Problem Chosen	2020 MCM/ICM Summary Sheet	Team Control Number
D		**2002118**

MoNGO: a detailed study of *Huskies*

Summary

Teamwork is believed to outperform individual efforts and a sequence of additive contributions of teammates with meticulous design and good strategies. As a representative of teamwork, sports, specifically football game, has gradually grabbed public attention.

To help team *Huskies* understand their team dynamics and further make approriate strategies, our paper provides a novel framework MoNGO (Modified Network Model with Group Dynamics and Oppponent Analysis) to fully evaluate different aspects from several scales, both spatially and temporally, which can be briefly inferred from the name *MoNGO*. Based on these, we provide strategies for *Huskies* in the hope that they can exprience an increase in winning rate.

We first analyze dynamic team performance constructing a network between players. With the knowledge of **Network Science** as well as **Graph Theory**, we develop a **Naive Network Model** to form a simple indicator *C-indicator* on account of passing behaviour among players to measure the cooperation.

Considering that passing behaviour cannot fully depict the degree of cooperation of a team, let alone providing strategies. Therefore, we propose MoNGO to make for the flaws. As mentioned before, MoNGO mainly consists of three parts which will be elaborated in the following paragraphs.

Modified Network Model aims at giving an indicator of the team. The indicator, **Rank(List)** is a sum of three parts: *PlayeRank* (measurement for individual ability), *TeamRank(List)* (measurement for global tactics) and *PartRank(List)* (measurement for local tactics). Besides, we used **GM** (Grey Prediction) to help evaluate physical ability for each player.

Group Dynamics Condition examined the leading coach, matching environment and member relationship for the team.

In Opponent Analysis, we design a score to represent the playing level of each opponent team and make a statistical result.

Then we use the framework to give strategies relating to the best squd and global and local tactic deployments. Also, we make suggestions about what squd and tactic to use facing different opponents.

In the end, we make sensitivity analysis and discuss strengths and weaknesses.

Keywords: Network Science; Graph Theory; Ranking System; Group Dynamics; Team Cooperation Strategy

MoNGO: a detailed study of $Huskies$

February 17, 2020

Contents

1 Introduction

1.1 Problem Background

This is the worst of times, with growing social ties comes more serious challenges; This is the best of times, with the development of technologies such as network science enabling us to scientifically organize groups of people from different fields and perspectives to solve problems.

It is universally acknowledged that teams equipped with smart strategies, ideal leadership and strong cohesion can usually outperform individual efforts and extra contribution of every member of the team, which is vividly demonstrated in case of team sports.

Figure 1: The 2009-10 season saw Barcelona as one of the best football teams in history. It is its superior teamwork and style of football that keeps it far ahead of other teams. [1]

Thanks to automated or semi-automated sensing technologies that provide high-fidelity data streams extracted from every game, we are now able to gain an insight into the understanding of certain of soccer team performance. We are provided with an abundance of data to evaluate and help home soccer team *Huskies* in future games.

However, challeges still exists and is stated as the following, and it is indeed what our paper is for:

1. How to construct and explore the passing network of players?
2. How to extract and analyze the data reagarding the team's staffing and tactical strategy from the data stream?
3. How to develop strategies based on the information obtained above to improve the team's winning percentage on the field?

1.2 Our Work

In this paper, we propose a novel framework mainly composing of three parts: Modified Network Model which gives a measurement about the degree of cooperation, Group

Dynamics Conditions and Opponent Analysis. And based on these, we give our stratagies.

The rest of the paper is organized as follows. In Section 2, we state the assumptons and symbols used in this paper. Section 3 provides sufficient details about our model. Section 4 carrys out result and analysis about our proposed model. Section 5 presents detailed stategies we give to *Huskies*. At last, we further study our model in Section 6 and make some conclusions in Section 7.

Our workflow can be briefly presented as follows:

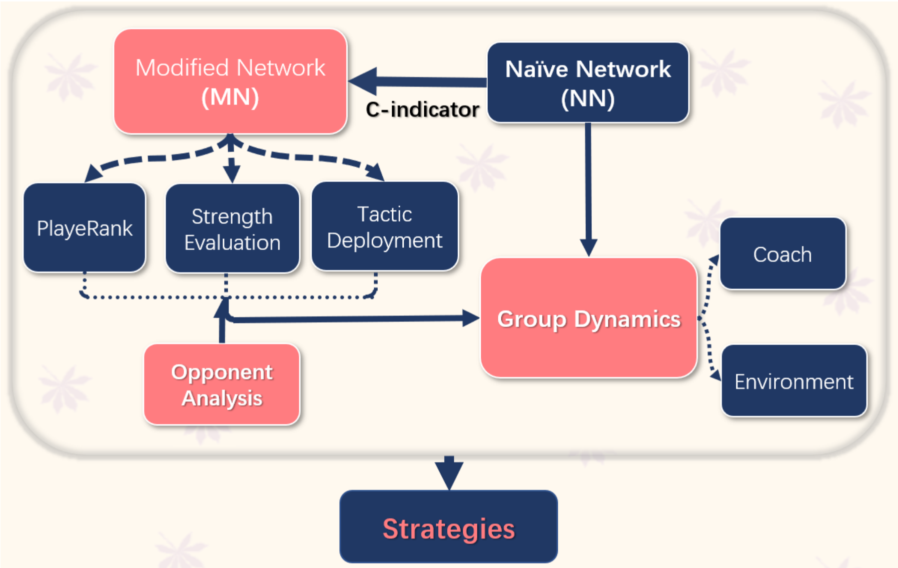


Figure 2: workflow of our team

2 Assumptions and Symbols

2.1 Assumptions

Our model makes the following assumptions:

1. **There are no injuries in each and every one of the matches.** The assumption just means that the status of all players only change with time.

2. **Factors of substitution during the match.** In our model, substitution is related with physical decline. We assume that substitutions are only used to replace players who have lost a lot of strength.

3. **The first half of the game is fixed for five minutes of supplementary time.** Since the specific end time of each half game is not determined, we assume for the purpose of approximation that the first half of the game is fixed for a total of $60 * (45 + 5) = 3000s$, and the second half of the game is represented by the time of the last event, i.e. the actual time we know from the event chart.

4. **Omitted data that indicates a player passing balls to himself,** since we cannot explain such situation.

5. **Squd of 4-4-2, i.e. 4 defenses, 4 midfields and 2 forwards, is the default squd.** This is based on the result of our analysis of regular squd.

6. **The time that player A and B are present at the same time in this game is represented by the overlapping part of the maximum time span of the events that A and B participate in.**

2.2 Symbols

In this paper we use the nomenclature in Table 1 to descibe our model. Other symbols that are used specifically in one model will be stated later.

Table 1: Symbols used in the model

Symbol	Definition
P_0	Number of times passing soccer to others
P_d	Number of times receiving soccer from others
P_c	Number of times grabbing soccer when passing
$P_{(A,B)}$	Passing soccer between menber A and B
$r_{P_0}(A)$	The rate of passing socccer of player A
$r_{P_d}(A)$	The rate of receiving socccer of player A
$r_{P_c}(A)$	The rate of grabbing socccer during passing player A
$r_{P_{(A,B)}}$	The rate of passing soccer between a pair A,B
$t_{P_c}(A)$	The average middle ball-pocession time for member A
$\sum t$	The total time for a player participates in matches
$System(A,B)$	Reflection of the number of $r_{P_{(A,B)}}$
$System(A,B,C)$	Reflection of the number of $r_{P_{(A,B)}}$ among three player

3 Statement of our Model

In this section, we will discuss all details about the three component separately.

After finding a best squd using only passing network, we further proposed an more complicated model which consists of three parts, **player-ranking**, **tactic development** and **player strength evaluation**. Along with group dynamic factors (coach-choosing, environment and member relationship) and opponent analysis forms our whole system.

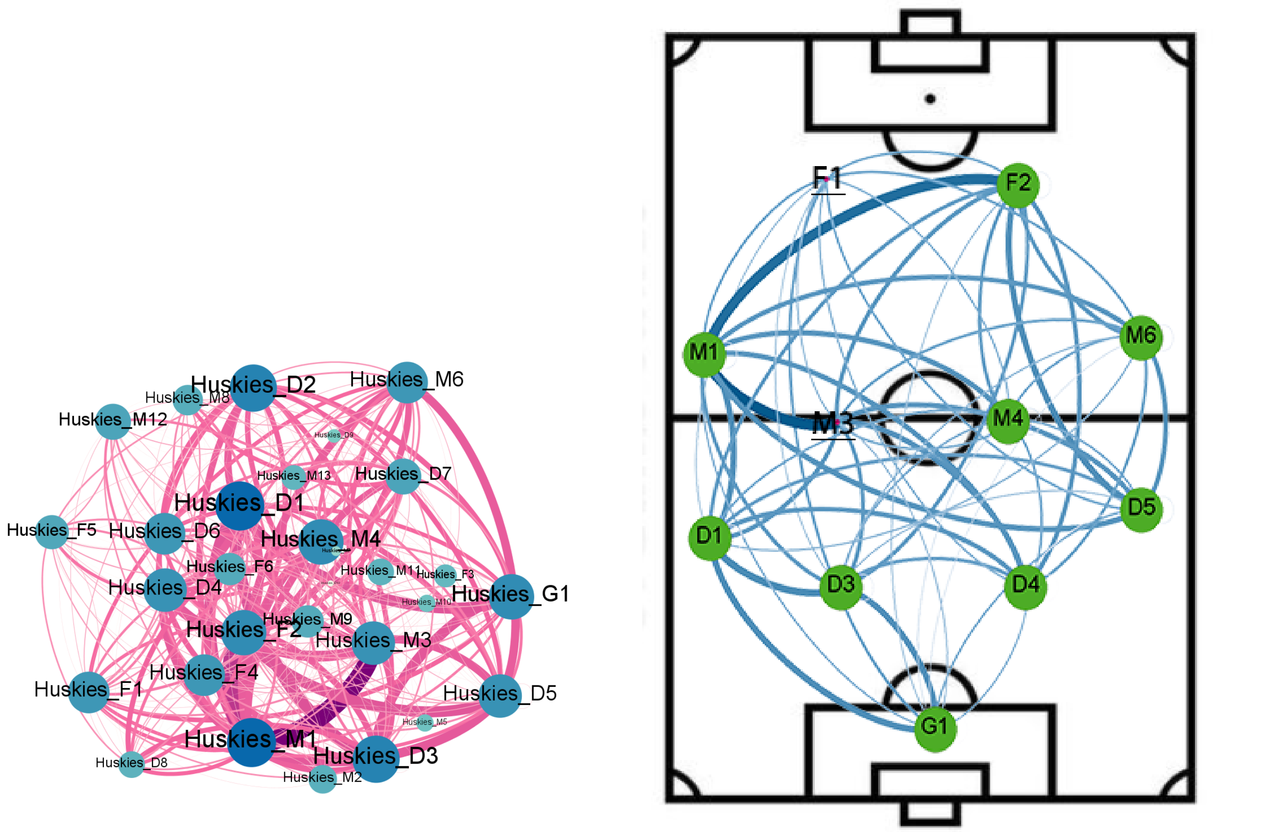
3.1 Naive Network Model

This model is based on graph theory. We formalize every team member as a vertex and each passing behaviour adds to an edge between two vertexes. [2]

Using data provided, we can obtain the following two figures with the help of *Gephi*. Figure 3a representes the situation of passing soccor among all the 30 players in *Huskies*. Figure 3b illustrates the passing situation of the regular squd 11 players of the 30 ones.

As for how to get the 11 players, we obtained the total number of events (both active and passive) that each member of the team had participated in during all events, by traversing all events of all 38 events. This data, we identified **the most active** (although the goalkeeper G1 is not one of the most active 11 people, but due to the uniqueness goalkeeper identity, must be in playing 11 people, so the substitute goalkeeper for the original 11th) in the 25 people team network 11 players, as to establish a list of players passing network.

According to symbols defined in section 2:



(a) Network for the 30 players in *Huskies*.

(b) Network for the most active 11 players and their position on a football field map

Figure 3: Passing behaviour of players during the whole season. The size of the vertex stands for the involvement of passing process and the thickness of edges symbols the number of times of passing soccor.

$$r_{P_0} = \frac{\sum_t p_0}{\sum t}, r_{P_d} = \frac{\sum_t p_d}{\sum t}, r_{P_c} = \frac{\sum_t p_c}{\sum t}, r_{P_{(A,B)}} = \frac{\sum_t p_{(A,B)}}{\sum t}, t_{P_c} = \frac{\sum_{pc} t}{P_c}$$

The model can be simply viewed as an undirected weighted graph, where the degree of each vertex indicates r_{P_0} and r_{P_d}, the weight of each vertex indicates $r_{P_{(A,B)}}$.Let $List = A, B, ...$ denote the 10 players except of the goal keeper. For the sake of clarity, we also listed the five additional definitions needed in this model.

Definition 1. *Dyadic Configuration: When $r_{P_{(A,B)}}$ achieves a threshold. Denote $System(A, B) = 1 \Leftrightarrow r_{P(A,B)} \geq r_1$ where r_1 is a parameter to be determined.*

Definition 2. *Triadic Configuration: When $r_{P_{(A,B)}}$ of pairs between three members achieves a threshold. Similarly, denote $System(A, B, C) = 1 \Leftrightarrow \min\left(r_{(P(A,B))}, r_{(P(B,C))}, r_{(P(A,C))}\right) \geq r_2$ where r_2 is another parameter to be determined.*

Definition 3. *Special Collaboration $PointE(List)$: A weighted sum of $System(A, B)$ and $System$ (A, B, C), i.e. $\sum_{A,B \in List} System(A, B) * K_{dya} + \sum_{A,B \in List} System(A, B, C) * K_{tri}$ where K_{dya} and K_{tri} are parameters to be determined.*

Definition 4. *Induvidual Score $PointP(A)$: A weighted sum of r_{P_0}, r_{P_c}, r_{P_d} etc. To be specific, $PointP(A) = K_{P_0} r_{P_0}(A) + K_{P_d} r_{P_d}(A) + K_{P_c} r_{P_c}(A) + K_t * t_{P_c}(A)$, where we introduce four parameters decided by ranking system only.*

Based on the above definition, we can obtain an indicator called *C-indicator* to measure the degree of coperation of a team, which is the sum of **Special Collaboration** and **Individual Score**, i.e. $C - indicator = \sum_{A \in List} PointP(A) + PointE(List)$, squd with the greatest C-indicator is presumed to be the best squd.

3.2 Modified Network Model

Although in the first part, we built a model based on passing networks, and gives a review of the team, we only consider about the relevant data of passing. However, the fact of football much complicated than this, we need to consider the elements of far more than mere passing, i.e. shooting, confrontation, foul, offensive, etc. Based on this, we modify the model of simple passing network.

In this part, we introduce the concept of PlayeRank [3] from the perspective of individual players. Also, with the help of strengh evaluation, we will be able to get an insight into an important attibute of players, $strength$. Finally, a tactic-deployment mechanism can design local tactics and global tactics for the team. After combining the added elements with the previous passing network, we have a new model in this section, which is used as a basis for evaluating the squad, predicting the optimal squad, and providing several possible tactical plays.

For this model, the notations are listed as table 2.

Table 2: Additional symbols used in *PlayeRank*

Symbol	Definition
$s(A)$	Number of shots for player A
$d(A)$	Number of duels for player A
$f(A)$	Number of fouls for player A
$P_h(A)$	Number of high passes for player A
$P_s(A)$	Number of smart passes for player A
$d_a(A)$	Number of air duels for player A
$r_s(A)$	The rate of shots of player A
$r_d(A)$	The rate of duels of player A
$r_f(A)$	The rate of fouls of player A
$r_{P_h}(A)$	The rate of high passes for player A
$r_{d_a}(A)$	The rate of air duels for player A
$PlayeRankF(F_i)$	The PlayeRank for forword F_i
$PlayeRankM(M_i)$	The PlayeRank for midfield M_i
$PlayeRankD(D_i)$	The PlayeRank for defense D_i
$k_{P_s}(A)$	The rate of smart passes of player A
$k_{atk}(A)$	The aggressive degree of player A

3.2.1 Ranking the players: PlayeRank

According to the definitions listed in 2, we will have the following expressions.

1. $k_{P_s}(A) = \frac{P_s(A)}{P(A)}$
2. $k_{atk}(A) = \frac{\sum_P \triangle_y}{\sum_P |\triangle_x|}$, where $P \in pass|\triangle_y > 0$

The major framework of PlayeRank is listed as figure 4.

According to the definition, the score for forward, midfield and defense can be expressed easily.

1. $PlayeRankF(F_i) = K_{PRFs}s(F_i) + K_{PRFP_0}P_0(F_i) + K_{PRFP_d}P_d(F_i) + K_{PRFt_{P_c}}t_{P_c}(F_i) + K_{PRFd}d(F_i) + K_{PRFf}f(F_i)$

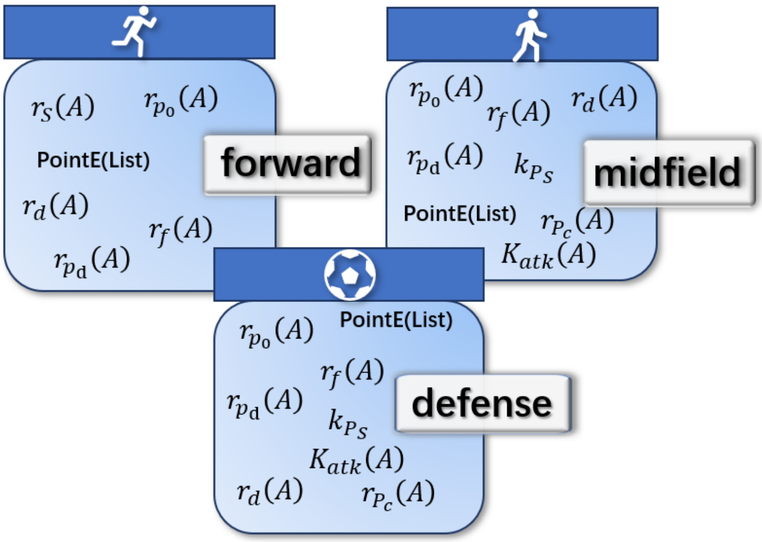


Figure 4: An illustration of the ranking system

2. $PlayeRankM(M_i) = K_{PRMP_0}P_0(M_i)+K_{PRMP_d}P_d(M_i)+K_{PRMP_c}P_c(M_i)+K_{PRMt_{P_c}}t_{P_c}(M_i)+K_{PRMk_{P_s}}k_{P_s}(M_i) + K_{PRMk_{atk}}k_{atk}(M_i) + K_{PRMd}d(M_i) + K_{PRMf}f(M_i)$

3. $PlayeRankD(D_i) = K_{PRDP_0}P_0(D_i)+K_{PRDP_d}P_d(D_i)+K_{PRDt_{P_c}}t_{P_c}(D_i)+K_{PRDk_{P_s}}k_{P_s}(D_i)+K_{PRDk_{atk}}k_{atk}(D_i) + K_{PRDd}d(D_i) + K_{PRDf}f(D_i)$

Where all the $K_x s$ are parameters to be determined. Therefore, one import indicator of this model which measures the individual ability can be expressed as:

$$PlayeRank(List) = \sum_{X\in(F,M,D)} \sum_{X_i\in List} PlayeRankX(X_i)$$

3.2.2 Physical Ability Evaluation

We have the set of additional symbols for this section in table 3.

Table 3: Additional symbols used in *PlayeRank*

Symbol	Definition
$Pass(A)$	Number of passes for player A during match i
$Duel(A)$	Number of duels for player A during match i
$Foul(A)$	Number of fouls for player A during match i
$P_A(A,i,t)$	The physical ability for A in match i at time t

Therefore, $P_A(A,i,t)$ can be defined as:

$$P_A(A,i,t) = \frac{\sum_{t\to endofmatch}(K_{Pass}Pass(A,i) + K_{Duel}Duel(A,i) + K_{Foul}Foul(A,i))}{\sum_i(K_{Pass}Pass(A,i) + K_{Duel}Duel(A,i) + K_{Foul}Foul(A,i))}$$

where all the $K_x s$ are parameters to be determined.

3.2.3 Tactic Deployment

In section 3.1, our roughness not only lies in the individual ability, but also in the tactical coordination of the team. The game of football cannot be too easily quantified as a contest of player strength because of the variety of tactics. Tactics are the concentrated display of team cooperation, and according to the size of the scope, we divided it into local tactics and global tactics. In this paper, we selected three representative types of local and overall tactics for analysis, and hope to show the team's teamwork ability through these indicators.

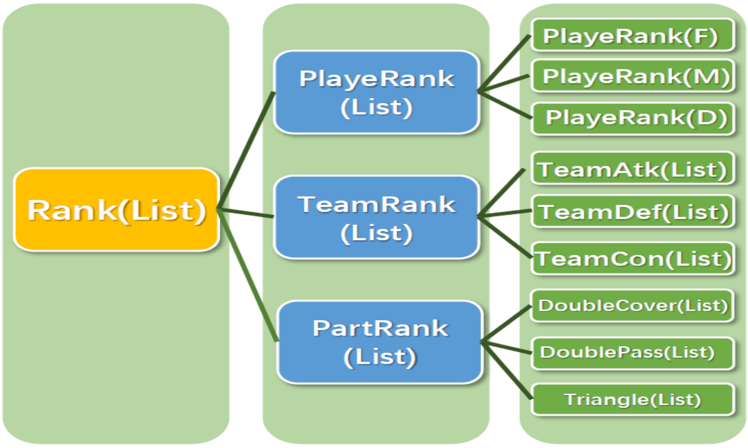
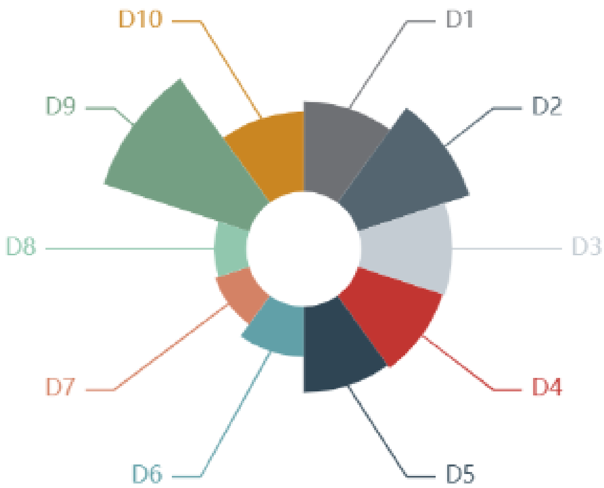


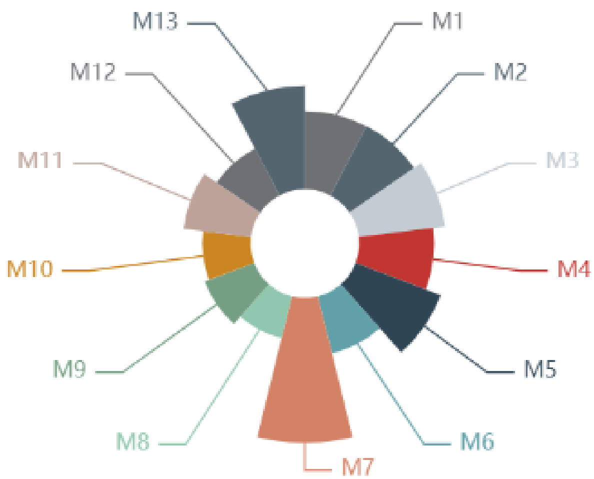
Figure 5: An illustration of the tactic deployment system

We will in the first place discuss global tactic deployment.

Qicik-Attack This strategy is generally used against weaker opponents (in the context of our question we use team score to indicate). Midfield and defense who are able to quickly pass the ball to the forward and active forward are chosen for this tactic.



(a) Attack degree of *defense*.



(b) Attack degree of *midfield*.

Figure 6: Representation of the degree of attack for *defense* and *midfield*. Every area symbols a player, and the bigger the area is, the deeper is the degree.

We quatify *Quick Attack* as the following expression:

$$TeamAtkF(F) = K_{TAFP_d} r_{P_d}(F) + K_{TAFd} r_d(F) + K_{TAFs} r_s(F)$$

$$TeamAtkM(M) = K_{TAMk_{atk}} k_{atk}(M)$$

$$TeamAtkD(D) = K_{TADk_{atk}} k_{atk}(D)$$

Where $K_{TAMk_{atk}}$, $K_{TADk_{atk}}$, K_{TAFP_d}, K_{TAFs}, K_{TAMP_d} are parameters to be determined.

Strong Defense Our attacks often fail to work when faced with a strong opponent, and overspending on them can make the defense vulnerable. One preferred solution against a strong opponent is to form a strong defensive team. Strong defensive lineups generally require the ability to hold the ball in the middle and back and the ability of all players to defense, achieving protection with the ball and counter without the ball.

Similarly, the measuring indicator can be expressed as:

$$TeamDefF(F) = K_{TDFd}r_d(F)$$

$$TeamDefM(M) = K_{TDMd}r_d(M) + K_{TDMP_0}r_{P_0}(M) + K_{TDMP_d}r_{P_d}(M) + K_{TDMt_{P_c}}t_{P_c}(M)$$

$$TeamDefM(D) = K_{TDDd}r_d(D)+K_{TDDP_0}r_{P_0}(D)+K_{TDDP_d}r_{P_d}(D)+K_{TDDt_{P_c}}t_{P_c}(D)+K_{TDDf}r_f(D)$$

All the $K_x s$ are also parameters to be assigned.

Passing Squd Centered by High Pass It has the advantage of being able to pass directly between distant points and is often combined with a variety of tactics to achieve quick point-to-point passing to enhance attack/defense. This tactic requires the ability of all players' ability to pass the ball (especially the ability to pass high passes).

It can be quatified as:

$$TeamControlF(F) = \sum_{Y \in (F,D,M)} \sum_{x \in (P_0,P_d,P_c,P_h,d_a)} K_{TCYx}r_x(Y)$$

Where there are 15 parameters to be determined.

In total, define $TeamRank(List) = TeamAtk(List) + TeamDef(List) + TeamControl(A)$.

For local tactic deployment, we mainly design three methods for tactics.

Double-cover only to be discussed among *midfield* and *defense*.

Football is a combination of attack and defence, and in defence, 1-1 is inevitable. As a member of the league, the condition of our players is also studied by other teams, in which the ability to duel is one of the important statistics. Against the ability of weak players, it is easy to become the other side's offensive breakthrough. And the duel ability is directly reflected in the data in our model is the duel rate (the number of successful duels of players/the time spent on the field).

In order to prevent the weak players from becoming the breakthrough of the other team and leading to the collapse of the whole defense, we studied the double-cover tactics.

Definition 5. *Class A: weakest players composed of the worst three players of defense and midfield. Specifically, D2, D3, D4 and M2, M6, M8.*

Definition 6. *Class B: strongest players composed of the best three players of defense and midfield. Specifically, D1, D6, D7 and M9, M10, M12.*

We say that for every class A player in the lineup, there is a breach in the defense, and every class B player can make for it. That is, each A and B in the lineup form a pair of **double-cover**, in which B is responsible for the breach caused by A.

By definition, score for measuring *double-cover* can be expressed as following:

$$DoubleCover(List) = K_{DCR} \min numberA(List), numberB(List)$$

where $numberA(List)$ and $numberB(List)$ denotes the number of players belong to A and B respectively. K_{DCR} is a parameter to be determined.

(a) *Defense* (b) *Midfield*

Figure 7: Passing rate, smart pass rate and duel rate for *defense* and *midfield*

Double-pass only to be discussed among *midfield* and *forward*.

Double-pass is a common attacking tactic in which two players pass through a defense in tandem. Since this tactic requires good team coordination, we assume that only players in *dyatia configuration* can perform this tactic. In addition, considering that this is an attacking strategy, both players need to be forward and midfield, and k_{atk} needs to be above average (here we assume that the team including the forward must surpass, because the forward's k_{atk} should be obviously better than the midfield).

Therefore, score for measuring *double-pass* can be expressed as:

$$DoublePass(List) = K_{DPR} numberDP(List)$$

where $numberDP(List)$ is the number of *double-pass* pairs and K_{DPR} is a parameter to be determined.

Triangle Attack and Defense System The triangle position has a strong versatility in football, can both attack and defend, having a high tactical value. Considering that this tactic requires a lot on the tacit cooperation of the three triangle players, we assume that only the *griadic configuration* can complete this tactic. Considering that the completion of this complex strategy requires strong individual ability, we think it is necessary that the sum of the PlayeRank of the three players greater than the sum of the corresponding positions'rank.

Score for measuring *Triangle Attack and Defense System* is:

$$Triangle(List) = K_{TR} numberT(List)$$

where $numberT(List)$ is the numberof *triangle attack and defense ystems* and K_{TR} is another parameter to be determined.

In summary, define $PartRank(List) = DoubleCover(List) + DoublePass(List) + Triangle(List)$.

3.2.4 Integration

As mentioned before, the measurement of this model consists of three aspects, i.e. individual ability, global tactic and local tactic.

$$Rank(List) = PlayeRank(List) + TeamRank(List) + PartRank(List)$$

3.3 Group Dynamics Conditions

Abstract theoretical approaches of findings from four of the most-researched topics within sport groups, namely, group size, group composition, group cohesion, and environment [4].

Previous two models considers the group composition and group cohesion, for this part, it is ituitive that we consider team leadership (coaches) and matching environment (away/home) and member relationship given such data.

3.3.1 Team Leadership: Coaches

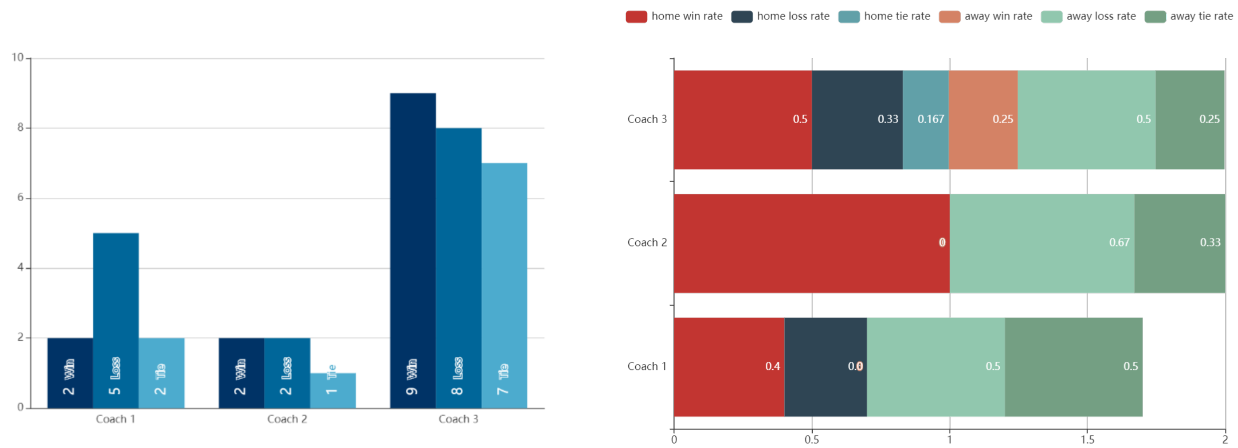
For the sake of delving into the relationship between team cooperation and team coaches, we use the result of the matches as the indication of team cooperation due to the lack of data.

Next, we conclude the data regarding coaches and the result of matches to form the following figure 8a.

From figure 8a, we can see that under similar conditions $Coach2$ and 3 have a relatively high percentage, where as $Coach1$ is slightly inferior.

3.3.2 Matching Environment

Same as the above section, we also look into the matching environment and the result of matches, which contributed to figure 8b.



(a) Different coaches

(b) Different environment

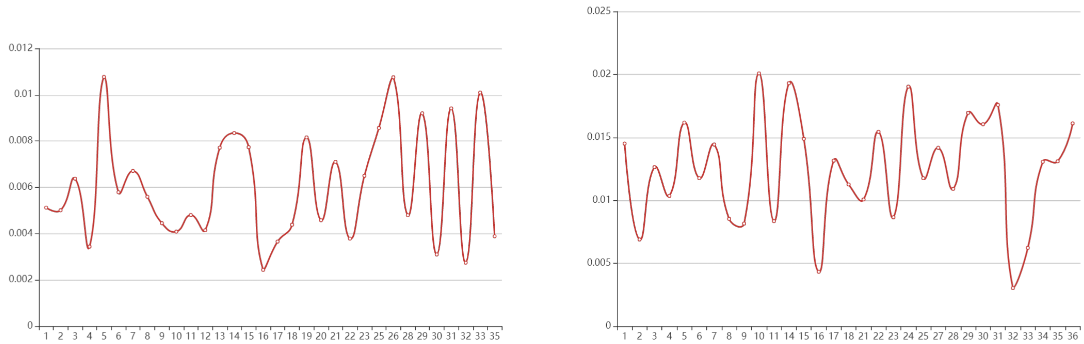
Figure 8: Win, loss and tie rate of different *Group Dynamics Condictions.*

For analyzing the condition of environment, we shall point to figure 8b. It can be easily inferred that when playing away from home, the winning rate tends to decline dramatically, both reaching 0 as $Coach2$ or $Coach3$ is leading.

3.3.3 Member Relationship

Using the model developed in 3.2, we are able to see one player's performance during the whole season for every match, by analyzing the situation for some striking points of the player, we will be able to see the relationship between him and others.

Evidently, passing with others is a direct indicator of the relationship with other team members. Therefore, we analysis the passing rate for every team member during the whole season. Take $F1$ and $F2$ as an example.



(a) $F1$ (b) $F2$

Figure 9: Passing rate for players during the whole season (only consider when he participates in the match).

It can be seen from figure 9a 9b, the passing rate for $F1$ and $F2$ both reached the lowest during $match16$. By analyzing the presence personnel for $match16$ and others, the absence of $D1$ may be the reason. Because $F1$, $F2$ and $D1$ is the star players of regular squd, it is possible that they spend a lot of time training together. Therefore, when $D1$ is changed to other player, the cooperation is largely reduced.

3.4 Opponent Analysis

In this part, we propose a model on estimating the strength of the 19 opponent teams.

$$Score = \frac{max + 1}{min + 1} * ori\_score * en\_indicator$$

where

- max is the lager score of a match.
- min is the smaller score of a match.
- The ori_score table is listed as table 4.
- The $en\_indicator$ table is shown as table 5.

Table 4: value of ori_score

Value	Situation
5	win
3	tie
1	loss

Table 5: value of $en\_indicator$

Value	Situation
0.75	home
1.25	away

4 Implementation

4.1 Naive Network Model

Firstly, we consider the basic graph theory definitions to measure the performance throughout the season for *Hukies* and visualize our results as shown in figure 10.

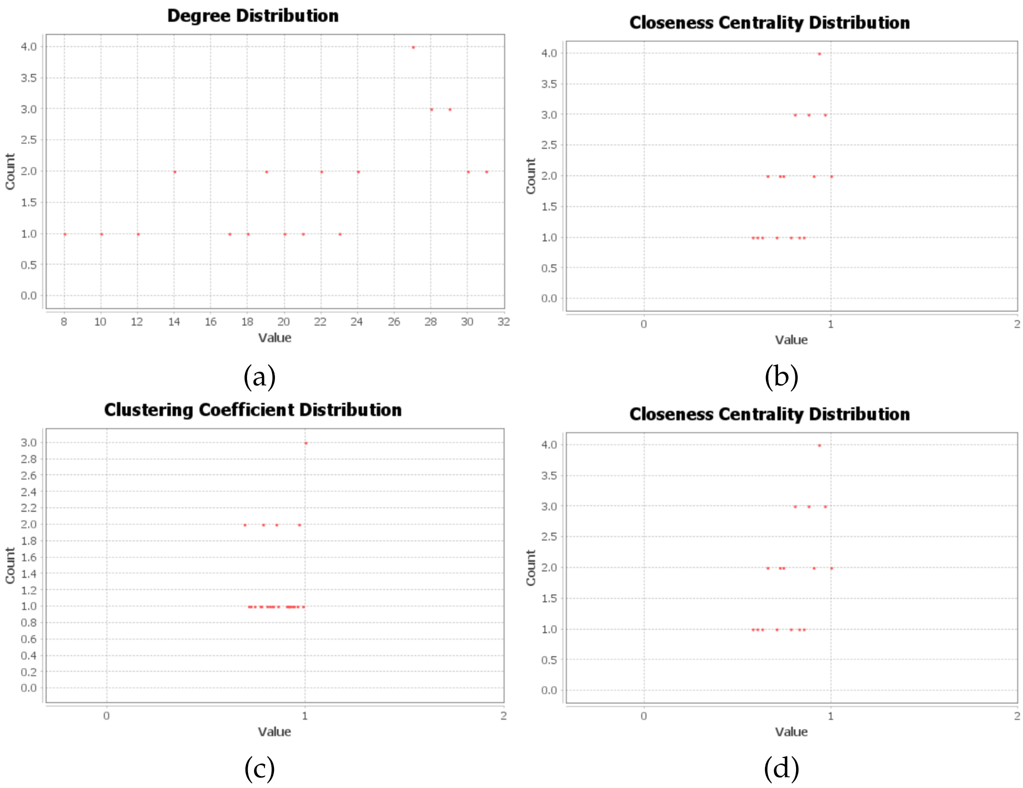


Figure 10: Basic properties of the graph in figure 3a. Figure 10a depicts the **degree distribution** *Huskies* players, the average degree is calculated to be 22.933. Figure 10b plots the **Eigenvector Centrality Distribution** of *Hukies*. The greater the score, the nearer a vertex is to other vertexes. Figure 10c pictures the **Clustering Coefficient Distribution** of *Hukies* with average clustering coefficient 0.863 and 2130 triangles in the graph [5] and Figure 10d illustrates the **Closeness Centrality Distribution** of *Huskies*. The graph is calculated with an average path length 1.2413 and the distance of the farthest vertexes is 2.

Inferred from the *Clustering Coefficient Distribution*, we can briefly conclude that the network is **Triadic Configuration** [6].

From the macro point of view, we can see that *Huskies* has a good degree of close contact and cooperation since the average numbers are relatively ideal, but the horizontal variance is still large which can be seen from the distribution.

Parameter Determination Now we began to determine the parameters mentioned in subsection 3.1.

- **Assignment for K_{P_0}, K_{P_d}, K_{P_c}, K_t:** After combining all the data and taking account of the score balance, we assigned the four parameters as $K_{P_0} = K_{P_c} = K_{P_d} = 1$ and $K_t = 0.0006$ (all of them are relative).

- **Assignment for r_1 and r_2:** Cconsider dyadic configuration and triadic configuration, their passing rate should be ranking top in the team [7], therefore, we choose the top 5% as the threshold for the dyadic configuration, and the top 10% as the threshold for triadic configuration, namelly, the value of r_1 is equal to the passing rate of 5% for the team, 10% for r_2 at the same time. After calculation, the exact value is: $r_1 = 0.00255, r_2 = 0.00177$

- **Assignment for K_{dya} and K_{tri}:** similar as above, with adjustment and balance, we assigned the two parameters as $K_{dya} = 0.004$ and $K_{tri} = 0.012$.

We have computed the C-indecator for the regular squd (besides player G1, whom must paticipate in every game):

$$C_r - indicator = 0.168905$$

In order to choose the best squd with greatest C-indicator, we traverse all the possible *4-2-2* squdsaround 3 million data to find the best one.

The best squd computed by our model is: $F2, F3, D9, D1, D2, D3, M1, M3, M5, M7$ with $C - indicator = 0.229424$ obviously greater than the regular squd.

4.2 Modefied Network Model

4.2.1 Parameter Determination

To better depict the overall ability for the team, the weight for *global tactic* should be greater than *PlayeRank* for tactical core players and lower otherwise, so that we can select a squd when playing certain tactics specifically.

- For PlayeRank, the parameters are assigned according to the different functions for forward, midfield and defense [8] as illustrated in table 6.

- For Local Tactic deployment, the parameters are assigned as following:

$$K_{DPR} = 0.03; K_{TR} = 0.005; K_{DCR} = 0.01$$

- As for Global Tactic deployment, the tactic core for *Quick Attack*, *Strong Defense* and *Passing Squd Centered by High Pass* is the forward, defense and the midfield respectively. Therefore the parameters are assigned according ly as shown in table 7.

Table 6: Parameters assigned for *PlayeRank*

Symbol	Value	Symbol	Value	Symbol	Value
K_{PRFs}	10	K_{PRMP_d}	2	K_{PRDP_0}	1
K_{PRFP_0}	1	K_{PRMP_c}	0.5	K_{PRDP_d}	1
K_{PRFP_d}	1.5	$K_{PRMt_{P_c}}$	0.5	$K_{PRDk_{P_s}}$	0.5
$K_{PRFt_{P_c}}$	0.0001	$K_{PRMk_{P_s}}$	0.2	$K_{PRDk_{atk}}$	0.0001
K_{PRFd}	2	$K_{PRMk_{atk}}$	0.00005	K_{PRDd}	2
K_{PRFf}	-5	K_{PRMd}	1	K_{PRDF}	-3
K_{PRMP_0}	2	K_{PRMf}	-5	$K_{PRDt_{P_c}}$	0.00005

Table 7: Parameters assigned for *Global tactic*

Symbol	Value	Symbol	Value	Symbol	Value	Symbol	Value
K_{TDFd}	2	K_{TDMd}	1.3	K_{TDMP_0}	0.6	K_{TDMP_d}	0.6
$K_{TDMt_{P_c}}$	0.004	K_{TDDd}	2.5	K_{TDDP_d}	1	K_{TDDP_0}	2
K_{TDDf}	5	K_{TAFd}	5	K_{TAFs}	15	K_{TAFP_d}	1
$K_{TAMk_{atk}}$	0.012	$K_{TADk_{atk}}$	0.008	$K_{TDDt_{P_c}}$	0.003	K_{TCFP_0}	0.7
K_{TCFP_d}	1.3	K_{TCFP_c}	0.5	K_{TCFP_h}	0.2	K_{TCFd_a}	5
K_{TCDP_d}	0.7	K_{TCDP_c}	0.5	K_{TCDP_0}	1.3	K_{TCDP_h}	0.35
K_{TCDd_a}	3	K_{TCMP_d}	0.8	K_{TCMP_c}	0.5	K_{TCMP_h}	0.35
K_{TCMP_0}	0.8	K_{TCMd_a}	3				

4.2.2 Physical Ability Evaluation

Parameter Determination In light of the consumption degree of $Pass$, $Duel$ and $Foul$, we set:

$$K_{Pass} = 1, K_{Duel} = 3, K_{Foul} = 10$$

Result Computation Considering the strength and activity of players, we seleted 6 main players evenly from forwards, midfields and defenses, inlight of *PlayeRank* and the total matching time [9], the attributes of them is shown as figure 11.

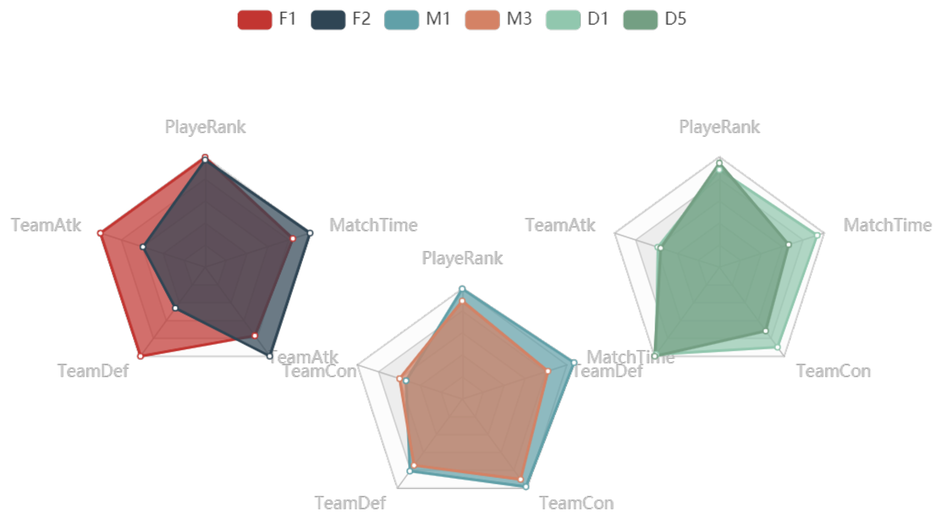


Figure 11: The distribution of different abilities of the main players in $Huskies$.

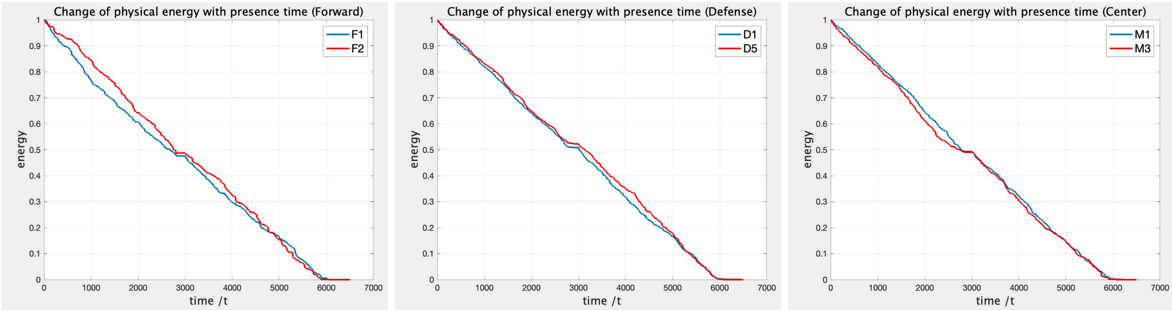
Then we choose such match that the aforementioned players presented the whole match, picturing their physically ability as shown in figure 12.

In figure 12, the lower the players are, the faster their physical ability decreases. From the figure, we can see that the physical ability decreases of F1, M3 and D1 are faster than that of F2, M1 and D5 respectively, which will be reflected in the subsequent substitution tactics.

4.2.3 Squd Optimization

The $Rank(List)$ for the regular squd is computed as:

$$Rank(Regular List) = 2.16757$$



(a) $F1$, $F2$ (b) $D1$, $D5$ (c) $M1$, $M3$

Figure 12: The PA(physically ability)-time plot of the main players.

With the same appoach as section 4.1, we traverse all the possible squds to find such squd with the highest $Rank(List)$.

The squd chosen is:

$$ListBest = (F1, F4; D1, D3, D9, D10; M1, M3, M7, M13; G1)$$

with $Rank(ListBest) = 2.29806$, which outperforms the baseline, i.e. the regular squd.

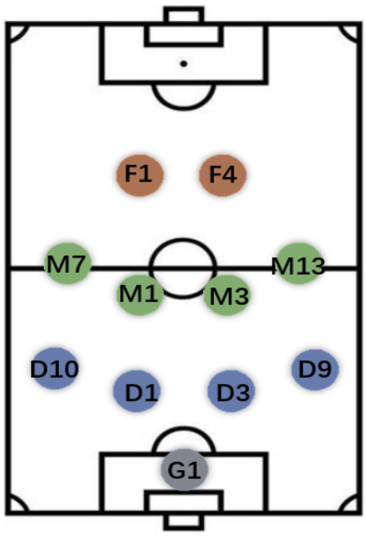


Figure 13: The best squd in general for $Huskies$.

4.3 Opponent Analysis

We calculated the score defined in 3.4, and the result is shown as the following figure 14, from which we can see that *opponent team 4* and *opponent team 5* both are strong competitors.

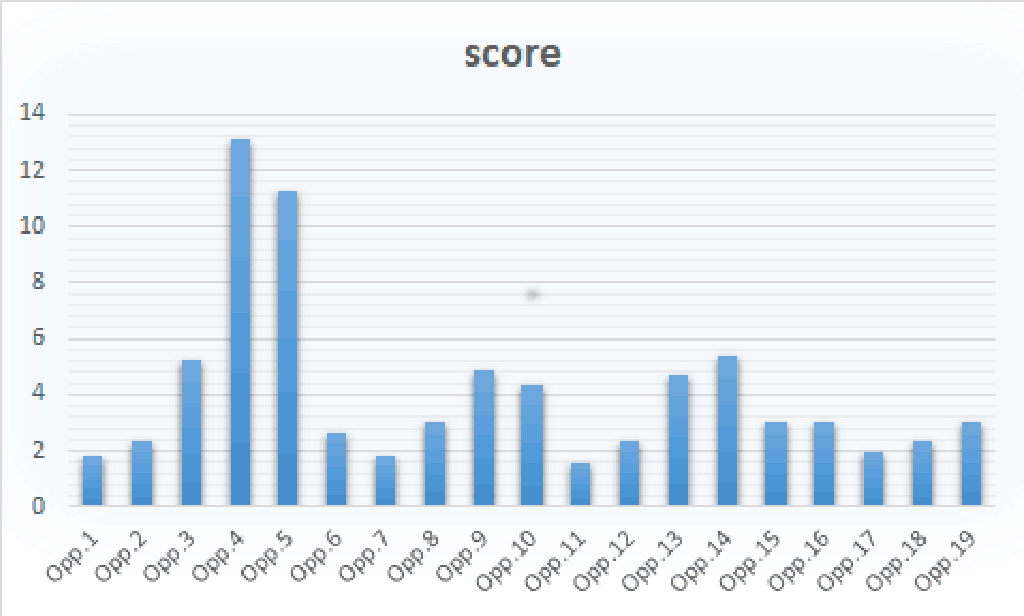


Figure 14: General situation of all the 19 opponent teams. Each team is represented by a bar.

5 Strategies

5.1 Strategies for *Huskies*

5.1.1 Tactics Specialization

From the above results, we can get the best team in the combination of individual ability and team tactics. This lineup can be called versatile or featureless. Although we subdivide the tactics into attack, defence and control in section 3.2.3, we give the same weight to all three in the aforementioned lineup, which doesn't really make much difference against different opponents. Therefore, we introduce a new concept, **specified best squd**.

The specified best squd is divided into best-attack, best-defense and best-control. Increase the weight of the three factors to show the emphasis on tactics. At the same time, the regular squd of $4-4-2$ is adjusted (specifically, attack squd $4-3-3$, defense squd $5-3-2$ and control squd $3-5-2$) [10]. On the basis of section 3, we only changed the weight of three global tactics and the position ratio of the lineup, then recalculated the results, and obtained the following three optimal squads as shown in figure 15.

The detailed squd and its corresponding $Rank(List)$ of the three classes ranked top three is listed as table 8

Table 8: squd and the related Rank(List) of three classes

squd(Rank)		
Attack-squd	Defense-squd	Control-squd
$(F_{1,2,4}, M_{1,7,13}, D_{1,3,5,9})$ (2.2648)	$(F_{1,4}, M_{1,5,7}, D_{1,3,5,9,10})$ (2.2700)	$(F_{1,2}, M_{1,3,5,6,7}, D_{1,3,9})$ (2.3538)
$(F_{1,4,5}, M_{1,7,13}, D_{1,3,5,9})$ (2.2640)	$(F_{1,4}, M_{1,7,13}, D_{1,3,5,9,10})$ (2.2695)	$(F_{1,2}, M_{1,3,6,8,14}, D_{1,3,9})$(2.3537)
$(F_{1,2,4}, M_{1,7,13}, D_{1,3,9,10})$ (2.2629)	$(F_{1,4}, M_{1,5,7}, D_{1,3,7,9,10})$ (2.2656)	$(F_{1,4}, M_{1,3,5,6,7}, D_{1,3,9})$(2.3510)

(a) Best attack-squd (b) Best defense squd (c) Best control squd

Figure 15: Three squds with respect to tactic deployment.

5.1.2 Opponent Consideration

From the discussion of global tactics, we can know that the squd of attack, defense and control are suitable for facing weak, strong and even opponents respectively. Figure 14 shows our analysis of the enemy.

Rank the opponents by their score, and devide them into three clusters, we have:

$$L1 = (Opp.4, Opp.5, Opp.3, Oppo.10, Opp.13, Opp14, Opp.9)$$

$$L2 = (Opp.6, Opp.8, Opp.15, Opp.16, Opp.19)$$

$$L3 = (Opp.1, Opp.2, Opp.7, Opp.11, Opp.12, Opp.17, Opp.18)$$

By our conclusion, squd of defense, control and attack are suitable for opponent teams in $L1, L2, L3$ respectively.

5.1.3 Physical Strength Situation

In section 4.2.2, we discussed the physical decline of several main players, and for this reason, we also believe that players with physical decline should be replaced in time to maintain their physical condition and the team's efficiency on the field.

As is shown in figure 12, the physical strength of the players will drop to 30% of the original physical strength at time $4000s$ ($1000s$ of the second half of the match). We take this as the limit and think that players should be replaced when their physical strength is lower than 30%. Take $F1$ as an example, we should replace him at $3976s$. Similarly, $F2$, $M1$, $M3$, $D1$ and $D5$ should be replaced at $4089s$, $4032s$, $4015s$, $4035s$ and $4307s$ respectively.

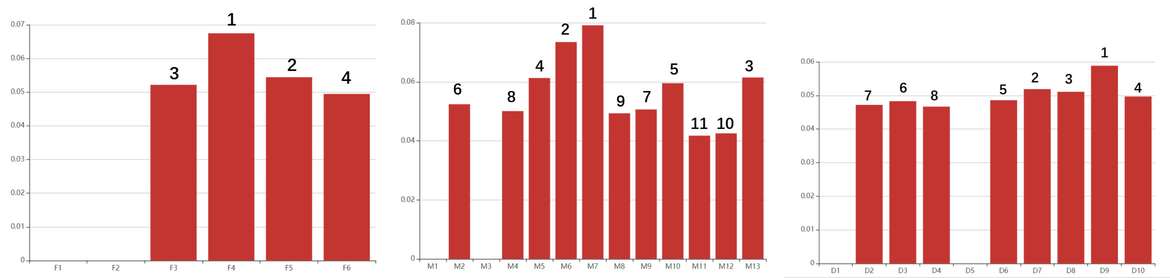
As for whom to replace the main player, we consider the indicator named *reinforcement*, denoted as $R(A)$, which is defined as :

$$R(A) = PlayeRank(A) + p(A)$$

. Where $p(A)$ is the profession-related indicator for player A. Specifically, for forwards, $R(A) = PlayeRank(A) + TeamAtk(A)$.

For example, when replacing $F1$, we should consider consider the highest score except $F1$ and $F2$, so $F1$ should be replaced with $F4$. Similarly, we should replace $F5$ for $F2$, $M7$ for $M3$, $M6$ for $M1$, $D9$ for $D1$, and $D7$ for $D5$.

Finally, we append $R(i)$ for non-main players as shown in figure 16.



(a) $R(i)$ for all forwards (b) $R(i)$ for all midfields (c) $R(i)$ for all defenses

Figure 16: $R(i)$ for all players. Rankings are stated on the figure, right above the bars. Main players are left in vacancy.

5.2 Advice for Building Effective Teams

For this part, we mainly give advice according to conclusions drawn based on *Group Dynamics*. The advice can be summerized as follows:

1. Leadership
   Although teamwork is about a group of people working together, the position a leader can not be undermined. In team building, the leader is the header. He is the core of the whole task, all members should be commanded by him, obey his orders, only by this will the team be in the optimal configuration. Also, one can see this by the study of coaches of $Huskies$.
   Therefore, a strong, determined and insightful leader is advised.

2. Environment
   The working environment is very important for teamwork. Just as environment is for individual working, it is the same for teamwork.
   A suitable working environment is also one of our suggestions.

3. Internal Relationship
   Teamwork is by no means a simple additive contributions of teammates, relationships among menbers also plays an essential role. It can act as positive catalyst or negative catalyst depending on its properties.
   We propose that when selecting team members, factors like characteristics of each person that could influence internal relationship should be taken into account.

In order to build a generalized model of team performance, we think that the following aspects are still needed:

- Personal Inclinations.
  We need to know the inclination for cooperation of every team member to organize team members, avoiding possible contradictions.

- Strengths and Weaknesses
  An insight into a teamate's strengths and weaknesses is necessary for allocating jobs inside team. In our model, we use the number of certain events to represent the ability of a player, shown in figure 11 this is not quite precise though. For example, a player with high passing numbers may not indicate he has a high passing ability, it may because his attack ability was suppressed by the opponents and has to go for passing.

6 Model Analysis

6.1 Sensitivity Analysis

Our Strategy Model is mainly designed for planning player line-up in different situations. As we know, on the fast-changing football field, its common to see that our players get injured and the replacement between them, which will certainly have an impact on the combat effectiveness of the entire team. Therefore, in this section, we would like to produce a sensitivity analysis to show whether our model is sensitive to the personnel turnover in different player positions.

It is worth nothing that when we discuss personnel turnover in a specific position, the best line-up is maintained in other positions. The possible combinations of the three positions are :

$$Forward : C_6^2 = 15, Midfield : C_{13}^4 = 715, Defense : C_{10}^4 = 210$$

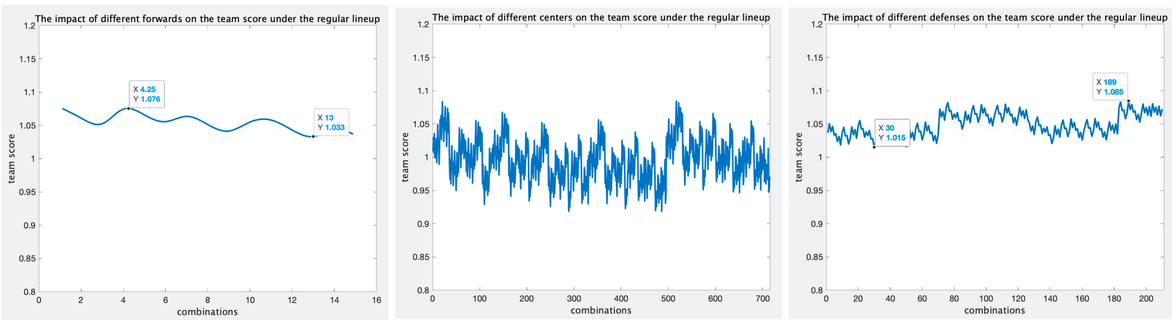
After processing the data and calculating the standard deviation of each group, we can see that the sensitivity ranking of our model to different positions is:

$$Midfield > Defense > Forward$$

With standard deviations 0.03216, 0.01641 and 0.01638 respectively.

Thus we can know that our model is most sensitive to the personnel turnover in midfield, followed by defense, and finally, the forward.

Figure 17: The changes in team score when a player at this position changes



(a) *Forwards* (b) *Midfields* (c) *Defenses*

Figure 18: Changes in team score for changing players in different positions.

6.2 Strengths and Weaknesses

6.2.1 Strengths

1. We have taken into account a number of data used to assess players' individual abilities and have devised a scoring mechanism *PlayeRank* to make our assessment fairer.

2. We made a deep refinement of team cooperation, observed *fullevent.csv* in detail, extracted some tactics with the knowledge of soccer of our team members and realized them through the deployment of the *Hukies* players.

3. We have done a lot of visualization to show the strengths and weaknesses of players' abilities, making it more intuitive to select players for different tactics.

4. We did a timeline analysis of some players' performances, both at different points in time and between games, which has a better analytical effect on team members' ability and performance as well as their interpersonal relationship within the team.

5. In terms of overall tactics, we chose one representative of the three position-centered plays, and one of the three most important plays, namely, attack and defense ball control. We also built an indicator (TeamAtk, TeamDef, TeamControl), which can intuitively reflect the team's three styles, not only intuitive, but also make the team's tactics more distinctive.

6.2.2 Weaknesses

1. Sometimes, some team data are not so accurate, because the players in the same position are usually interchangeable, which makes their cooperation appearing less than that between different positions

2. Due to the limitation of the data given, we are unable to obtain a lot of important information (for example, whether the goal is scored successfully or not is not reflected in *fullevent.csv*, and the starting list of each half-court game also needs to be calculated).

3. Some players, e.g. D9, M7, rarely get the chance to play as substitutes, which makes them cherish the chance to play and make efforts to perform, resulting in the data per unit time very high (even much higher than the previous main players), and it is difficult to exclude the influence of this situation after player modeling.

4. In the actual game, although the number of overall tactics is small, the number of local tactics is very large (even several times the number of overall tactics), but in our analysis, only three common ones are selected.

5. It is very difficult to quantify the events of team members into team contributions (the selection of parameters in the model), because it is difficult for us to accurately obtain the weight of various events, so we can only assign different proportions to the parameters when calculating various data to show the team's ability.

7 Conclusion

Our paper provides a modeling method to the evaluation of professional football teams and the selection of strategies. In this model, we quantified the individual and cooperative abilities of players into passing ability, duel ability and so on. Based on the knowledge of graph theory and combined with the processing and use of data, we first weighted a series of data of passing ability, obtaining a simple passing network model.

On the basis of above, we considered various other data besides passing ability, and built the "MaNGO" model by combining PlayeRanks (considering individual abilities) of three professions, physical strength (using *GM*), and tactics (for real football matches).

Strategies are fully discussed both internally and externally, a generalization is made for our model to give some suggestions for various teamwork other than sports.

References

[1] Mufasa and sansha, “Internet science has revealed the secrets of the top football team barcelona,” Sohu, 2019. [Online]. Available: http://www.sohu.com/a/345517471_100104269

[2] P. Cintia, F. Giannotti, L. Pappalardo, D. Pedreschi, and M. Malvaldi, “The harsh rule of the goals: Data-driven performance indicators for football teams,” in *2015 IEEE International Conference on Data Science and Advanced Analytics, DSAA 2015, Campus des Cordeliers, Paris, France, October 19-21, 2015*. IEEE, 2015, pp. 1–10. [Online]. Available: https://doi.org/10.1109/DSAA.2015.7344823

[3] L. Pappalardo, P. Cintia, P. Ferragina, D. Pedreschi, and F. Giannotti, “Playerank: Multi-dimensional and role-aware rating of soccer player performance,” *arXiv preprint arXiv:1802.04987*, 2018.

[4] W. N. Widmeyer, L. R. Brawley, and A. V. Carron, “Group dynamics in sport.” 1992.

[5] M. Latapy, “Main-memory triangle computations for very large (sparse (power-law)) graphs,” *Theoretical computer science*, vol. 407, no. 1-3, pp. 458–473, 2008.

[6] C. K. Mengüç, “A solution of a mathematical model which simulates football game as a logistics network,” in *Proceedings of the 7th International Conference on Sport Sciences Research and Technology Support, icSPORTS 2019, Vienna, Austria, September 20-21, 2019*, J. Vilas-Boas, P. P. Correia, and J. Cabri, Eds. ScitePress, 2019, pp. 105–112.

[7] G. S. Fleisig, R. F. Escamilla, J. R. Andrews, T. Matsuo, Y. Satterwhite, and S. W. Barrentine, “Kinematic and kinetic comparison between baseball pitching and football passing,” *Journal of Applied Biomechanics*, vol. 12, no. 2, pp. 207–224, 1996.

[8] E. Ozceylan, “A mathematical model using ahp priorities for soccer player selection: a case study,” *South African Journal of Industrial Engineering*, vol. 27, no. 2, pp. 190–205, 2016.

[9] D. Strnad, A. Nerat, and Š. Kohek, “Neural network models for group behavior prediction: a case of soccer match attendance,” *Neural Computing and Applications*, vol. 28, no. 2, pp. 287–300, 2017.

[10] D. Vallejo, G. Alises, J. Albusac, C. Glez-Morcillo, and J. J. Castro-Schez, “SIDANE: towards the automatic analysis of football tactics and actions,” in *Proceedings of the 3rd International Congress on Sport Sciences Research and Technology Support, icSPORTS 2015, Lisbon, Portugal, November 15-17, 2015*, J. Cabri and P. P. Correia, Eds. SciTePress, 2015, pp. 334–341. [Online]. Available: https://doi.org/10.5220/0005601503340341

Appendices

Appendix A Simulation for Modified Network Model

```
% Please import the .csv(Personal Data) and two over one matrix(Too_matrix)
% into number matrix first.
K_cd = 0.005;       %co-defense
K_too = 0.01;       %two over one
K_tad = 0.03;       %Triangle assisted defense
player_score(29) = zeros;
sub(10) = zeros;
max_score = 0;
max_sub(10) = zeros;
cnt = 0;            %loop counter
for i = 1 : 29
    player_score(i) = PersonalData(i, 1) + PersonalData(i, 2) +
                    PersonalData(i, 3) + PersonalData(i, 4);%personal score
end
% Simulate number of combinations
F_list = nchoosek([1 2 3 4 5 6], 2);
D_list = nchoosek([7 8 9 10 11 12 13 14 15 16], 4);
M_list = nchoosek([17 18 19 20 21 22 23 24 25 26 27 28 29], 4);
% Iteration Process
for i = 1 : length(F_list)
     sub(1) = F_list(i, 1); sub(2) = F_list(i, 2);
     % Choose 2 forwards each iteration
     for j = 1 : length(D_list)
         sub(3) = D_list(i, 1);
         sub(4) = D_list(i, 2);
         sub(5) = D_list(i, 3);
         sub(6) = D_list(i, 4);        % Choose 4 defenses each iteration
         for k = 1 : length(M_list)
             sub(7) = M_list(i, 1);
             sub(8) = M_list(i, 2);
             sub(9) = M_list(i, 3);
             sub(10) = M_list(i, 4);    % Choose 4 midfields each iteration
             cnt = cnt + 1;          % counter update
             disp(cnt);
             tmp_score = 0;
             for t = 1 : 10
                tmp_score = tmp_score + player_score(sub(t));
             end
             tmp_score = tmp_score + K_cd * co_defense(sub);
             % Co-defense situation
             tmp_score = tmp_score + K_too * TwoOverOne(Too_Matrix);
             % Two-over-one situation
             tmp_score = tmp_score + K_tad * tri_defense(sub);
             % Triangle defense situation
             if(max_score < tmp_score)
                 % Update the max score and record its line-up
                max_score = tmp_score;
                for t = 1 : 10
                    max_sub(t) = sub(t);
                end
             end
         end
     end
end
function res = TwoOverOne(Matrix)%Return the number of two-over-one couples
    res = 0;
    for t = 1 : 29
        for v = 1 : t-1
            if(Matrix(t, v) == 1)
                res = res + 1;
            end
        end
    end
end
function res = tri_defense(sub)%if existed - return (res = 1)
    triangle_de = [7 9 23];%Triangle assisted defense
    flag(3) = zeros;
    for i = 1 : 10
        switch sub(i)
            case(triangle_de(1))
                flag(1) = 1;
            case(triangle_de(2))
                flag(2) = 1;
            case(triangle_de(3))
                flag(3) = 1;
```

```
        end
    end
    if(flag(1) == 1 && flag(2) == 1 && flag(3) == 1)
        res = 1;
    else
        res = 0;
    end
end
function res = co_defense(sub)%Return the number of co-defense pairs
    weak = [8 9 10 18 22 24];%M2, M6, M8, D2, D3, D4 (co-defense)
    strong = [7 12 13 25 26 28];%M9, M10, M12, D1, D6, D7 (co-defense)
    cnt1 = 0; cnt2 = 0;
    status(30) = zeros;
    for i = 1 : 10
        if(sub(i) >= 7)%Not Forward
            status(sub(i)) = 1;
        end
    end
    for i = 1 : 6
        if(status(weak(i)) == 1)
            cnt1 = cnt1 + 1;
        end
        if(status(strong(i)) == 1)
            cnt2 = cnt2 + 1;
        end
    end
    res = min(cnt1, cnt2);
end
```

Appendix B Physical Ability Evaluation

```
% Please import the .csv(a player's event) into a string matrix
% (named 'M3_Data') first. Here we take the fitness changes of
% M3(Midfield No.3) over time for example.
% energy(t) = (the sum of event scores from moment t to
% the time of the last event played by this guy) / (the sum of all event
% scores made by this guy in this game)
global times
times = 1243;
% Number of records. We filter out matches played for more than
% 5500 seconds in order to ensure the robustness of the model
score_pass = 1;    % a pass = 1 point
score_duel = 3;    % a duel = 3 points
score_foul = 10;   % a foul = 10 points
score(6500) = zeros;  % first half 3000s, second half 3500s for a game
% Iteration Process
for i = 1 : times
     tmpHalf = char(M3_Data(i, 2));
     if(strcmp(tmpHalf, '1H') == 1)
         t = time(char(M3_Data(i, 5)));   % first half of the game.
     else
         t = time(char(M3_Data(i, 5))) + 3000; % second half of the game.
     end
     switch char(M3_Data(i, 4))
         case("Pass")
             score(t) = score(t) + score_pass;
         case("Duel")
             score(t) = score(t) + score_duel;
         case("Foul")
             score(t) = score(t) + score_foul;
     end
end
tmp_sum = 0;
% Further processing of the data.
for i = 1 : 6500
     score(6501 - i) = score(6501 - i) + tmp_sum;
     tmp_sum = score(6501 - i);
end
% Normalized
for i = 1 : 6500
    score(6501 - i) = score(6501 - i) / score(1);
end

% Drawing operation
x = 1 : 1 : 6500;   % We divide a game into 6500s.
values = spcrv([[x(1) x x(end)]; [score(1) score score(end)]]);
% Curve smoothing
f1 = plot(values(1,:), values(2,:), 'lineWidth', 2);
title("Change of physical energy with presence time (Midfield)",
```

```
'fontname', 'Time New Roman', 'fontsize', 15);
xlabel("time /t", 'fontname', 'Time New Roman', 'fontsize', 15);
ylabel("energy", 'fontname', 'Time New Roman', 'fontsize', 15);
h = legend(f1, 'M3');
set(h, 'fontsize', 15, 'color', 'w', 'edgecolor', 'k', 'textcolor', 'k');
% legend setting
grid on;

function res = time(str)
% To transfer the type of time record (from string to double)
    t = str2double(str);
    res = floor(t);      % Round down
end
```