Problem Chosen	2022	Team Control Number
D	MCM/ICM Summary Sheet	0000

Football Team Performance Evaluation and Strategy Analysis

Summary

As the world grows interdependence, mankind is confronted with complicated challenges. Therefore, it is of great necessity for members with different skills to address complex problems collectively. To explore the appropriate structure of teamwork, we develop four models as follows:

First, we establish **The Football Passing Network Model** for Task 1. To analyze the properties of our football passing network, we propose structural indicators from two main aspects: macro (Network Connectivity, Centrality Distribution, Spatial Features) and micro (Network Patterns, Clustering Coefficient). Furthermore, we analyze these indicators combined with time series in one match and the whole season to describe dynamic change of the network.

Second, we build **The Comprehensive Hierarchical Analysis of Teamwork Model** for Task 2. We propose 8 ability indicators, which measure the abilities of four types of players respectively and use **Analytic Hierarchy Process (AHP)** to calculate the weights. In order to evaluate teamwork in a match, we quantify another five indicators. Using **Binary Logistic Regression**, we calculate the weights for 9 indicators in total and acquire the comprehensive indicator.

Third, we construct **The Strategic Analysis Model** for Task 3. In order to zone the football field, we draw **heat maps** to reflect the pass and shot frequency at each coordinate. Structural strategy and pass motifs are designed for the Huskies. We use the sliding window method to analyze 9 teamwork indicators and propose improvements for the Huskies by the Mann-Whitney hypothesis test.

Finally, we create **The Generalized Successful Teamwork Model** for Task 4. In this model, we propose five aspects which evaluate the performance of a general team.

Our model passes the model test and sensitivity analysis, indicating that the model has strong stability and can be widely used in explore the performance of teamwork.

Keywords: Network Analysis; Binary Logistic Regression; Analytic Hierarchy Process (AHP)

关键词要包含研究问题（主题）

Contents

1 Introduction

1.1 Background

With the increasing trend of globalization, human faces more complete challenges. To cope with a variety of problems and pursue interconnected development, individuals with various expertise form teams. Team success depends on not only individual ability and effort, but also the cooperation between them. Favorable cooperation can cause to synergy effects, which means that one plus one is greater than two. Therefore, team cooperation, which determines the efficiency and success of the team is extremely important. Football, a common competitive team sport is a representative example to explore teamwork. Numerous indicators and methods can be used to judge the team performance.

et.al

In 2010, Pena and Touchette [1] firstly attempted to analyze the matches of FIFA 2010 World Cup using typical network analysis tools. Although the data are limited, the conclusions mean a new world of team analysis. Before long, Cintia and others [2] adopted passing networks showing more effective predictions of the outcomes of football games. To the author's knowledge, the amounts of competitions they considered are limited and indicators they considered are unitary. Although the accuracy of the results they received are acceptable, the rationality is doubtful [3, 4].

1.2 Problem Analysis

Task 1 requires us to build a football passing network between teammates based on dataset "passingevents.csv" and propose structural indicators, such as network patterns between teammates, time and space to quantify the network. It is of great importance to identify the network in multiple dimensions.

Task 2 requires us to propose perfomance indicators and team level processes. A model, which is created by indicators is applied to test strutural, configurational and dynamical aspects of teamwork.Additionally, it requires us to clarify the application range of strstegies.

Task 3 requires us to illustrate structural strategies which are able to improve the Huskies's success. It is of great essence to explore the teamwork model that we build in Task 2.

Task 4 requires us to generalize our analyses in the Huskies to other teams in our societies. Aimed to design more effective teams and evaluate teamwork, we should develop a generalized model based on indicators we create in above Tasks and other new aspects.

1.3 Views of Our Work

这里最好再概括一下，细节看图
或者写一个引出，因为有...问题，我们做了下面这些工作

The workflow of this paper is shown in Figure 1.

2 Assumptions and Justifications

- **We only consider the problem on a two-dimensional plane.** Because the vertical movements of players in a football match has negligible influence on the match.
- **We treat each player as a mass point in a match**, since all we focus is their position in the match, their division of labor and their connection to each other.

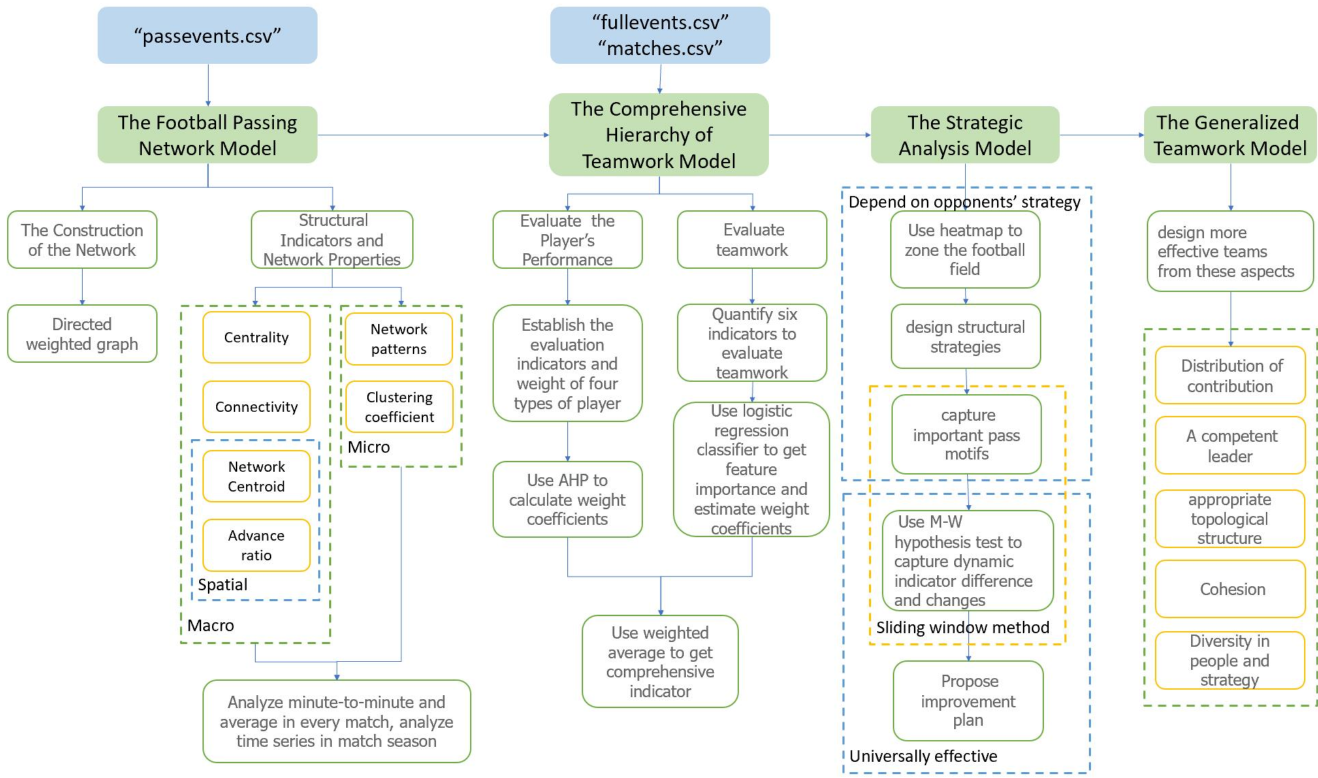


Figure 1: The workflow of this paper

- **The data provided in the topic are true and reliable to a certain degree.** Because the model we built is based on the data provided in the topic, only the high validity of the data can guarantee the high reliability of our model. 用了线性加权、线性回归就要有一条线性假设

3 Notations Description

The primary natations are shown in Table 1.

4 Model Construction and Solution

4.1 The Football Passing Network Model

Taking the first match in the last season as an example, the data for Huskies of the first match in dataset "passingevents.csv" is used to show the football passing network. Meanwhile, all the explanations of the structural indicators and the network properties are also based on the first match.

4.1.1 The Construction of the Network

We consider football players in a football team as nodes and collect all players in this team together to make up the set $V = \{v_i\}_{i=1}^{n}$. If player i passes the football to player j, then an edge from node i to node j generates. All the edges make up the set $E = \{e_j\}_{j=1}^{n}$, , and the edges that

Table 1: Variables

Symbols	Description
L	the average shortest path length
DC_i	the degree centrality of node i
$DC(Fig.1)$	the graph degree centralization of Figure 1
(x_i, y_i)	the coordinates in football passing network
(x_0, y_0)	the network centroid
AR	the advance ratio
$C_w(i)$	the weighted clustering coefficient of node i
C	the clustering coefficient of the network
CR	the consistency ratio
D_{ij}	the topological distance between node i and node j
w	the characteristic coefficient in the logistic regression
b	the characteristic intercept in the logistic regression

Coeff,全局变量，全局要用到，要有明确的指代性
局部变量，要","后面加解释
函数，自变量和因变量要标注清楚
单位unit

attach to node i make up set $N(i)$. Based on the dataset, we build a complex network including n=14 nodes in total as shown in Figure 2.

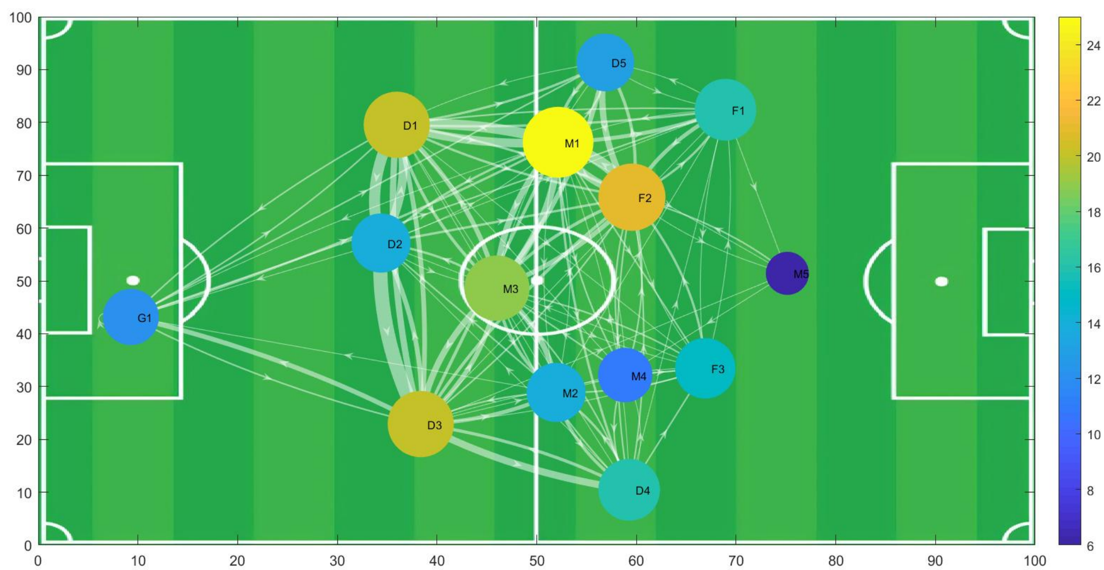


Figure 2: The football passing network of the Huskies for match 1

The standard court is scaled down. Both x-coordinate and y-coordinate are in the range [0,100]. Additionally, Each circular node in the network corresponds to one player. The total number of nodes means that the number of players that play in the first match. Since nodes D5, M4 and M5 are substituted during the match, the total number of nodes is 13 which is larger than 11. The coordinates of the node are the average coordinates of all passes made by one player during this match. The size and color of the node represents the number of elements in $N(i)$ The size of the node is positively correlated with the number of passes made by the player. The color and the ruler indicate the specific passing times. The arrow on edge represents the direction of passes. The width of the edge between node i and node j depends on the number of passes between them. The more passes occur, the wider the edge is. The playing time also affects the width.

4.1.2 Structural Indicators and Network Properties

In order to analyze the properties of our football passing network, we propose structural indicators from two main aspects that are in one match (take the first match as an example) and the whole season. In the first match, we focus on network connectivity, centrality distribution and spatial features in macro (teamwork) level and network patterns and local clustering coefficient in micro (pairwise) level. During the whole season,we concern the total number of passes and the variation of centrality distribution.

I. In one match

- **Macro Level: Network Connectivity**
  Network Connectivity indicates the correlation between tew nodes in the network. It is quantified by average shortest path length

$$L = \sum_{s,t \in network} \frac{d(s,t)}{n(n-1)}$$

  where s and t are two nodes in our network. d is the topological distance that the ball must go through to connect any two nodes.

- **Macro Level: Centrality Distribution**
  Centrality is the importance of a node in a network. More elements in $N(i)$ means a more important position in the network. In our network, centrality is affected by the number of passes made by a player. Centrality distribution is the gap between players in the team. We apply three indicators to define the centrality distribution of the team.
  Degree Centrality: Degree is a significant concept in network theory. A naive idea to measure the local importance of the node is in degree and out degree of the node. Similarly, getting the ball and giving the ball are equally valuable when we measure the importance of a player. In other words, we use the number of passes to measure the local importance of a football player. Degree centrality of node i is created as follow:

$$DC_i = \frac{(indegree_i + outdegree_i)}{n}$$

  where n is the total number of edges in the football passing network.
  Eigen Centrality: The basic notion of eigen centrality is to regard the importance of a node as a function of local importance of its adjacent node. To be specific, the more important the player, who is directly connected to the local important player is, the greater the eigen centrality of the local important player is. Eigenvector centrality computes the

eigen centrality based on the centrality of its adjacent nodes. Eigenvector centrality for node i is the i^{th} element of the vector x defined by the equation

$$Ax = \lambda x$$

where A is the adjacency matrix of our football passing network with eigenvalue λ. By virtue of the Perron-Frobenius theorem, there is a unique solution x, all of whose entries are positive, if λ is the largest eigenvalue A.

Graph Centrality Dispersion: The definition of centrality on the node level can be extended to the whole graph, in which case we call it graph centralization. The graph centrality dispersion is the distribution of graph centrality. In other words, graph centrality dispersion reflects the difference of importance between players. The smaller the graph centrality dispersion is, the closer contributions they make. Let $v*$ be the node with highest degree centrality in Figure 1. Let $X = (Y, Z)$ be the Y-node connected Figure 2 that maximizes the following quantity with $y*$ being the node with highest degree centrality in X. Then, we define

$$H = \sum_{j=1}^{Y}[DC(y*) - DC(y_i)]$$

Graph degree centralization of Figure 2 is as follows:

$$DC(Fig.1) = \frac{\sum_{i=1}^{V}[DC(v*) - DC(v_i)]}{H}$$

Correspondingly, other types of graph centrality are acquired. The graph cantrality dispersion is computed.

Ruselt: In the first match, the graph centrality dispersion of the Huskies and its opponent is around 0.14and 0.17, respectively. This means that the teamwork of the Huskies is beteer since the importance of each player and the contributions of each player is more similar.

- **Macro Level: Spatial Features**

  As for the spatial features, we propose two indicators: network centroid and advance ratio.

  Network Centroid: The advange coordinates (x, y) of all passes made during the match difine the network centroid (x_0, y_0):

  $$(x_0, y_0) = (\frac{\sum_{i=1} x_i}{N}, \frac{\sum_{i=1} y_i}{N})$$

  where (x_i, y_i) is the position of a pass. The network centroid in x-coordinate reflects the state. For the Huskies, the network centroid in x-coordinate between [0,50) means trend of defensive state while it between (50,100] means trend of attack state. It is inverse for the opponent 1.The network centroid in x-coordinate around 50 indicates no preference for any of the sides of the state. More attack states means that the probability of success is larger.

  Result: Figure 3 is the statisitc of the passing position in the first match. The real-time state is directly shown. By calculating the network centroid, the Huskies shows the attack state in the first match. The network centroid of opponent 1 proves the logic that opponent 1 adopt

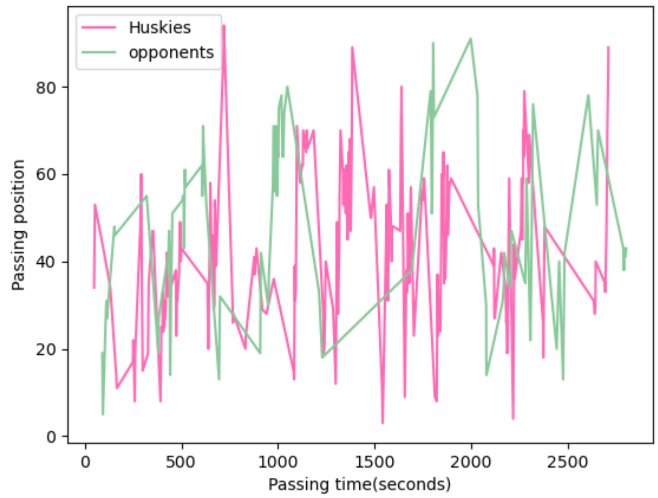


Figure 3: The passing position for the Huskies and opponent 1 in the first match

the defensive strategy, correspondingly.

Advance Ratio: Average ratio between the passing distance parallel and perpendicular to the opponent's goal is created as the direction of the passes of a team.

$$AR = \frac{\Delta y}{\Delta x}$$

where $\Delta y = y_2 - y_1$ of a pass is the difference between the y-coordinates at the fnal y_2 and initial points y_1 of a pass,while Δx is defned, accordingly, for the x-coordinate. If $\Delta x = 0$, then $AR = \Delta y$.

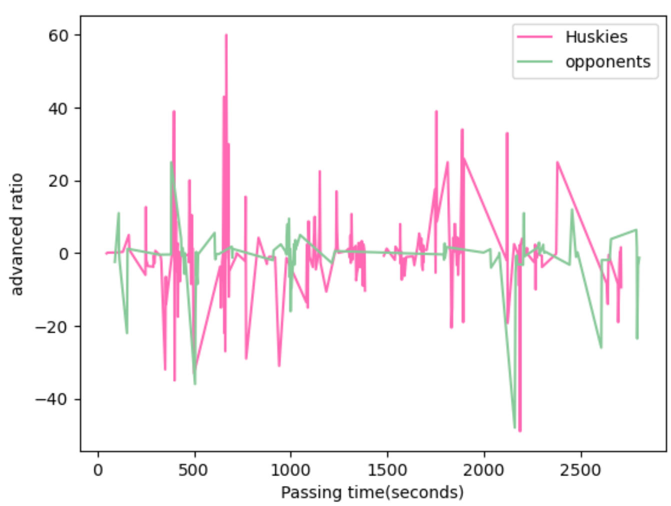
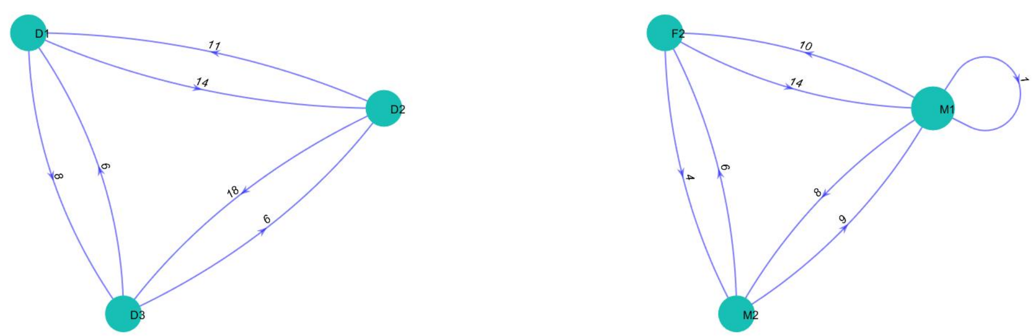


Figure 4: The passing direction for the Huskies and opponent 1 in the first match

Result: In Figure 4, we observe that has a ratio of advance much higher than the opponent 1, which reveals that passes are more parallel to the opponent's goal than the opponent 1. AR indicates that the Huskies is not concerned about advancing directly towards the goal, but on moving the ball in parallel, probably to fnd the most adequate moment to advance.

- **Micro Level: Network Patterns**
  Dyadic configuration refers to the relationship between two nodes and triadic configuration refers to the relationship between three nodes. In our football passing network, these configurations are subnetworks where occur frequent passes between two or three teammates. To acquire the subnetworks and find the interactions between teammates, we use network motifs [5]. Network motifs are subgraphs that frequently occur in a complex graph. The frequency of occurrence is the importance of the motif. We use Z-Score statistics to quantify the importance of each motif. Since the relationship in network motifs are undirected, retrofits including directed and weighted are made.



(a) (b)

Figure 5: The dyadic and triadic configuration of the football passing network

Result: Figure 5 (a) shows the dyadic configuration between D2 and D3 on account of the most passing time between them. Figure 5 (b) shows the triadic configuration between M1, M2 and F2. Additionally, we only take the topological distance in consideration instead of the actual distance.

- **Micro Level: Clustering Coefficient**
  The clustering coefficient of a node express the possibility that its adjacent nodes are connected to each other. A higher value of the clustering coefficient is an indicator of robustness [6]. This means that random deletion of a node or edge in the network has little impact on the nectwork connectivity [7]. In other words, one player's error has little impact on the passes in the team.
  The weighted clustering coefficient of node i, which measures the likelihood that adjacent

nodes of node i are connected mutually is

$$C_w(i) = \frac{\sum_{j,k} w_{ij} wjk w_{ik}}{\sum_{j,k} w_{ij} w - ik}$$

where where j and k are any two nodes in the network and w_{ij} is the number of edges between node i and node j. Finally, the clustering coefficient of the whole network is

$$C = \frac{1}{N} \sum_{i=1}^{N} C_w(i)$$

II. The Whole Season

- **The Total Number of Passes**
  The total number of passes is the most basic indicators to indicate the total number of edges n in our network. In a football match, the higher number of passes of a team means that the higher percentage of ball possession. The basic of goal is that players in team possess the ball. Generally, the higher a team's ball possession rate, the better the team's performance in the match.
  Result: Figure 6 indicates the number of passes of the Huskies and its opponent in each match in the last season. It is obvious that the number of passes in a team is antagonistic.

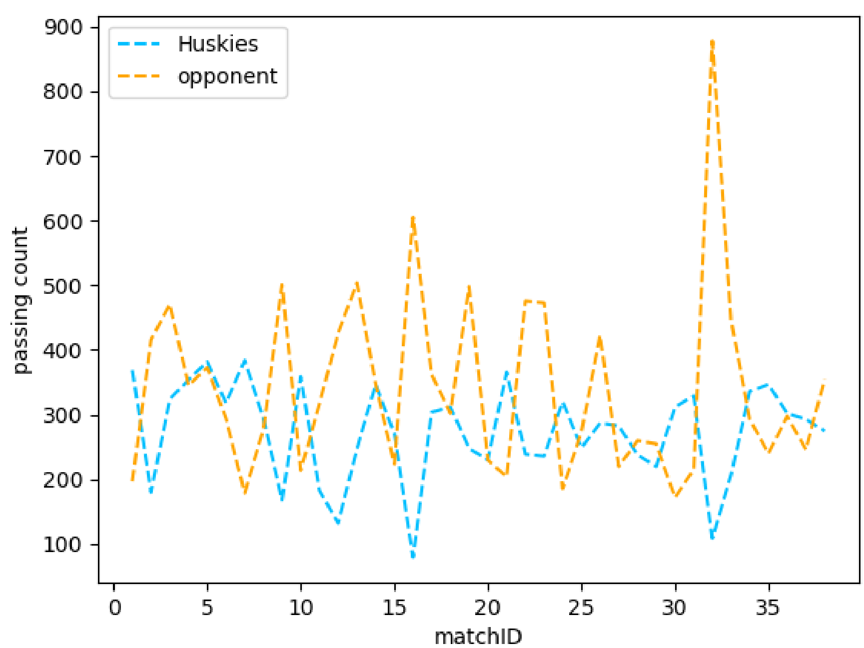


Figure 6: The total number of passes the Huskies and 38 opponents in the last season.

- **The Variation of Centrality Distribution**
  The Centrality is define in the "Macro Level: Centrality Distribution" above. However, we focus on the variation of centrality in the last season in this section. This variation shows the cooperation when the Huskies face different opponents.

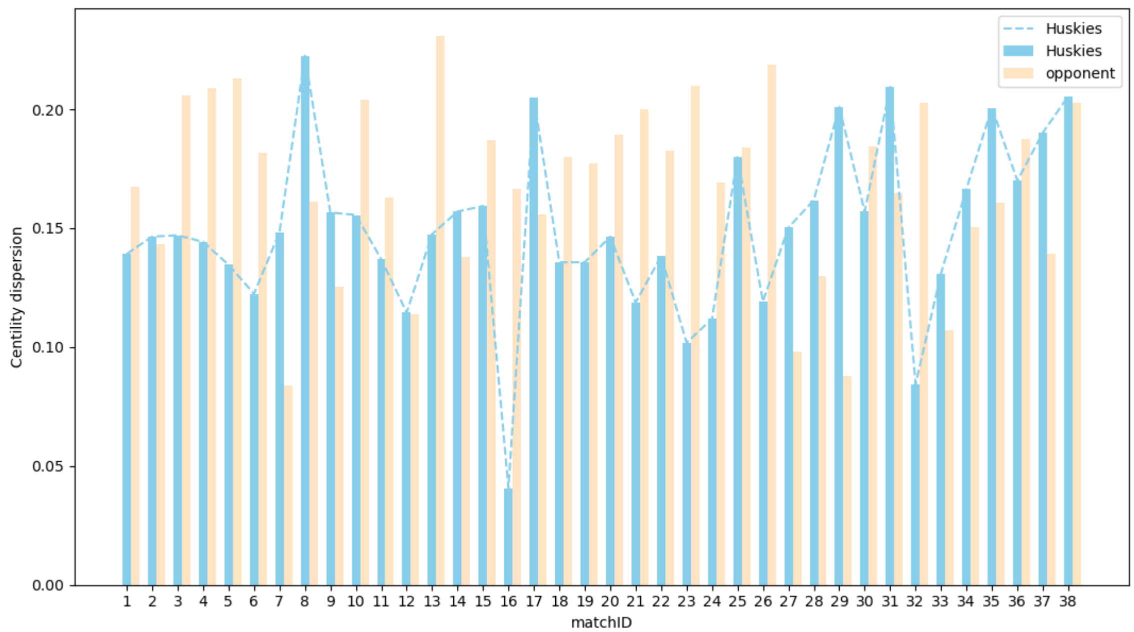


Figure 7: Centrality dispersion of the Huskies and 38 opponents in the last season.

Result: Figure 7 shows an obvious oscillation in the last season.

4.2 The Comprehensive Hierarchical Analysis of Teamwork Model

Good teamwork depends on the advantage of each player and the perfect cooperation between players. This is obviously hierarchical. Therefore, we build a comprehensive hierarchy of teamwork model. Firstly, we divide players into four categories that is forwards, midfields, defenders and goalkeepers. Several indicators are set to measure ability for each type of player. The player's ability can be extended to the team level by calculating the weighted ability. The following step is to quantify five team indicators: contribution degree, adaptability, flexibility, tempo and flow ability. Finally, we receive the comprehensive indicator by calculating the weighted average of all indicators. The weight is determined by the characteristic coefficients of the logistic regression model trained by the winning and losing matches

4.2.1 Data Pre-processing

First of all, we analyze the data in "fullevents.csv".

- The final coordinates of the event type shot is either (0,0) or (100,100). These coordinates are invalid since whether goal or not cannot be confirmed.
- The initial and the final coordinates of save attempt are either (0,0) and (100,100). These data are invalid.
- Due to the low data dimension and the independent occurence of various event, Principal Component Analysis and Analytic Network Process are not used.

4.2.2 The Player's Ability Model

After analysing the conditions which contribute to the goal, we propose six indicators to measure a player's ability showing in Table 2.

Table 2: The ability indicators of a player

单位，无量纲就画斜线

Indicators	Description
γ_1	the number of smart pass per minute
γ_2	the number of duel per minute
γ_3	the number of shot and the accuracy of shot per minute
γ_4	the number of acceleration per minute
γ_5	the number of clearance per minute
γ_6	the number of touch per minute

Smart pass is a peneteative pass that attemts to break the opponent's defensive lines to gain a significant advantage in attack. The accuracy is defined by $goal/shot$. Clearance is a defensive action where a player kicks the ball away from their own goal with no intended recipient of the ball. Touch is the ability of controlling the movement of the football.

Analytic Hierarchy Process (AHP) are used to determine weight of different indicators. We set up a matrix to calculate a list of the relative weights to measure the importance of the indicators. Normalization is a dimensionless process for indicators.

$$\gamma_{ij}^{'} = \frac{\gamma_{ij} - min(\gamma_i)}{max(\gamma_i) - min(\gamma_i)}$$

where γ_{ij} is the i_{th} row and the j_{th} column of the matrix. The quality of the weights are then measured by consistency ratio

$$CR = \frac{CI}{RI}$$

where $RI = 1.24$ when $n = 6$. If the $CR < 0.1$, then the judgements are trustworthy.

Since the responsibility of each type of player is different, we divide the players into four categories: goalkeeper (G), forward (F), midfielder (M) and defender (D). The responsibility of the goalkeeper is different from those of other players, therefore, we firstly consider the ability indicators for the players except the goalkeeper. Moreover, different responsibility result in different weight on indicators for different types of player. Forwards are the players in a team who play nearest to the opponent's goal and are therefore principally responsible for scoring goal. γ_3, γ_4 and γ_6 have larger weight. Midfields are the players whose usual positions are in the midfield, behind the forwards. γ_1, γ_4 and γ_6 occupy an large weight. Defenders are players who attempt to prevent the opponent from scoring. γ_2, γ_4 and γ_5 are put on larger weight. Therefore, the ability indicator of forwards are

$$I_F = w_{F1} \times \gamma_1 + w_{F2} \times \gamma_2 + \cdots + w_{F6} \times \gamma_6$$

where w_{Fi} is the weight of indicator γ_i for forwards. Correspondingly, I_M and I_D are the the ability indicator of midfields and defenders, respectively.

The Goalkeepers' only responsibility is defend the goal. Therefore, we separately set four indicators for them, namely the number of duel γ_2, clearance γ_5, pass γ_7 and save attempt γ_8 per minute. The following processes are the same and the final ability indicator is I_G.

4.2.3 The Successful Teamwork Model

Based on players' performance, the performance of the team is measured. We propose five team indicators, namely contribution degree, Adaptability, flexibility, tempo and flow ability.

- Contribution Degree β_1: Centrality Distribution
  As we define in the football passing model, centrality distrubution is centrality dispersion DC. This nagative index indicate that the smaller the value is, the more even each player's contribution is.

- Adaptability β_2: Clustering Coefficient
  As mentioned above, clustering coefficient$C$ indicates the robustness of the network. Therefore, it is applied to measure the adaptability of teamwork.

- Flexibility β_3: The shortest Topological Distance
  In order to measure flexibility, we consider the shortest topological distance that is the football must go through to connect any two nodes. Since our passing network is weighted, which means that the number of edges between nodes is different), we take different weights of the edges into account. Hence, the higher the weight is, the shorter the topological distance between two nodes is. The topological distance between node i and node j is defined as

  $$D_{ij} = \frac{1}{e_{ij}}$$

  where e_{ij} is the number of edges between node i and node j.

- Tempo β_4: Statistical Analysis
  Tempo is defined as the frequency of events during the match. During the equal period of time, the equal number of each event means that an even peace of match. The first half of the first match for the Huskies is taken as an example to explain the tempo. We choose the four events which have the high frequency and can indicate behaviors, namely pass, duel, freekick and shot. Then the freauency histograms are shown in Figure 8. It is clear that the peak of pass and shot frequency is basically the same, indicating that the Huskies has a fierce attack in this period. In addition, there is always duels before the Huskies's attacks. Therefore, the tempo is fast.

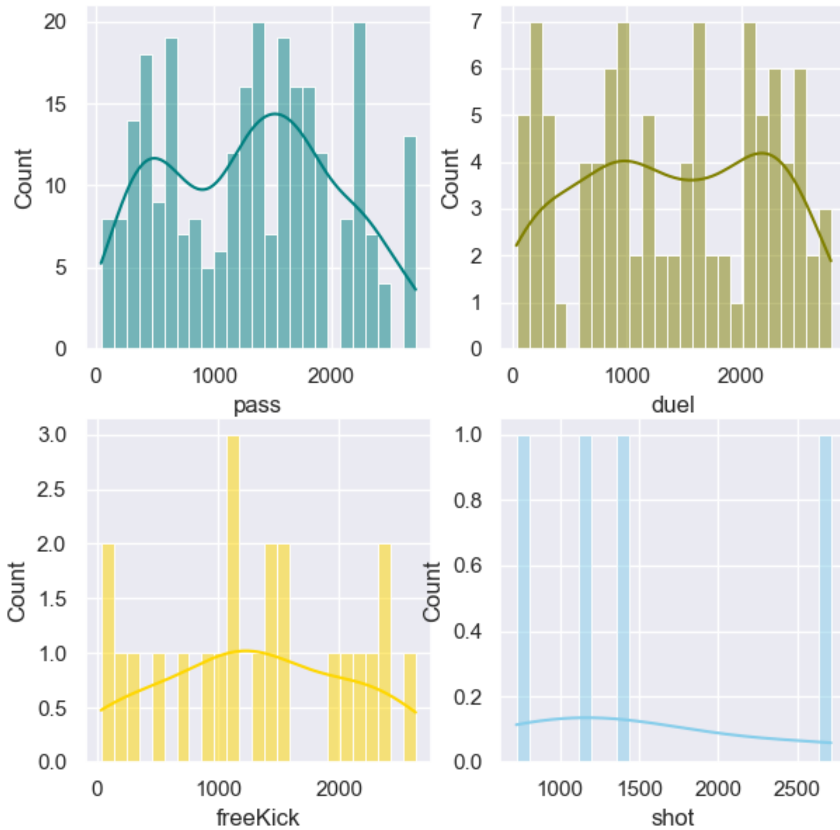


Figure 8: The frequencies of pass, duel, freekick and shot

- Flow Ability β_5: Position and Advance Ratio
  The football passing network is dynamic during a match. Hence, flow is peoposed to describe that change. Flow ability is quantified by the football passing position which we mentioned in "Centrality Network" and advance ratio $\Delta y/\Delta x$.

Winning or losing reflects the success of team to a certain extent. Considering that the indicators that affect the winning of the match are important and all indicators are added linearly, this process coincides with the logistic regression. Therefore, we delete tie data and use winning and losing data to train logistic regression classifier. The characteristic coefficient **w** and intercept **b** we receive through trainning are

$$\mathbf{w}^T x + \mathbf{b} = ln\frac{P(Y=1|x)}{1-P(Y=1|x)}$$

$$P(Y=1|x) = \frac{1}{1+e^{-(\mathbf{w}^T x+\mathbf{b})}}$$

w is used as weights of ability and team indicators to output the comprehensive indicator.

4.2.4 Result: The Comprehensive Indicator for Teamwork

According to AHP, the weights for I_F, I_M, I_D and I_G are shown in Table 3.

Table 3: The weights for the ability indicators of players

	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8
I_F	0.0465	0.0830	0.4254	0.1333	0.0445	0.2673	-	-
I_M	0.1051	0.0307	0.0739	0.3003	0.0454	0.4446	-	-
I_D	0.0708	0.4284	0.0419	0.1407	0.2456	0.0726	-	-
I_G	-	0.1547	-	-	0.2523	-	0.1170	0.4761

The visualization of the ability of forwards, midfields and defenders is shown in Figure 9.

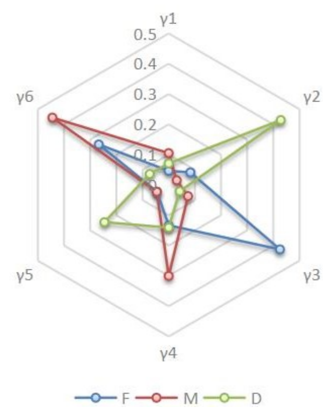


Figure 9: The ability of forwards, midfields and defenders

Finally, based on the characteristic coefficients of the logistic regression model, the weights of the comprehensive indicator are shown in Table 4.

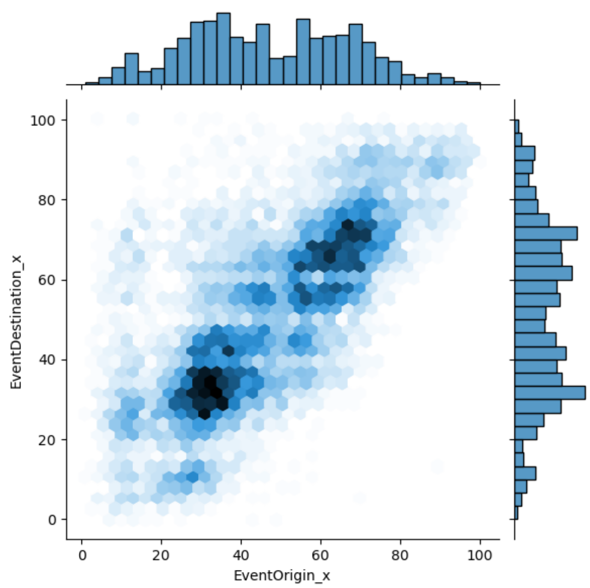
Table 4: The weights for the comprehensive indicator

	I_F	I_M	I_D	I_G	β_1	β_2	β_3	β_4	β_5
weight	0.0989	0.1194	0.1303	0.0404	0.1199	0.2120	0.0578	0.1272	0.0942

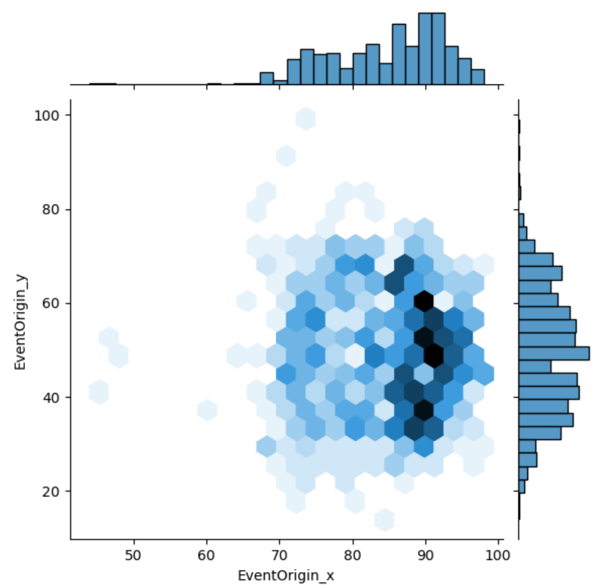
4.3 The Strategic Analysis Model

4.3.1 The Zone and Formation Model

We analyze the position of pass and shot which are two representative events and draw their distributions showing in Figure 10 . The darker the colour on the graph is, the higher probability the event occurrs. As Figure 10 shows, the Huskies passes frequently in the lower left and upper right of the court and excells in diagonal pass. Based on these heat maps, we divide the whole



(a) The distribution of the pass position



(b) The distribution of the shot position

Figure 10: The distribution of the pass and shot position

match into five zones: save zone, midfield, defensive zone, attack zone and shot zone. Shot zone is determined by the area that 70% shots occur.

In Figure 11, we propose 4-4-3 formation for the Huskies. Firstly, the player make the most of his strengths and avoid his weaknesses in the position that we arrange. The positions of the players we arrange are based on the ability indicators that we proposed in the player's ability model and the zone. After we analyse the ability of each player in the Huskies, 11 players are chosen and arranged to the position which matches them best. Additionally, this formation indicates the teamwork of the Huskies fully. This formation slightly tends to attack. The ability of forwards, midfields and defenders for the Huskies are basically equal and the ability of wingers is outstanding. The diagonal long pass can be realized easily when this formation is applied. Meanwhile, the acceleration ability which means stamina of our players is preeminent. This superior ability makes up for the disadvatage, a long back-up distance when defending of this formation. Furthermore, furthermore, this formation has several inherent advantages. It is flexible, therefore, can be changed to 4-2-3-1 and 4-3-2-1 formation which maximize the former and minimize the latter of the Huskies.

4.3.2 Strategies and Improvements

Passing Strategies

Several cooperations between players are received by identifing important motifs. The process is shown in Figure 12. We use sliding window to analyze the dynamic change of passing network over time. Since the number of pass in a half of the match is between 100 to 180, we set the sliding step to 20 passes and draw the passing network by 50 passes. Slide one step for every 5 passes analysis. By repeating the above steps, we analyse the dynamic change of the passing network and the influence of the previous pass on the current pass. The important and frequent passing motifs

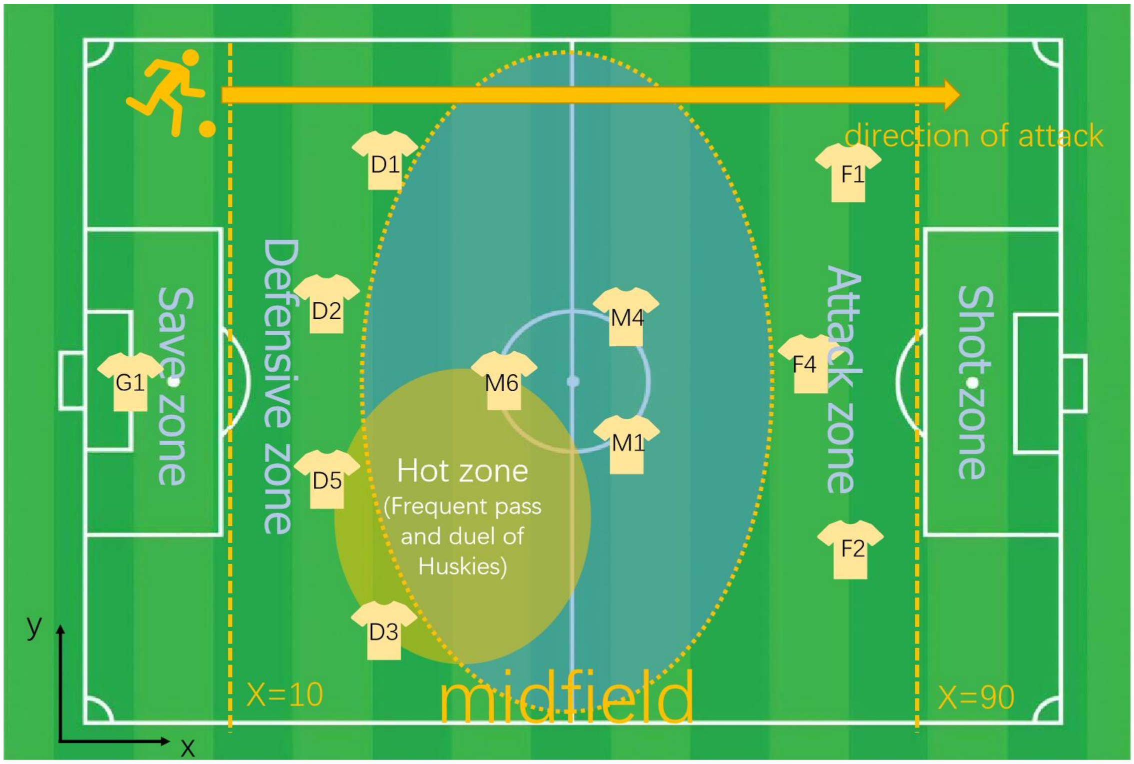


Figure 11: The zones and th strategy for the Huskies

are then acquired.

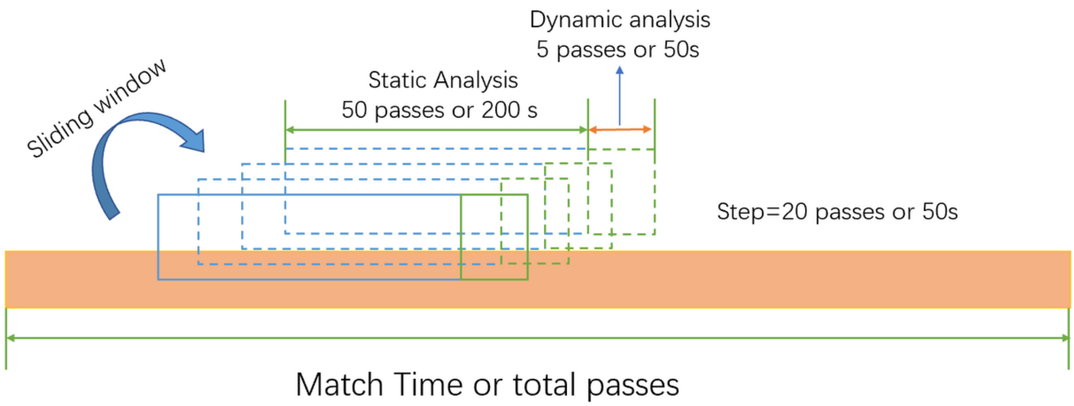


Figure 12: The process of sliding window

Ruslts: By appling the sliding window on the ninth match with maximum goals, we conclude three passing strategies showing in Figure 13. The second motif occurs most frequently (46.15%) and the frequency of third motif is 38.46%, the second largest.

Improvements

We use Mann-Whitney hypothesis test to analyze 8 teamwork indicators, showing the conclusion in Table 4. In Mann-Whitney hypothesis test, the validity of statistics does not required distributional

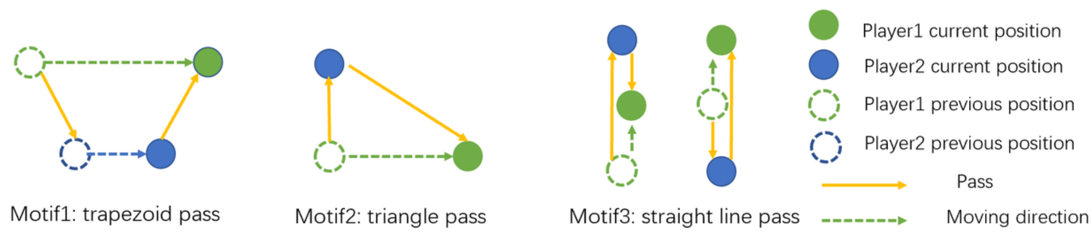


Figure 13: The important passing motifs

assumptionsand use sliding window to anlyse other indicators. In this problem, the distributions of indicators in 38 matches are far away from the normality.

Table 5: Mann-Whitney hypothesis test for ability and team indicators

	I_F	I_M	I_D	I_G	β_1	β_2	β_3	β_4	β_5
statistic U	1056	720	1080	555	242.0	369.5	246.0	315.0	313.0
p-value	0.003	0.689	0.001	0.040	0.013	0.718	0.017	0.209	0.198

We propose four changes based on the above graphs and analyses.

- Figure 6 indicates that the number of passes in the Huskies is less than its opponents' average passes. The number of passes is intimately related to the possession a team has and the result of a match. The Huskies may have flaws in their passing skills. As shown in Figure 10(a), there are frew passes along the back diagonal. Hence, the skills of passing, especially along the back diagonal should be improved in the Huskies.

- The advance ratio shown in Figure 4 illustrate the vertical movement of the Huskies is frrequent. Their main strategy is conservative, generally moving in parallel to opponent's goal to find opportunities to attack. Additionally, the p-value of tempo β_4, which is less than the average p-value inicates a slow tempo. Therefore, the Huskies can try to enrich their strategies, for example more offensive and fast tempo.

- In Table 5, the Huskies lack of flexibility β_3 and have an advantage on contribution degre β_1, compared with their opponents. Hence, the players in the Huskies have almost equal abilities and contributes to teamwork equally. However, the flexibility of the Huskies is supposed to be enhanced.

- The average values of I_F and I_M in the Huskies are higher than their opponents' average while the average values of I_D and I_G in the Huskies are dramatically lower. This means the abilities of defenders and goalkeepers are lower, resulting in a poor defence. Therefore, defenders and goalkeepers need targeted training.

4.4 The Generalized Successful Teamwork Model

Individual Contributions

In a football team, each player plays a positive role in his position and makes contributions as much as possible. The football team then can win the match. An efficient team needs each member's contributions in positive direction. The distribution of members' contributions is an indicators to evaluate the team performance.

Leader

In a football team, different coaches result in different formations and players playing in a match. The team performance and the result of the match may have significant difference. This is proved by analyzing the dataset “matches. csv”. An efficient team requires a leader with leadship and decision-making skills. He is able to assign tasks based on different members' features. The leadership of leader is an indicators to evaluate the team performance.

Cohesion

A football team with cohesion means that there are close connection between players and they collectively make efforts focusing on the goal. The cohesion affects the result of a match. Correspondingly, an efficient team should has a high level of cohension. The clustering coefficient, degree centrality and flow centrality is used to measure the cohension of a team.

Topological Structure

In a football match, the frequency of passes between players forms a weighted oriented graph. An efficient team has favourable topological structure. The accurancy of topological structure reflects the team performance.

Diversity

Diversity of members: In a football team, different types of players excel in different aspects and contribute in different aspects. Therefore, a efficient team requires members who are expert in different aspects. The extent of the coverage is an indicator.
Diversity of stategies: According to different opponents and real-time situation, football team adopt and adjust pass, attack and defendence stategies. Similarly, a efficient team is able to propose various strategies to cope with changes in the environment.

5 Model Test and Sensitivity Analysis

Model Test

The accuracy of losgistic regression classifier is 0.821 with I_2 regulation and inverse of regularization strength=5, and the correlation coefficient between the comprehensive indicator and score in match is 0.587. Therefore, the nine indicators can predict the result of the match to a certain extent. In addition, the comprehensive indicator is linearly correlated with the score of the match, which has great reference value for strategies and improvements.

Sensitivity Analysis

When regulation strength is constant, we adjust the weights w_1 to w_9 in the range of 0.02. The accuracy of losgistic regression classifier fluctuate in the range of 0.741 to 0.821, indicating that our model is relatively stable. Therefore, our modle passes the sentitivity test [8].

6 Conclusion

By constructing a passing network, extracting and analyzing metrics to evaluate player performance and team cooperation, we concluded that the Huskies' game strategy was conservative

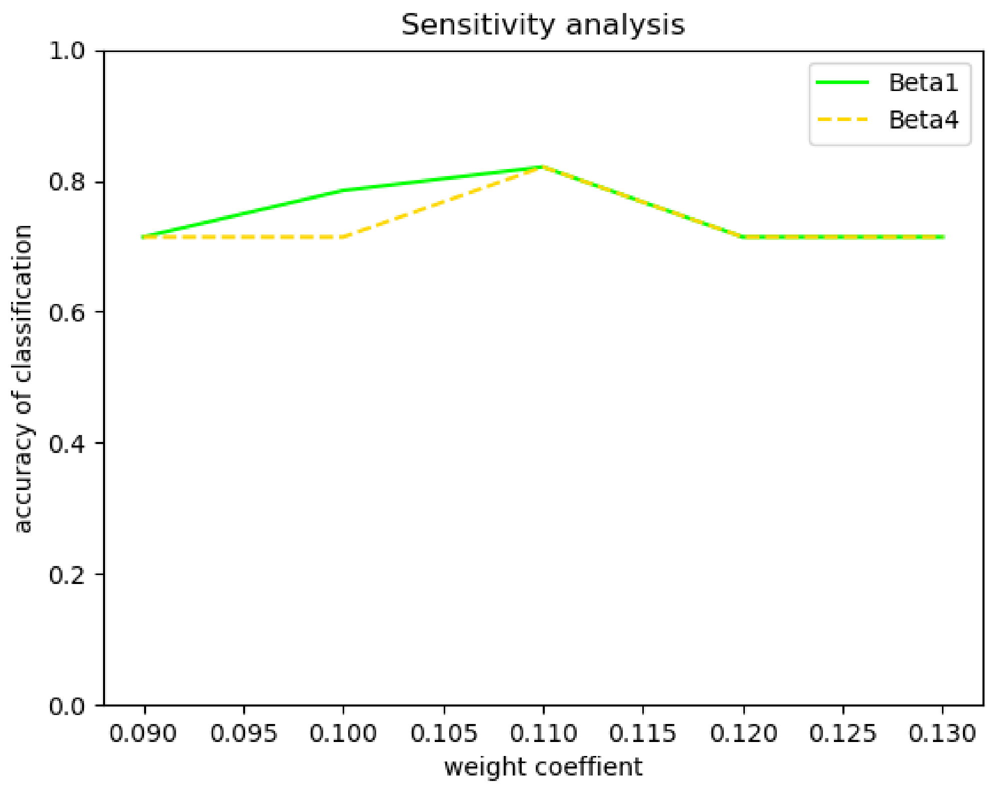


Figure 14: Sensitivity analysis of our model

and their cooperation was less flexible, shortcomings in passing led to a lower possession of the ball.Thus, we provide coaches with improvement strategies: strengthen training in passing in the opposite diagonal direction and make an effort to improve the ability of Defender and Goalkeeper. We also generalized universal teamwork model, concluding that building a successful team requires consideration of five aspects as follow: distribution of contributions, the ability of leader, the topology of teamwork, diversity and cohesiveness within the team.

7 Strengths and Weaknesses

7.1 Strengths

- To detailedly analyze our football passing network, we create the indicators in time dimension and space dimension, in which the space dimension includes micro and macro. Additionally, we apply sliding window and frequency analysis to analyze the dynamic characteristics of our network. Thereout, we propose improvements tailored to the Huskies's strengths and weaknesses.

- In order to dertermine the weights of indicators, we use losgistic regression classifier to estimate the importance of these indicators. This mothed not only overcomes the shortcoming of the entropy weight method that ignores the importance of the indicator itself, but also makes up for the subjectivity of AHP. The comprehensive indicators, we propose reflects the closely relation between the match result and the ability and level indicators. Moreover, it serves as a reliable reference for strategies and improvements.

7.2 Weaknesses

- Due to the limitation of data, we measure the ability of a player by the shooting average of a team instead of himself.
- Only simple linear superposition is considered when we calculate the comprehensive indicator for teamwork. Actually, there may be complex nonlinear relationship between some indicators and the score of the goal. Therefore, simple linear superposition cause a large error that the maximum accuracy of the logistic regression classifier is 0.741. Using nonlinear methods, such as decision tree will improve the accuracy.

References

[1] J. L. Pena and H. Touchette, "A network theory analysis of football strategies", arXiv preprint arXiv:1206.6904, 2012.

[2] P. Cintia, S. Rinzivillo, and L. Pappalardo, "A network-based approach to evaluate the performance of football teams", in Proceedings of the Machine Learning and Data Mining for Sports Analytics workshop, ECML/PKDD 2015, 2015.

[3] F. M. Clemente, M. S. Couceiro, F. M. L. Martins, and R. S. Mendes, "Using network metrics in soccer: A macro-analysis", Journal of human kinetics, vol. 45, no. 1, pp. 123–134, 2015.

[4] L. Pappalardo and P. Cintia, "Taca la bala says the wizard: a trip into the world cup 2014", http://bigdatatales.com/blog/2014/07/12/ taca-la-bala-says-the-wizard-a-trip-into-the-world-cup-2014/, 2014.

[5] Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U. Network Motifs: Simple Building Blocks of Complex Networks. Science. 2002; 298(5564): 824-827.

[6] Deng Guishi, Wu Peijian, Tian Wei. Robustness and Vulnerability of Global Shipping Network [J]. Journal of Dalian University of Technology, 2008, 48(5): 765-768.

[7] D. J. Watts and Steven Strogatz (June 1998). "Collective dynamics of "small-world' networks". Nature. 393 (6684): 440–442. Bibcode:1998Natur.393.440W. doi:10.1038/30918. PMID 9623998

[8] Zhang Xiaolin, Wang Chongming. New Advances of Research on Group Perforamce and Team Effectiveness[J]. Applied Psychology, 1997(02): 58-64.