them. Therefore, the social relation model is quite suitable to analyze the passing network. Meanwhile, unicet 6 can be used to analyze the adjacency matrix and cohesive subgroups, so the ball-passing network can be used to analyze passing interaction from the whole network, partial-network and ego-network.

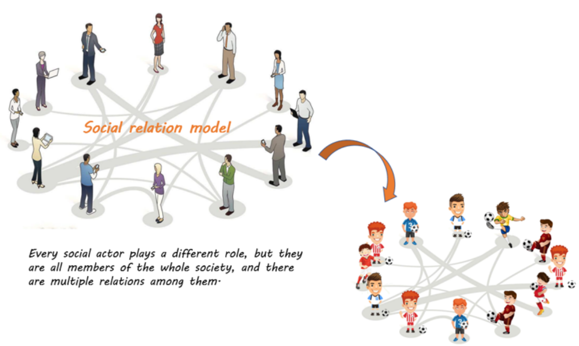

Figure 1: social relation model

3.2 Social relation model

3.2.1 The model construction: the Average Passing Network

• Adjacency matrix of passing

We define that a_{ij} means passing from player i to player j (i,j could be any player)

$$a_{ij}=\begin{cases}1 & pass\ completed\\0 & no\ pass\end{cases}$$

Regardless of the variety of passing methods, the number of passes between two football players can be expressed as an n*n(n≤ 30)adjacency matrix

$$\alpha = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix}$$

Transpose the pass adjacency matrix to obtain:

$$\alpha^T = \begin{bmatrix} b_{11} & \dots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{n1} & \dots & b_{nn} \end{bmatrix}$$

Because one-way passing has directivity, the adjacency matrix from this action also has directivity. In order to eliminate the influence of passing statistics caused by passing direction, a new adjacency matrix A is obtained:

$$A = \begin{bmatrix} a_{11}+b_{11} & \dots & a_{1n}+b_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1}+b_{n1} & \dots & a_{nn}+b_{nn} \end{bmatrix}$$

To sum up A of the whole game, and then input it into unicet6 to get the pass topological graph without weight.

- **Weighed pass**

When we quantify the passing performance of a player, the passing number only represents one aspect of the interactions between players, since every ball passing differs in distance and degree of difficulty. For example, 'Simple pass' is not difficult to be done. 'Smart pass', however, is something more than a simple pass, and the player win some advantage for his/her teammates with this pass.

Let L denote the distance of passing, d denote the coefficient of ball-passing difficulty, and v denote the value of passing:

$$v = d_i * \alpha + d_j * \alpha^T c$$

Add up the value of passing throughout the game, and we can get the adjacency matrix, flowmatrix(FMX), for passing throughout the game:

$$FMX = \Sigma v$$

Turn to next page to find the table of ball-passing difficulty coefficient

- **Gephi visualization and generation of network parameter**

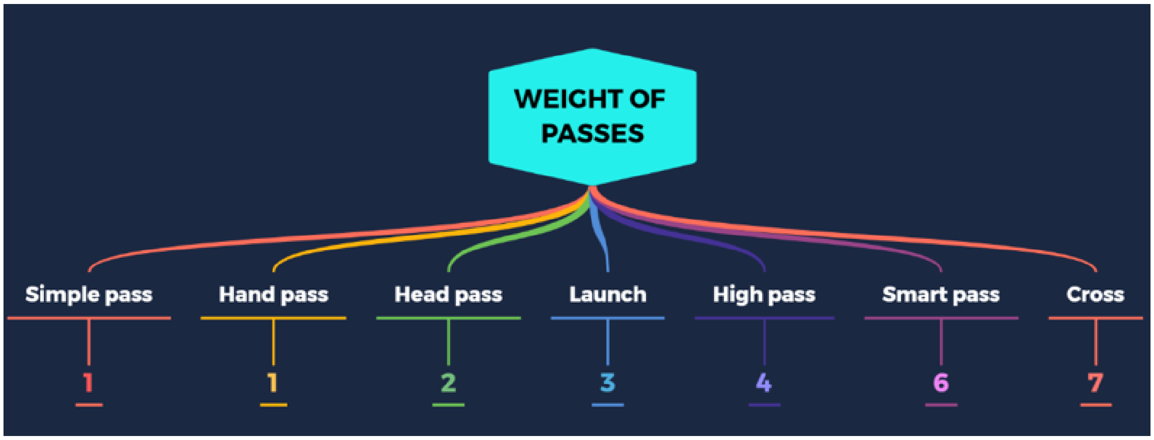

Figure 2: In the figure, different types of passing and corresponding difficulty coefficient d are shown respectively

With the help of gephi, FMX could be visualized. We download the hd pictures of the soccer field from the Internet, and use OpenCV to carry out the projection changes and mapping of the coordinates. From the coordinate data, we can calculate the average coordinates of the location of the passing event. The position obtained is the average position of each player participating in passing, in another word, the position of nodes in the passing network. Similarlyafter changing the coordinate, we can get opponents average passing network.

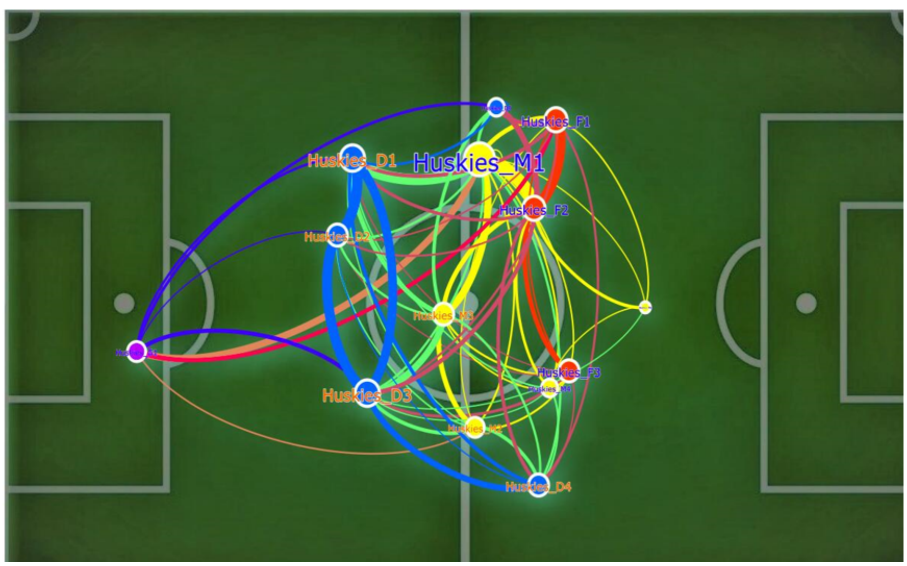

Figure 3: The nodes denote players. The position of each node represents the average range of ones motion.

The size of nodes is proportional to the total number of passes and receptions. Different colors of the nodes correspond to different soccer positionsredforward, yellow: midfield, blue: defense, purplegoalkeeper. The size of the node label is clarified by eigenvector, and different colors of the label denote different modularity class. The arc between two nodes represents the interaction between two players, and the width of the arc is proportional to the weight according to v.

Here are some parameters of the network may be used later:

Table 2: parameters of the network

parameter
network diameter
modular degree
average clustering coefficient
centrality distribution
degree distribution
average path length

3.2.2 Model solution and analysis

In the social network analysis method, the research can be divided into three levels according to the network type: the whole network, partial-network and ego-network. The whole network analysis to the passing characteristics of the whole team; partial-network analysis refers to pass coordination characteristics in one part; ego-network, however, reflects the role of individuals in the passing network.

• **the whole network anlysis**

Since passing is a one-way relationship, the original matrix is a directed multivalued matrix. When it comes to measuring player performance and the interaction level of the team as a whole, this kind of directional transfer is somewhat redundant. Therefore, we adopt the method of transpose the original matrix to transform the directed multi-valued matrix into a symmetric matrix based on the sum of the output and input degrees. And then draw the topology of passing: From the whole network, we can

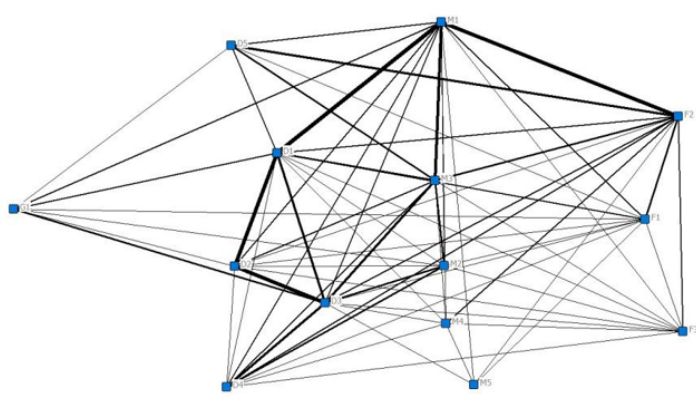

Figure 4: Nodes represent different players, and the line between nodes represents the number of direct passes made by players

see the general distribution of passing position and passing frequency. At the same time, by giving weight to different types of passing, we can intuitively see the distribution area of valuable passing, and then judge whether the team's playing style is attack or defense. In addition, observe the passing line. If there are thick and dense lines in a certain position, To a large extent it indicates that this area is the focus of our attack or defense, so further tactical adjustments should be made based on the result.

- **partial-network analysis**

the analysis of cohesive sub-groups the analysis of cohesive sub-groups is a kind of substructure analysis method of social network, which can simplify the complex social network structure and help us find the substructures contained in the network, so as to visualize the network structure more effectively and concisely. In the passing network, there are some cooperative teams whose players often pass the ball to each other. The analysis of cohesive sub-groups can help us analyze the structure and characteristics of them. Players who often cooperate with each other pass more often, forming factions in the passing network. This analysis is based on reciprocal faction analysis, which is carried out along the passing network in Ucinet6. Here are some sub-groups.

1: M3 D4 F3 D2 M1 M2 D3 D1
2: M3 D4 F2 F3 D2 M1 D3 D1
3: D4 F2 F3 D2 M1 D3 D1 F1
4: D4 F3 D2 M1 M2 D3 D1 F1
5: M3 D4 F2 F3 M4 M1 D3 D1
6: G1 D2 M1 M2 D3 D1 F1
7: M5 F2 M4 M1 D3
8: M5 F2 M1 D3 F1
9: M3 F2 D5 M1 D1
10: F2 D5 M1 D1 F1
11: G1 D5 M1 D1 F1

There are 11 factions in the passing network. Within the subgroup, there is reciprocity and accessibility between members, but the interaction between members of the subgroup is far greater than that between the internal and external. Therefore, players who appear in multiple subgroups at the same time are likely to be star players, who play a role in connecting multiple subgroups.

3.2.3 Optimization of model: N-pass Network

Through the characterization of passing number and passing value vwe build Average Network. However, the interaction of a team may be interpreted as a consequence of the higher number of passes between players, which could lead to statistically significant differences in a diversity of network metrics. But is just the number of passes behind the differences of the network parameters? and is it enough to look at the average values of the network metrics? According to what Buldu has said in his paper, on the one hand, we are going to define passing networks as non-static entities, thus evolving in time, and we will track the evolution of their parameters. On the other hand, we are going to exclude the importance of the number of passes, in order to just focus on the topological organization of the networks. With these two objectives in mind, we construct the Npass networks of a team as the networks containing n consecutive passes, with n L, being L the total number of passes during the match.

First of all, let's set the value of N, let's refer to the setting in the literature, and take n = 50. since it is a value low enough to allow a tracking of the network evolution along the match and, at the same time, high enough to guarantee the creation of a network between players

Figure A shows how the position of the team moves forward and backward during the match. In this particular case, Huskies plays, most of the time, more advanced than opponent14, which did lead

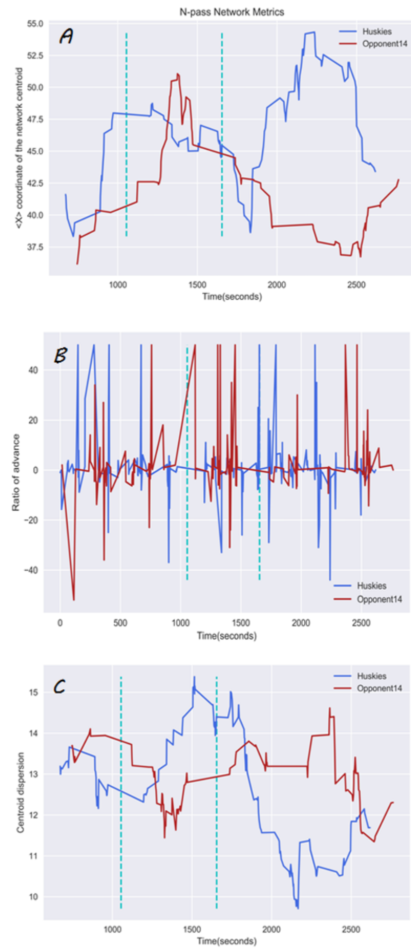

Figure 5: Huskies vs. opponent14 Game 14 1H.Temporal evolution of the network parameters: (A) X coordinate of the networkscentroid, (B) ratio of advance δ_y/δ_x and (C) the centrality dispersion ECdisp. Cyan vertical dashed lines indicate the two moments at which Huskie perform a shot(opponent14 did not make any shot in the 1H).

to an advantage in the result. Note how the centroid of both Huskies and opponent14 seem to be not so stable, while Huskies has higher top point, arriving to its maximum value around 2200s. Also note how Huskies is the first team to construct the 50-pass network around 200s, while opponent14 required 400s. In Fig. B, we plot the ratio of advance of the 50-pass networks of both teams. Again, we can see fluctuations of the parameter during the match. Specifically, opponent14 has a lowest value during the first part of the match, but unfortunately, they did not make an effective attack or shot. Moreover, we can observe the advance ratios of the two team have no obvious upward or downward trend. Finally, Fig. C shows the fluctuations of the centrality dispersion of the players of both teams. We can observe how Huskies has a high level of the centrality dispersion between 1500s and 1700s, which seems to be related with the period where the centroid of the team advances towards opponent14s goal (shown in Fig. A). This change of the centrality distribution could be related to a change of the style of playing. Since centrality dispersion increases, there is a higher heterogeneity in the importance of the players in the passing networks, which could be related to the fact that a few players are taking the lead of the team. It is worth mentioning that this change in the organization of the passing network does seem to be effective, since the second shot of Huskies comes around to the maximum of centrality dispersion.

4 The BPNN-teamwork evaluation model

4.1 Concepts introduction

Just as we can't judge a team's success just by its final work, we can't judge whether a team's success only depends on whether it wins or scores. We should look for the factors that affect a team in the process of team work. In a football match, the team's flow in the game, the success rate of passing, whether the game is played home or away, and who is the coach will affect the quality of team work. Considering synthetically, we use BP neural network to model and analyze team cooperation level.

BP (back propagation) neural network is a multilayer feedforward neural network based on error back propagation, which is the adjustment rules of weight and threshold adopt error back propagation algorithm, which is a learning algorithm of neural network with tutor. BP network can learn and store a large number of input-output mode mapping relationships without revealing the mathematical equations describing the mapping relationship in advance. The learning rule of the network is to use the steepest descent method to adjust the weights and thresholds of the network through back propagation, so as to minimize the sum of squares of the network errors. The topological structure of BP neural network model includes input layer, hidden layer and output layer. The hidden layer can be extended to multiple layers. As long as there are enough neurons in the hidden layer, this network can be used to approximate any function. Here we don't know how many team influencing factors contribute to the whole team work evaluation, so using BP neural network is the most scientific and efficient modeling method.

4.2 Back propagation neural network model

4.2.1 The model construction

• **influence factor**

•Adaptability: Adaptability refers to the degree of reflection and coordination of the players in the face of the game situation. Therefore, we choose the performance of the players home or away field as well as in the coaching of different coaches to discuss the adaptability of the team.

Home or away: we record home as 1 and away as 0. Coach: the three coaches are recorded as 1, 2 and 3.

•Flexibility: Flexibility refers to the running and passing of players on the field. We select the following two factors to investigate the level of team cooperation.

Passing times: We have weighted the number of passes according to the difficulty coefficient of different passes.

Tempo: To quantify tempo, we divide the total time of each shot by the number of shots, and then process the total number of matches, which represents the degree of tempo.

•Team coordination factor: The cooperation of team members also determines the success of a team to a certain extent. We choose the rate of success passing between team members for analysis.

Pass success rate: Compare the total pass times of players with pass success times.

•Other factors: There may be other factors in measuring the level of cooperation of the team, but they are not the main factors, so we ignore the influence of these factors.

• **Data processing**

The Ratio of Pass Completion

We use Pandas to read the fullevents.csv. In the game of iwe screen out the data of MatchID=i, TeamID=Huskies,EventType=Pass. Count the number of NaNs in the column of DestinationPlayerID', and recorded it as INincorrect number, which represents the number of inaccurate passes, and the length of the 'OriginPlayerID' column, and recorded it as TNtotal number, which is the total number of the pass event. After that the pass completion rate is:

$$PCR = 1 - \frac{IN}{TN}$$

Weighted Passes

According to the weighting rules mentioned in figure, we define the weights of different passing events in the EventSubType column. Then, we add a new column called PassWeight for each game to store these weighted passes. By summing the values in this column, we can get the total weighted passes of each game. See ball-passing network model model for the detailed derivation process

Game Tempo

Considering that the number of shots has a great impact on the rhythm of the game, when the average shooting time is smaller, the rhythm is faster, but on the contrary, the pace of the game is slower. At the same time, the success rate of passing is also a major factor affecting the flow of the game. The higher the success rate of passing, the faster the pace of the game. We filter the data in file fullevents.csv whose 'EventType' equals to Shot, by adding up the time of the first half and the second half for each game, we can get the total time of a game TT(total time)

Obviously, each value in the 'EventTime' column of the filtered data denotes the time when a shot event happened, and the length of the 'EventType' column is the number of shots.

We record that the time point of each shot is st, so the sum of shot time(TST) is

$$TST = \Sigma ST$$

Note that the total number of shots is m, so the rhythm of the game is

$$GT = \beta_1 * \frac{TST}{M} + \beta_2 * PCR$$

$\beta_1, \beta_2$are the weight factors Through the data processing, the five data in 38 sessions are obtained and tabulated at the bottom of last page:

We take whether the team has beat the opponent as the main measurement basis. The success of the game is recorded as 1, the tie is recorded as 0, and the failure is recorded as -1. Then we use BP neural network model to model and analyze.

• **Building neural networks**

Input layer

There are 5 input layer parameters in the system. And the figure shows the network layer structure:

Input sample normalization

Because of the difference of the value size and dimension of the selected nodes, if it is directly used as the network input, it may lead to the network training can not meet the accuracy requirements or over fitting and other problems. Before the network training, each eigenvalue must be normalized or

TEAMWORK INDICATORS

	1	2	3	4	5	6	7	8	9	10
Pass Completion Rate	0.766	0.652	0.773	0.794	0.814	0.768	0.766	0.696	0.629	0.787
Weighted Passes	766	492	637	755	679	764	893	696	478	768
Game Tempo	0.232	0.265	0.330	0.331	0.244	0.275	0.267	0.220	0.221	0.300
Coach	1	1	1	1	1	1	1	1	1	2
Side	1	0	0	1	0	1	1	0	1	0

TEAMWORK INDICATORS (CONTINUED)

	11	12	13	14	15	16	17	18	19
Pass Completion Rate	0.642	0.550	0.815	0.842	0.783	0.503	0.749	0.752	0.775
Weighted Passes	538	431	486	667	583	336	636	654	560
Game Tempo	0.276	0.306	0.342	0.296	0.185	0.488	0.249	0.294	0.335
Coach	2	2	2	2	3	3	3	3	3
Side	1	0	0	1	1	0	0	1	1

TEAMWORK INDICATORS (CONTINUED)

	20	21	22	23	24	25	26	27	28	29
Pass Completion Rate	0.710	0.794	0.677	0.688	0.693	0.752	0.808	0.759	0.713	0.67
Weighted Passes	628	800	627	664	771	638	663	640	667	607
Game Tempo	0.301	0.237	0.216	0.237	0.278	0.364	0.227	0.193	0.243	0.28
Coach	3	3	3	3	3	3	3	3	3	3
Side	0	0	1	0	1	1	0	1	0	0

TEAMWORK INDICATORS (CONTINUED)

	30	31	32	33	34	35	36	37	38
Pass Completion Rate	0.772	0.765	0.599	0.657	0.830	0.740	0.740	0.755	0.806
Weighted Passes	711	743	394	574	673	766	714	674	562
Game Tempo	0.248	0.246	0.209	0.278	0.300	0.204	0.243	0.245	0.287
Coach	3	3	3	3	3	3	3	3	3
Side	1	0	1	1	0	1	0	1	0

Figure 6: the five indicators in 38 sessions

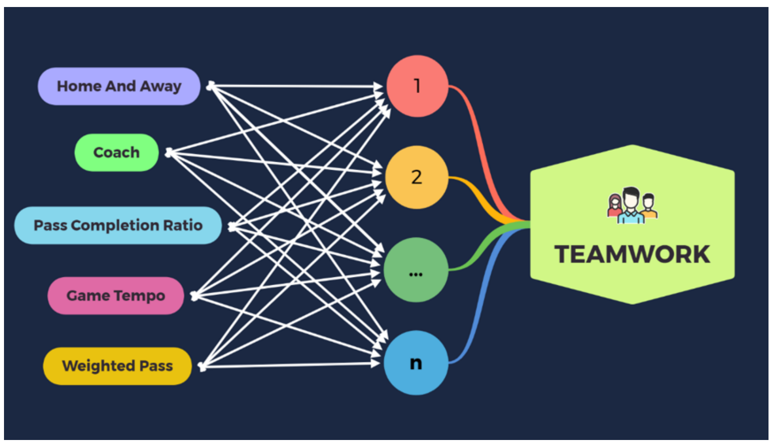

Figure 7: the five data in 38 sessions

standardized. In this paper, the method of min-max standardization, that is deviation standardization, is selected, which maps all data in the range of 0-1. The function formula is:

$$x^* = \frac{x - min}{max}$$

Max refers to the maximum value of data measured by the same indicator, and min refers to the minimum value of data measured by the same indicator.

Number of neurons in the hidden layer

The hidden layer of BP neural network is very important for network training. Improper setting of network nodes will directly lead to network training failure. Proper setting of the number of nodes in the hidden layer can not only reduce the number of learning times, accelerate the convergence speed of the network, but also affect the network performance if the hidden nodes are too large or too small.

Generally, the number of nodes in the hidden layer is set according to the empirical formula. The empirical formula used in this paper is:

$$l = \sqrt[2]{m + n} + a$$

Among them, the value of a is generally rounded from 1 to 10, and then the value of a is changed according to the effect of network training. According to the test, when the number of nodes is 9, the error is the lowest.

Output node settings

In essence, neural network is a kind of mapping. Using network training to approach a difficult relationship, the output of network is the result we need to get. The result we need to get is the level of teamwork (based on the success or failure of the competition, so we can set a node.

Training network

Select the elastic gradient descent method in BP network training function, which is trainrp, to establish prediction model. Before using neural network for data association, first train the network. The network coverage should be improved as much as possible, and the network error should be reduced so that it can adapt to a variety of special associated goals. The training error of the network model is 0.01, the maximum training times is 1000, the iterative process is 20, and the learning rate is 0.05. The network training diagram is as shown in the figure .

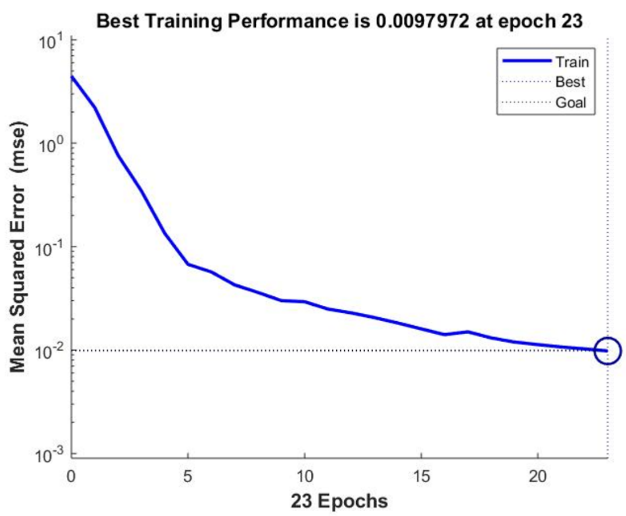

Figure 8: It can be seen from the figure that the training requirements can be basically met after training.

4.2.2 Model solution and analysis

Network test and verification

After 38 training samples of neural network are trained, 4 randomly selected verification samples are brought into the trained neural network to get the prediction results. The team cooperation level obtained through neural network analysis is consistent with the actual situation, and the error is within the acceptable range, so the model is relatively successful.

Table 3: prediction results

Actual group number	4	9	16	28
Actual level of teamwork	-1	-1	0	-1
Predicted level of teamwork	-0.8632	-0.7684	0.1646	-0.8524

Conclusion

There are many factors that affect team cooperation. This model constructs a reasonable evaluation system, and proposes to use BP neural network to simulate it. Experiments show that BP neural network can build a stable evaluation system and achieve good experimental results.

5 Recommendations

Dear Mr. Coach,

Hello! Firstly, we sincerely appreciate it that you give us a chance to do something for Huskies, our hometown team. And after the analysis of the whole season data of Huskies, we do get a lot of findings which is according to average network, the N-pass network and the BPNN-teamwork evaluation model. Now I will share these with you as below.

5.1 According to Average Network

We choose two representative matches where Huskies played against Opponent14, i.e. the $14^t h$ and the $38^t h$ match for analysis. Details of the two matches are as follows:

Table 4: two matches against opponent14

14: Huskies	4: 0	win	home	Coach2
38: Huskies	1: 3	lose	away	Coach3

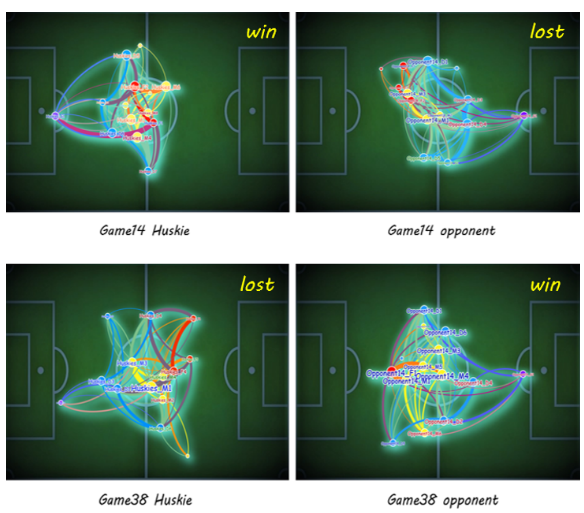

Figure 9: The specific graphic description has been mentioned above and thus will not be repeated

5.1.1 Analyzing the network

First, we consider the macroscale of the network and draw some conclusion based on its outline. In game 14 and game 38, the winning team beat the other by a large margin. As is shown in the figures, the shape of the winning teams network is roughly an inverted trianglethus we vividly name it as the god blessing triangle (GBT). In comparison, the arcs of the losing teams network are in a mess and hardly can we abstract an inverted triangle from the network. Therefore, we suppose that GBT may be an ideal structure which represents the trade-off between attack and defense, and indicates better interaction between teammates. Thus, it would probably be safe to say that GBT is a scientific organization.

Second, in terms of the mesoscale of the winners network, we can find some thick links which denotes strong interactions. This means that some star players act as the mainstay of the team during the match. It is these star players who connect players in different positions and help to construct the GBT structure.

Besides, we discover that the star players of the winning team are mostly forwards, while the star players of the losing team are mostly midfields or defenses. As a result, we suppose that it is of vital importance to guarantee the leadership of the forward in the field. Thus, when assigning a player to the forward position, the coach is expected to choose the one who can interact with others well and is full of energy. Whats more, reasonable formations should also assure that forwards are not restrained by opponents.

Finally, we move on to the microscale i.e. the individual level of the network. Analyzing the role that goalkeeper plays from the four networks, we find that in game 14, the goalkeeper undertook the responsibility as a star player while in game 38, the goalkeeper showed a low degree of engagement, which could have given rise to the lose of game 38. Consequently, the interactions with goalkeepers is of equal importance to those with forwards, midfields and defenses.

We visualize the average network (undirected) of Huskies and Oppoent14 respectively.

5.1.2 Average clustering coefficient analysis

Fig Huskies clustering coefficient in game 14. The height of the columns denotes the concentration of the nodes in the network i.e. the interaction of the team

Clustering coefficient corresponds to a measure representing the node clusters, representing the players frequently interacting in the game-play network. That is to say, the measure quantifies how important the player has been when the team is in possession of the ball. If node i hasn_i adjacent edgesand between these adjacent edges, there are m links. Thus the clustering coefficient is defined as:

$$C_i = \frac{m}{n_i(n_i - 1)/2}$$

While the average clustering coefficient is the mean value of individual coefficientswhich represents the cohesion of the whole network. As is shown by the figures abovein game14 and game38the winners average clustering coefficient and total triangles are larger than those of the loserswhich denotes that the cohesion of the winners whole network is higher than that of the losers and that players are tied more closely. So the interaction between players is of crucial importance to the match.

5.2 According to N-pass Network

Using N-pass network, we can analyze interaction indicators apart from the number of passes. Still, we compare the performance of Huskies and Opponent14 in game 14 and game 38 to generalize essential factors which make for the success of the match according to three metrics of the N-pass network.

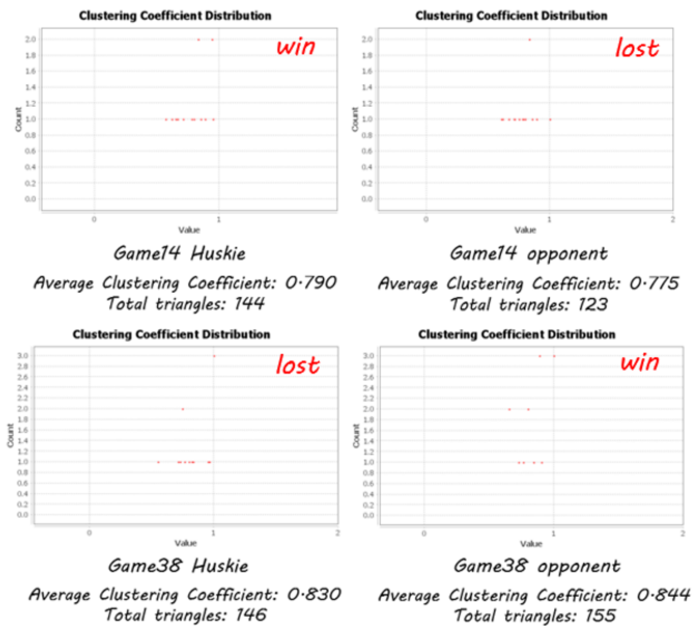

Figure 10: Average Clustering Coefficient

Figure A shows how the position of the team moves forward and backward during the match. In game14, Huskies plays, most of the time, more advanced than opponent14, however, in the game38 the case is quite the opposite, which lead to score gap in the result. Huskies has higher top point in game 14, while opponent14 reach its top point at the beginning of the game. Maybe we can draw a conclusion that a moderate shift of <X> coordinate of network is quite benefit the team to some degree. Also note how the winning side of two games constructed the 50-pass network first, while the lost side required longer time. Therefore, we can find that the larger the number of passes is, the better the tempo of game is and the greater the chance of goal is.

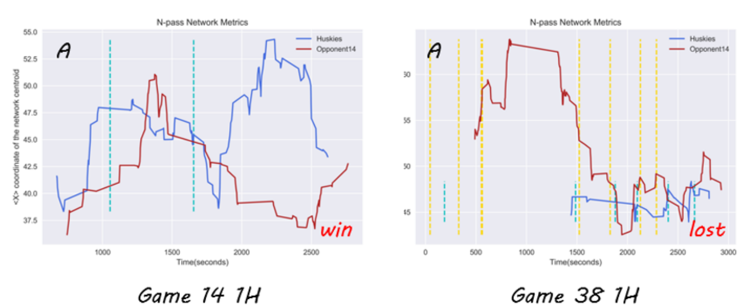

Figure 11: <X> coordinate of network

In Fig. B, we plot the ratio of advance of the 50-pass networks of both teams. Again we can see fluctuations of the parameter during the match. Specifically, comparing the advance ratio of Opponent14 in game 14 and game 38, we can find the reason why Opponent14 could beat Huskies in game38. Compared with the advance ratio in game 14, the ratio of advance is less volatile in game 38, and the fluctuation happened mostly when players shot. Thus Opponent14 seize the opportunity to

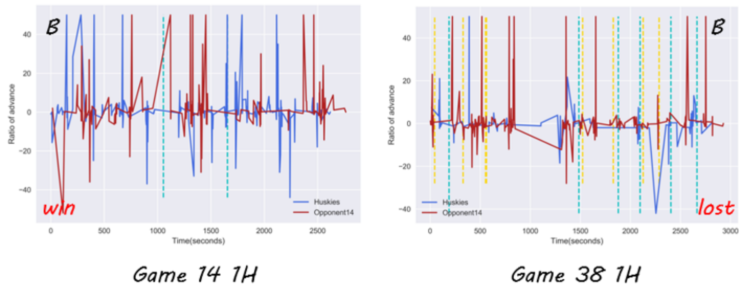

Figure 12: <X> ratio of advance

advance in game 38. This result shows that the coach should stress the significance of controlling the ratio of advance. Players are expected to advance towards opponents goal when the time is right and pass balls paralleled to opponents goal otherwise.

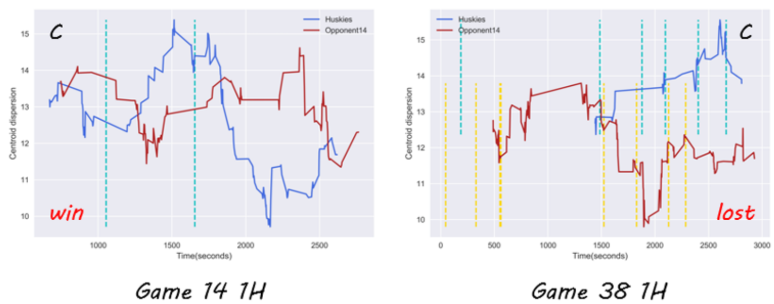

Figure 13: centrality dispersion

Finally, Fig. C shows the fluctuations of the centrality dispersion of the players of both teams. We have known that since centrality dispersion increases, there is a higher heterogeneity in the importance of the players in the passing networks, which could be related to the fact that a few players are taking the lead of the team. In the game38, the centrality dispersion of Huskies increased over timeand is higher than that of Opponent14, which could be ascribed to the contribution of the star player. But unfortunately, this change in the organization of the passing network does not seem to be effective, which reflects that individualistic heroism may lead to the lack of coordination when the team attacks and loopholes when the team defenses and is adverse to the improvement of teamwork.

5.3 According to the BPNN-teamwork evaluation model

We selected the first game of 38 games to analyze the influence of different factors on huskies' team cooperation level. Using the established neural network model, the control variable method is adopted to explore how these five factors change to get the optimal team cooperation level:

• **Change the Coach**: It can be seen from the table that the influence of different coaches on team cooperation factors is quite different. For the coach, coach 1 may be more strategic and for the team members, coach 1 may be more suitable for their taste.

• **host or guest**: From the table, we find that the team is very demanding for the host and guest field operations. The team cooperation factor at home is significantly higher than that away, which greatly

Table 5: Change the Coach

coach	1	2	3
Predicted level of teamwork	0.9236	0.8928	0.6557

Table 6: Host or Guest

host or guest	1	0
Predicted level of teamwork	0.9236	0.5692

reflects the adaptability of husky team needs to be improved.

• **Change the PassCompletionRate** : It can be seen from the table that when the rate of success passing is significantly reduced, it will be greatly reduced, but when the pass success rate is near a certain range, it will not significantly change the level of team cooperation.

• **Change the GameTempo**: We can see that with the change of the tempo of the game, the level of team cooperation fluctuates up and down. The guess may be that for different games, it is suitable to adopt different competition tempos. More data is needed to further study the choice of the best tempo.

• **Change the WeightedPasses**: It is not difficult to find out from the table: the more effective passing there is, the stronger team cooperation level it will be. When the effective passing reaches a certain number, the effect on the level of cooperation will be not significant. Therefore, it proves once again that it is very important to attach importance to the interaction among players.

Some constructive suggestions are as follows:

• **Player training and Selection**

In the 38 games of the whole season, Huskies's players have different performances. Some of them have performed well, achieved a good interaction with their teammates and played a key goal at a critical time. However, the interaction level of some team members is relatively low, resulting in "heroism", ignoring the cooperation with their teammates, or their own ability is not outstanding, which does not play a positive role in the team. According to our ball-passing network model, we need star players not only to have their own professional skills, but also to be able to interact with other players to pass more valuable balls, as to drive the whole team to win. At the same time, according to our good team evaluation model in normal player training, we should pay attention to the accuracy of player's passing, let them do more passing and shooting exercises, and encourage players to actively cooperate with the coach to make corresponding training and improvement; in order to improve the adaptability of players, the team should participate in more away games in training games, so as to improve the psychological quality of players in away games, and ensure stable development wave. Therefore, in the next season to make some personnel adjustments, in order to ensure that star players can get a higher salary, and timely terminate the contract performance of the poor players, to improve the overall level of the team.

Table 7: Change the PassCompletionRate

PassCompletionRate	0.7655	0.7155	0.6655	0.8155
Predicted level of teamwork	0.9236	0.8728	0.7124	0.9327

GameTempo	0.2321	0.1821	0.2821	0.3321
Predicted level of teamwork	0.9236	0.9178	0.9198	0.9218

Table 8: Change the WeightedPasses

WeightedPasses	766	866	666	566
Predicted level of teamwork	0.9236	0.9288	0.8479	0.6981

• **Tactical adjustment**

Good tactics are half of success. Having good tactics can not only give full plays to the role of star players, but also fully mobilize the enthusiasm of team members, so that each player's ability can be fully exerted on the field. Not only that, good tactics can also effectively contain the fierce attack of the opposing team and ensure that our situation is always good. According to our ball passing network model, we can know that in the process of players' interaction, we need to be fully close to GBT mode to ensure efficiently offensive posture and excellent defensive ability. At the same time, according to our good team evaluation model, it can be known that the key to win is to ensure the proper operation tempo.

• **Learning from Coach 1**

A good coach can fully improve the team's ability. Through the analysis of BP neural network, it can be seen that coach 1 has a greater advantage in improving the overall team cooperation factor of the team. It is suggested to learn from coach 1 next season.

6 Generalized models of team performance

Through the establishment, solution and optimization of the two models above, we have found some evaluation factors that affect team work in the process of improving Huskies. In the network above, we find that the interaction between team members is very important. In the team work evaluation model, we consider about adaptability, flexibility and the team's coordination. In addition, due to some limitations of team cooperation in football projects and the incompleteness of the data given, we can not only extract all the influencing factors from the team cooperation of football, which is a competitive game. To develop a common model of team performance, we need to capture other factors, and then promote the model to make it more universal. If we want to generalize it, we first need to discuss what characteristics a good team should have?

• **Excellent leadership core**: the leadership ideas and ideas within the leadership core should be consistent, and the leadership core should have outstanding decision-making ability to make the best decisions in response to different situations; the keen observation ability should not only be familiar with the strengths and weaknesses of each member of the team and assign corresponding tasks, but also master the team information of competitors; strong coordination ability to ensure team awards, the rationality of punishment mechanism, how to mediate the conflicts when the players have conflicts, and how to maintain team harmony.

• **Efficient division of labor**: Team members perform their duties and cooperate with each other

• **Unified goal**: Members have a good vision for the future, have a common belief to support their progress, and make them have a strong executive power to complete tasks on time or even ahead of schedule.

• **High adaptability**: Whether the team can play normally or even super normally under different conditions.

• **High flexibility**: Whether the team can enter the state quickly and adjust the rhythm timely

• **Stability of operation system**: Whether the team can maintain stable and normal operation with the passage of time and the change of personnel.

• **Degree of interaction and cooperation**: High frequency interaction and cooperation can enhance tacit understanding and cooperation among team members, so as to improve work efficiency and quality.

• **Good team atmosphere**: Team members trust each other, care for each other and think for each other.

Therefore, we need to take the information above into account. In the question 2, we need to add relevant items and adjust parameters in the team evaluation model built by the neural network to get the general model of team performance.

7 Strengths and Weaknesses

7.1 Strengths

7.1.1 Strengths 1

The social relationship model simplifies the complex connections in the social relationship, constructs the whole relationship network in the form of nodes and edges, and can carry out modular analysis, which greatly simplifies the difficulty of problem, and provides a feasible analysis path for the diversified relationship analysis.

7.1.2 Strengths 2

Using BP neural network. Neural network has the ability of self-adaptive, which can give an objective evaluation to the multi-index comprehensive evaluation problem, which is very useful for weakening the human factors in the weight determination.

7.2 Weaknesses

7.2.1 Weaknesses 1

In the social relationship model, there is no differentiation among the members of the factions, that is, all members are equal in graph theory and no individual data can be obtained from the core edge analysis.

7.2.2 Weaknesses 2

The biggest problem encountered in the application of ANN is that it can't provide analytical expression, the weight can't be interpreted as a regression coefficient, and it can't be used to analyze causality. At present, it can't explain the significance of the weight of ANN from theory or practice. A large number of training samples are needed, the accuracy is not high, and the application range is limited. The biggest obstacle is the complexity of the evaluation algorithm. People can only deal with it by computer, but the commercialized software is not mature enough.

References

[1]Buldú, J.M., Busquets, J., Echegoyen, I. et al.Defining a historic football team: Using Network Science to analyze Guardiolas F.C. Barcelona. Sci Rep, 9, 13602. (2019).
[2]Cintia, P., Giannotti, F., Pappalardo, L., Pedreschi, D., Malvaldi, M.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. (2015).
[3]Duch J., Waitzman J.S., Amaral L.A.N. Quantifying the performance of individual players in a team activity. PLoS ONE, 5: e10937. (2010).
[4]GÜRSAKAL, N., YILMAZ, F., ÇOBANOLU, H., ÇALIYOR, S. Network Motifs in Football. Turkish Journal of Sport and Exercise, 20 (3), 263-272. (2018).
[5]Krzysztof Palczewski. Identification of the football teams assessment model using the COMET method,2491-2501,(2019)
[6]William B. Sheehan. Using cooperative networks to analyse behaviour in professional Australian Football,291-296,(2020)
[7]E.Arriaza-Ardiles. Applying graphs and complex networks to football metric interpretation,236-243,(2018)
[8]Frost, Lionel. Labour market regulation and team performance: The Victorian Football Leagueš Coulter Law, 19301970,276-287,(2012)
[9]Montoliu, Raúl. Team activity recognition in Association Football using a Bag-of-Words-based method,165-178,(2015)
[10]Narizuka, Takuma. Statistical properties of position-dependent ball-passing networks in football games,157-168,(2014)
[11]Takuma Narizuka. Statistical properties of position-dependent ball-passing networks in football games.(2014)

Appendix

Here are some HD passing network pictures：

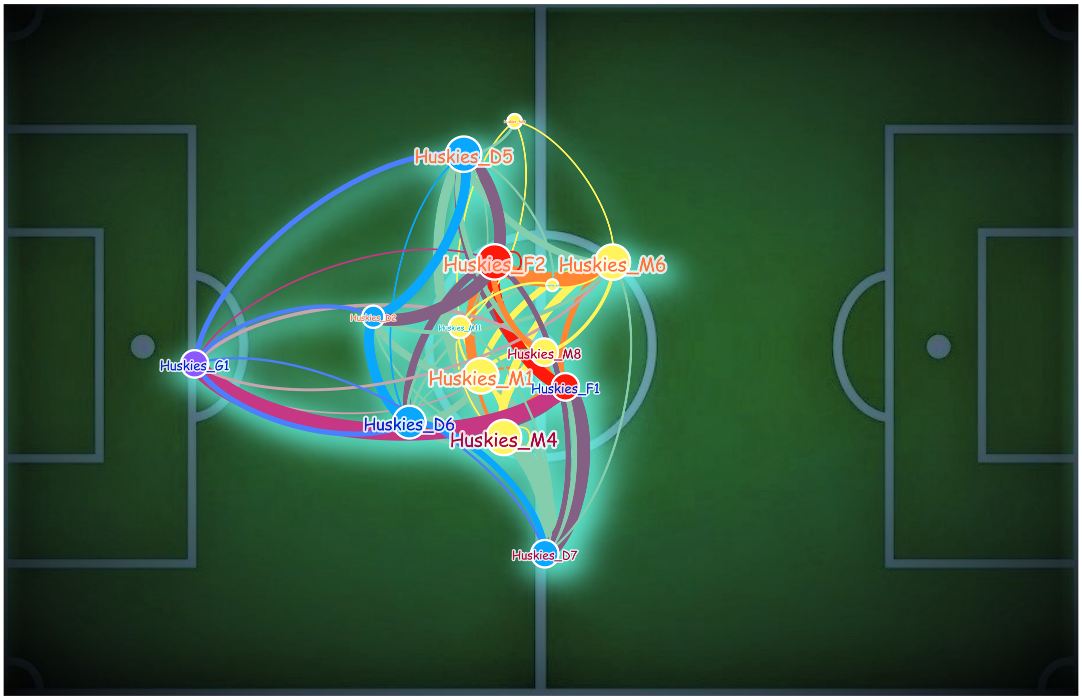

Figure 14: HD 1

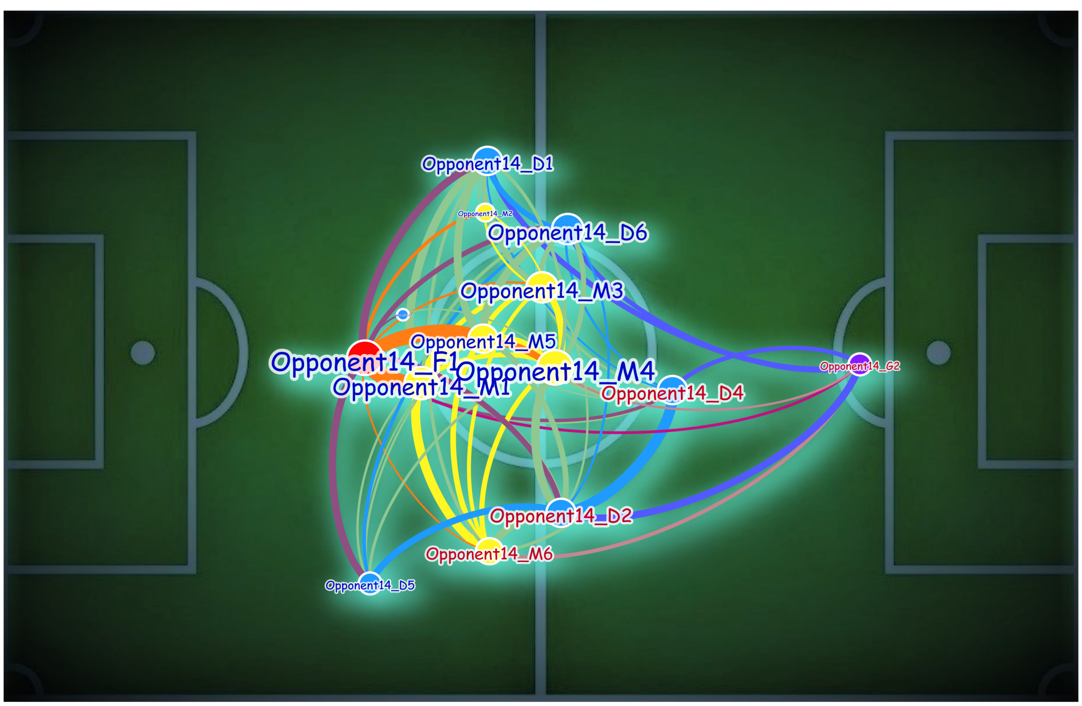

Figure 15: HD 2

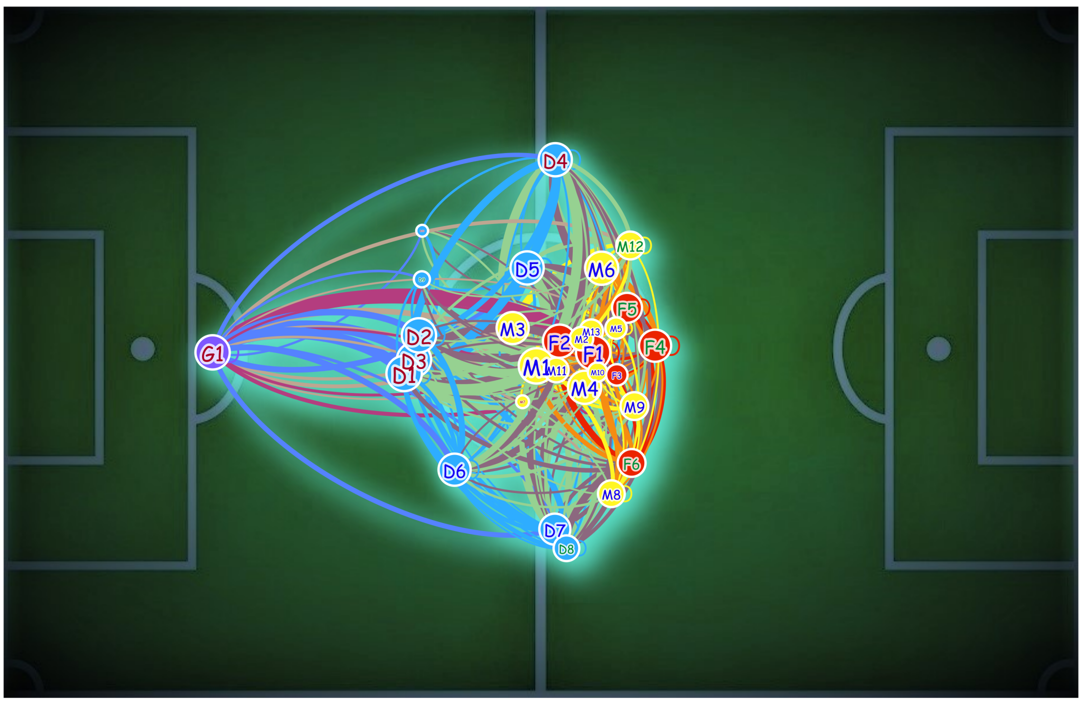

Figure 16: HD 3