Teaming Wisdom Coming from Soccer Field

Summary

Nowadays teamwork is playing an increasingly important role in all walks of life, especially team sports. Team success is much more than the sum of the abilities of individual players. Rather, it is based on many other factors that involve how well the teammates play together. The purpose of this paper is to create a model that captures performance indicators that reflect successful teamwork, and generalize the model to other domains.

In TASK 1, we establish a *ball passing network* based on the passing events, in which each player is a node and each pass constitutes a link between players. We define several indicators helping us know about the network on the whole. Then we utilize *K-means Algorithm* to classify *dyadic and triadic configurations* in the network, which help to identify the network patterns. Moreover, we explore the network with other scales like time and region to further study the network.

In TASK 2, we first identify performance indicators that reflect successful teamwork, such as coordination, distribution, tempo, flexibility and pressing. Then an *Evaluation model based on Adversarial Regression* is built to score the teamwork performance of both our team and opponent in one match. To make scores of both teams satisfy the final result of each game, we regulate the weights of five indicators, and record every possible solution. Furthermore, we choose the weight with LowestError as the regression equation. Finally, we notice that only by adjusting the strategy based on the opponents' strategy, trying to make the total score higher or repress the score of opponents, can the Huskies more likely to win.

In TASK 3, we analyze the formations and strategies Huskies adopt, along with the indicators and final results. Considering the actual situation of the Huskies, the most effective strategy is *Defensive counterattack*. Then comparing Huskies with Typical representative of this strategy (*Italy National Team*), we get to know that the huge gap with the strong teams is mainly reflected in *Coordination*, *Flexibility* and *Pressing*. Based on the conclusion we draw, we offer the coach some advice on *flexible formation*, *core players* and *coherence* that help them succeed in the next season.

In TASK 4, we generalize our model to more teamwork, not limited to team sports. With reference to the indicators we identify for football match, we come up with the complex set of factors that make some groups perform better. In addition, other aspects of teamwork are taken into account to organize more effective teams.

Finally, we conduct a sensitivity analysis in order to gain some deep understanding of our model, and verify the robustness of the model in many cases. Additionally, we analyze the strengths and weaknesses of our model.

Keywords: Passing network, Adversarial Regression, Defensive Counterattack

Contents

1 Introduction

1.1 Problem Background

With the development of social interconnections, the challenge people faced has become complex increasingly. Some reports point out the best strategies for team formation, the best interaction between teammates, and the ideal leadership style. A strong team across departments and domains can complete complex tasks that cannot be accomplished through individual efforts or additional contributions from a series of teammates.

The coach has asked ICM to quantify and formalize the structural and dynamical features that have been successful (and unsuccessful) for the team. The team are provided with the following tasks:

- Create a network for the ball passing between players, and explore multiple scales such as regions and time.

- Identify performance indicators that reflect successful teamwork.

- Make suggestions to the coach using the opinions gained from the model.

- Indicate how to build an effective team and give other important factors for forming an dynamic team.

1.2 Our work

1. Firstly, we establish a *ball passing network* based on the passing events.Then we utilize *K-means agAlgorithmm* to classify dyadic and triadic configurations in the network.

2. Then we build an *Evaluation model based on Adversarial Regression*, and get the regression equation. Taking the opponents into account, we have a discussion about whether strategies are universally effective.

3. Furthermore, We analyze the formations and strategies Huskies adopt and recommend the team to adopt *Defensive counterattack*. What's more, we get to know the gap with strong team after comparing with *Italy National Team*.

4. Last but not least, we generalize our model to more teamwork and indicate how to build an effective team and give other important factors for forming an dynamic team.

2 Preparation of the Models

2.1 Assumptions and Justifications

- The ball passing network is an undirected graph.

- The incoming player substitutes the outgoing player directly, without changing the team formation.

- There is no difference between the opponents' strength level, which lies in the opponents' counter strategy.

- Each player, either of Huskies or of opponents, has the same physical quality and individual ability.

2.2   Notations

The primary notations used in this paper are listed in Table 1.

Table 1: Notations

Symbol	Definition
L	Total links of network
ρ	Network Density
w_{ij}	Number of passes
d_{ij}	topological distance
D	Network Diameter
$C(i)$	Clustering Coefficient
f	ratio of goals to shots
d	ratio of defenses to losses
φ	Distribution of contributes
t_b	50-ball Passing Time
μ_i	Number of shots
ν_i	Number of defenses
S	Score of teamwork
β_i	Weight of indicators
γ	Coordination among players

3   TASK 1: Ball Passing Network Model

3.1   Construction of Ball Passing Network

3.1.1   Ball Passing Network

Considering the relationship between players on the football field, the network for the ball passing can be described as an *adjacency matrix*[1] in Data Structure. An adjacency matrix can be generated to perform network analysis, which represents the connections between a node (player) and an adjacency node (teammate). To generate an adjacency matrix for network analysis, you must define criteria that characterize the connections. We define w_{ij} as the number of passes between player i and player j. The adjacency matrix between players can be expressed as:

$$
\begin{array}{cccc}
 & 1 & 2 & \cdots & 11 \\
\begin{array}{c}1\\2\\\vdots\\11\end{array} &
\left(\begin{array}{cccc}
w_{11} & w_{12} & \cdots & w_{1,11} \\
w_{21} & w_{21} & \cdots & w_{2,11} \\
\vdots & \vdots & \cdots & \vdots \\
w_{11,1} & w_{11,2} & \cdots & w_{11,11}
\end{array}\right),
\end{array}
\tag{1}
$$

In this paper, we assume that the ball passing network is an *undirected graph*, which means the edge has no direction. Therefore, we have the number of passes between players $w_{ij} = w_{ji}$

Meanwhile, the ball passing network can be vividly illustrated in **Figure 1** by *Gephi*.
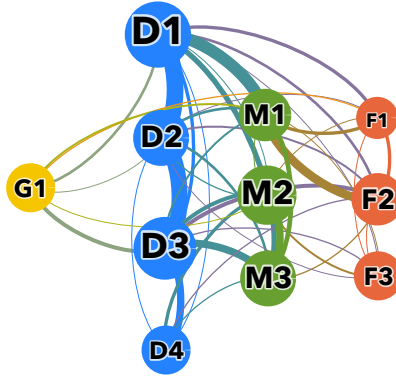

Figure 1: An example of the ball passing network

Where width of the links are proportional to their weights, while the size of the nodes correspond to the number of passes.

3.1.2 Determination of Network Parameter

In order to further understand the network model, we should select some relevant indicators. The ball passing network analysis of the Huskies was focused on the following four measures based on the connections between teammates: *total links*, *density*, *diameter*, and *clustering coefficient*.

- **Total Links** L

The sum of the elements of each row of the adjacency matrix $\sum_{j=1}^{11}$ is the total number of passes between player i and all other teammates, which is the *degree* in *Graph Theory*. A node with a higher degree is a player who made more considerable passes with most of its teammates. This shows that the player is involved in the attacking development of his team..

In an undirected graph, the *Total Links* (L) between each team player is a significant parameter in the analysis, which the value is half of the total sum of each adjacency matrix row sum.

$$
L = \frac{1}{2}\sum_{j=1}^{11}\sum_{i=1}^{11} w_{ij}.
\tag{2}
$$

The *Total Links* is very useful when evaluating the performance of a team, as it is the absolute number of total interactions between teammates in a match. Therefore,

the team's average total link index is higher than the average, which indicates a strong cooperation among team members.

- **Network Density** ρ

Total Links is the number of absolute interactions between one player and another player. However, the *Network Density* of team is a relative parameter, and it can also measure the overall emotion between teammates.[2]

In graph theory, the density of a graph is proportional to the maximum number of possible adjacency edges. Let's consider the number of edges in the graph as n. Thus, the *Network Density* ρ is defined as the ratio of the *Total Links* L to the maximum number of possible adjacency edges as follow:

$$\rho = \frac{L}{\mathbf{C}_n^2}, \tag{3}$$

where in this paper the number of edges $n = 11$, so the density $\rho = \frac{L}{55}$.

- **Network Diameter** D

The *Network Diameter D* of a network reflects how far, at most, two nodes in the graph are. The topological distance d_{ij} of the link between two players i and j is defined as the inverse of the link weight.

$$d_{ij} = \frac{1}{w_{ij}} \tag{4}$$

The *Diameter D* of a graph is the maximum topological distance[3] (the length of the largest geodesic) between any two players and can be expressed as:

$$D = \max_i \max_j d_{ij}. \tag{5}$$

In case of team players, a small diameter reflects a low maximum distance between teammates, which may reveal that the team's passing game was diffused among most of its players (rather than a few acting as central ones).

- **Clustering coefficient** $C(i)$

Considering the network is weighted, we can not simply account for the number of nodes connected between them but how the link weights are distributed. In the case of the *ball passing network*, the number of passes between pairs of players is not constant. As a consequence, we use a *Weighted Clustering Coefficient*[4] $C_w(i)$ to measure the possibility that a given player i's teammates will also create a connection between them:

$$C_w(i) = \frac{\sum_j \sum_k w_{ij} w_{jk} w_{ik}}{\sum_j \sum_k w_{ij} w_{ik}}. \tag{6}$$

Where j and k are the two teammates of player i, w_{ij} and w_{ik} are the number of ball passes between player i and both them. At last, the *Clustering Coefficient* C_i for the entire network is obtained by averaging $C_w(i)$ for all players, i.e.

$$C = \frac{1}{n} \sum_{i=1}^n C_w(i), \tag{7}$$

where in this paper the number of edges $n = 11$.

3.2   Identification of Network Patterns

3.2.1   Team Formation

The formation of each game can be easily obtained by looking through the player list. In the dataset provided for us, all players are placed in four positions, which are *goalkeeper*, *defense*, *midfield* and *forward*. So when we study the formation in one game, we place the players in the position his ID indicates, thus obtaining the formation. To simplify the model, we assume that the substituted player replaces the player substituted directly, without changing the formation. We take the 1st, the 2nd and the 4th match as examples. The formation of these matches means a teamwork strategy. Take 4-3-3 as an example, this formation means we have 4 defenses , 3 midfields and 3 forwards.

The 1st match Huskies take 4-3-3 formation, while the formations of the 2nd and the 4th matches are 5-3-2 and 4-4-2 respectively. In each game, the coach will take different formations according to the strategy, so figuring out the formation is important to understand the strategy, thus better identifying the network pattern. However, formation just gives us a general view of the network. To identify the network pattern more comprehensively, we need more details of the network.

3.2.2   Dyadic and Triadic Configurations

Considering the difference in opponent and strategy, the ball passing network we build varies from match to match. Therefore, in order to identify the network pattern, we need to analyze a specific game in detail. In this chapter, we take the 1st match as an example.

Figure 2: Topological Distance Matrix of the 1st match

First, we apply our model to the 1st match and build the ball passing network. From the adjacency matrix, the *Total Links L* and *Network Density* ρ can be easily obtained by the equation (2) and (3) we give, which are 369 and 6.71 respectively. Moreover, we can use the equation (4) to obtain the topological distance matrix (see **Figure 2**).

In the topological distance matrix, ∞ means the distance between the two nodes is infinite, which suggests that they don't have any passing in the game. There is no doubt that the statistics marked by red box is smaller than other statistic, which means the two nodes are very close in topological distance.

In *Graph Theory*, we call these nodes with *higher similarity (distance) cluster*. Therefore, the passing network has many clusters in it and we use a *K-means Algorithm*[5] to classify the clusters in our passing network. The results have been shown in **Figure 3**.

Figure 3: Dyadic and Triadic Configurations

Consequently, we classify (D1, D2, D3), (D1, M1, F2) and (M2, M3) into three clusters, which are so-called *Dyadic and Triadic Configurations*. Also, we can conclude that the passing network patterns involve one dyadic configuration and two triadic configurations.

Then we explore more deeply on (D1, M1, F2). According to the passing network, these three players make 227 passes in total, while the passes between these three account for 26.7%. Moreover, the *cluster coefficient C* of the three equals 7.31, much smaller than that of the whole team (7.62), which means the connectivity between the (D1, M1, F2) is much higher than that between the three and other players. In this case, it's more likely to form clusters just like these three players.

3.3   Other Structural Indicators

3.3.1   Time Scale

In a gesture to explore other structured indicators to evaluate the ball passing network, we next analyse the time scale. As small as every minute and every second, and as large as the entire game and season, the *time scale* can undoubtedly reflect the team's game status. In this part, we mainly discuss two aspects: *10-minute network passes* and the *entire game network passes*.

- **10-minute Network Passes**

When it comes to the relatively small period of time throughout the game, we have to introduce the *tempo* in football. Tempo, in a word, is the speed of any movement or activity, which mainly includes the speed of advancement, transmission, and running. Players in the field determine when to be fast and when to be slow through factors such as technology, vision, anticipation, tactics, and the actual situation of the game.

An important indicator of tempo is the number of passes. Team with more passes in a unit time are those generate more passes in less time, and indicates that they have a well capability of controlling tempo. In our case, we analyzed the effects of using different time intervals t of passes and chose $t = 10 \min = 600 \sec$ as a trade-off value.

At the same time, we take the 1st match as an example to study the number of passes of the Huskies every ten minutes, which is defined as α. Considering the interval between two half, we didn't talk about the number of ball passes between 40 - 45 min in both half.

Figure 4: 10-minute Network passes of Huskies in the 1st match

As is clearly depicted in **Figure 4**, Huskies' 10-minute Network passes is relatively stable. From the timeline, we can find that in about 30 minutes, the Huskies changed the tempo, the attack speed accelerated, and the rhythm showed a simple and bright. In the last 10 minutes, the Huskies were relatively conservative and had fewer passes, and we can attribute it to exhaust.

- **Entire Game and Entire Season Network Passes**

In this part, we explore the passing network at the level of the whole game. Using the ball passing matrix and equation (2), we get the following **Figure 5**.

Six matches are chosen from different period of the whole season. As is shown in **Figure 5**, we can conclude that at the beginning of the season, the teamwork performance is *unstable*.

However, As the season progresses, the Huskies' performance has gradually stabilized, and a stronger connectivity between the players has also formed.
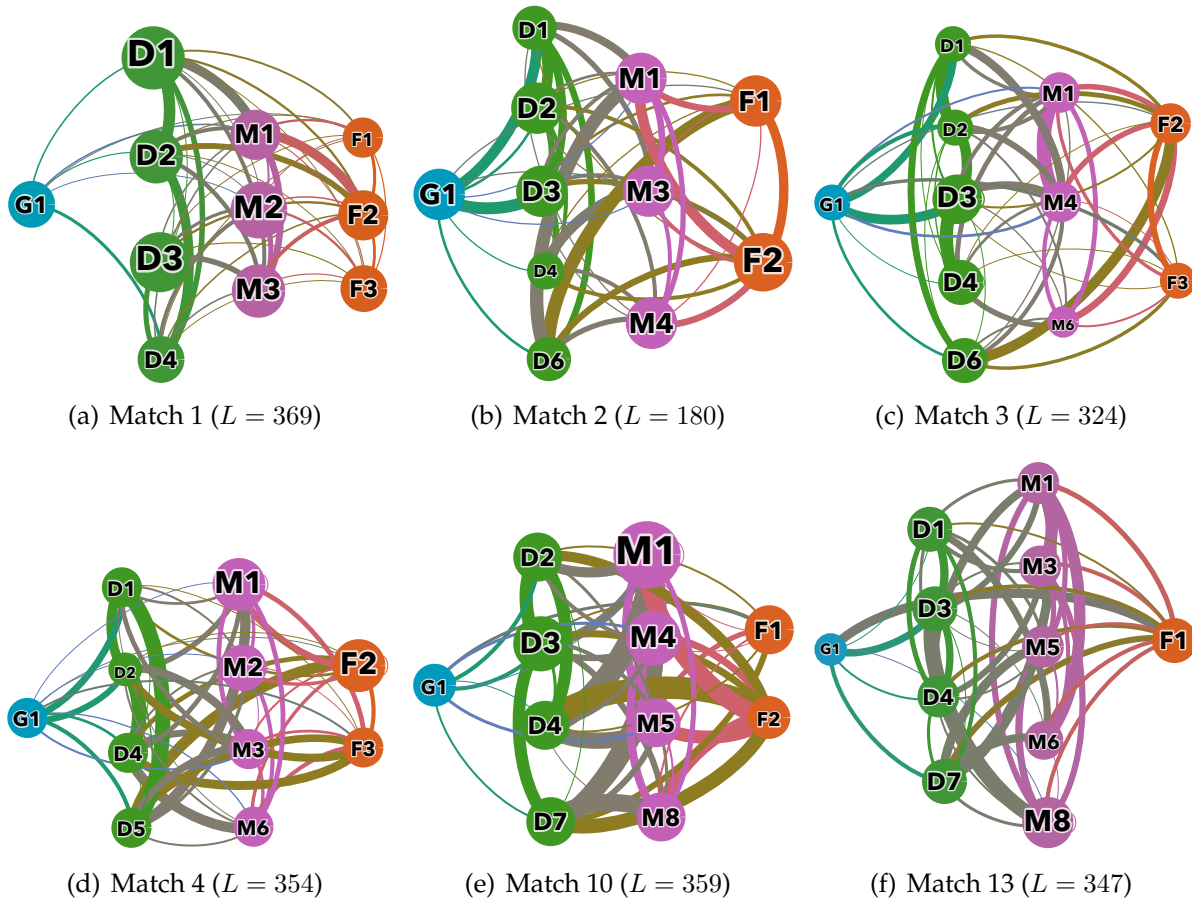
(a) Match 1 ($L = 369$) (b) Match 2 ($L = 180$) (c) Match 3 ($L = 324$)

(d) Match 4 ($L = 354$) (e) Match 10 ($L = 359$) (f) Match 13 ($L = 347$)

Figure 5: Entire Game and Entire Season Network Passes

3.3.2   Region Scale

In spite of time scale, we can also analyze the passing network by region scale. The football field can be divided into eight regions averagely. Then we use the location where each passes occur to have an overview.

Figure 6: An overview of the passes in football field
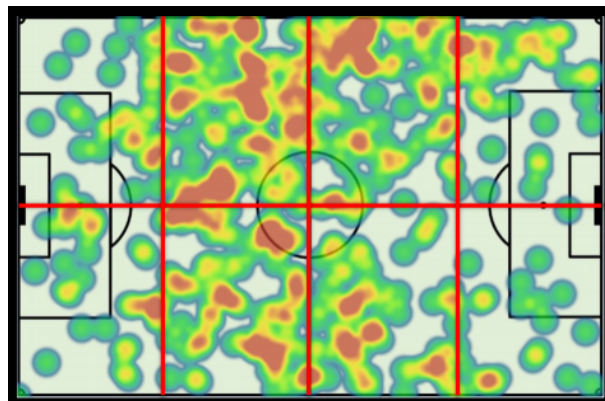
As is shown in **Figure 6**, Huskies make their most passes in the backcourt, which is in line with their highlight in defense. Moreover, the connectivity between defense players is higher than that of forward players. However, **Figure 6** also shows that Huskies hasn't made full use of the width of the field. They can try to make more passes in those regions with light color.

4   TASK 2:   Evaluation of Successful Teamwork

4.1   Identification of Performance Indicators

To further evaluate whether the teamwork is success, we should select some relevant indicators. This paper simplifies the performance indicators into two aspects: *Static* and *Dynamic*. The *Static* performance indicators mainly contain: *Coordination among players*, *Distribution of contributions*. And the *Dynamic* performance indicators involve: *tempo*, *flexibility*, *Pressing*.

4.1.1   Static Layer

- **Coordination among players**

In terms of coordination among players, we use *cluster coefficient C* to quantify it. One player's cluster coefficient means his connectivity with other teammates. The team's average cluster coefficient embodies the level of coordination in the team. Therefore, the higher average cluster coefficient C is, the stronger coordination γ is.

- **Distribution of contributions**

Considering the contribution distribution, we may consider the contribution of each player into two aspects, namely forward and defense. Set the number of shots per player i to μ_i and the number of defenses to ν_1.

Below we discuss forward and defense success rates. Forward success rate can be obtained by the ratio of the number of goals to the number of shots, which can be described as:

$$f = \frac{\mu_{\text{goal}}}{\sum_{i=1}^{11} \mu_i}, \tag{8}$$

Defense success rate can be obtained by the ratio of the number of defenses to the number of losses plus one, which can be expressed as:

$$d = \frac{\sum_{j=1}^{11} \nu_j}{\nu_{\text{loss}} + 1}. \tag{9}$$

Where we add one to the denominator of f to prevent the denominator from being zero.

Then we get a parameter φ about *Distribution of contributions*:

$$\varphi = f \cdot d. \tag{10}$$

Where Larger φ means fewer mistakes and more success.

4.1.2   Dynamic Layer

- **Tempo**

In **TASK 1**, we have talked about tempo by means of *10-minute Network Passes*, which is the number of passes per ten minutes. However, when we discuss team performance evaluation, the performance of *10-minute Network Passes* is not very obvious,

we choose *50-ball Passing Time*[6]. We construct the *50-ball Passing Time* with the aim of accounting for the temporal evolution of the game. *50-ball Passing* contain only 50 consecutive passes and are assigned the time of the last of these passes.

The 50-ball Passing Time t_b is the time required to construct a 50 ball passes. It is obtained subtracting the time of the first pass of the network from the time of the last pass. Teams with shorter t_b are those that generate more passes in less time.

- **Flexibility**

In this section, we use the y-coordinate standard deviation σ_y to represent team *flexibility*. The better the team's flexibility, the greater the standard deviation σ_y, and the more the off-center disturbances occur. The flexibility can be marked as:

$$\sigma_y = \frac{1}{n} \sum_{j=1}^{n} (y_j - \bar{y}) \tag{11}$$

As is shown in **Figure 7**, the team with higher σ_y has more attack strategy, either from both sides or in the middle. In contrast, the team with smaller σ_y has fewer attack strategy. As a result, if the opponents focus on the defense of the middle, they will not be flexible enough to change their attack strategy like attacking from left or right.

(a) Match with higher Flexibility $\sigma_y = 33.70$ (b) Match with smaller Flexibility $\sigma_y = 26.66$
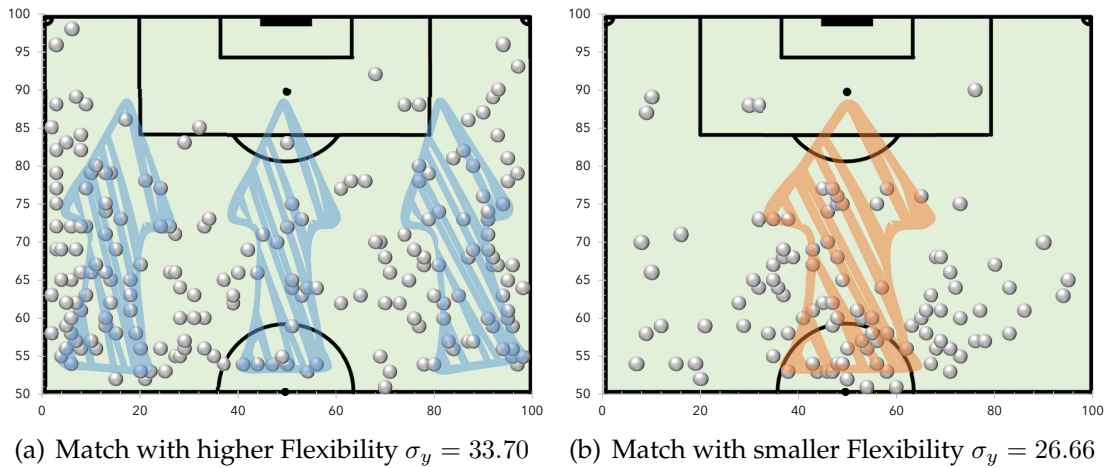
Figure 7: Comparison of matches with different flexibility

- **Pressing**

Pressing is an indicator of offense and defense status. *Gegenpressing* refers to the premise of the entire defense of the team, forcing the opponent to rush the ball from the opponent's backcourt. Pressing can be estimated by the expectation of the x-coordinate.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{12}$$

When $\bar{x} = 50$, we can consider that there is no gegenpressing and the situation on the court is balanced; when $\bar{x} > 50$, there is a situation of gegenpressing on football field.

Figure 8: Situation of Gegenpressing on football field

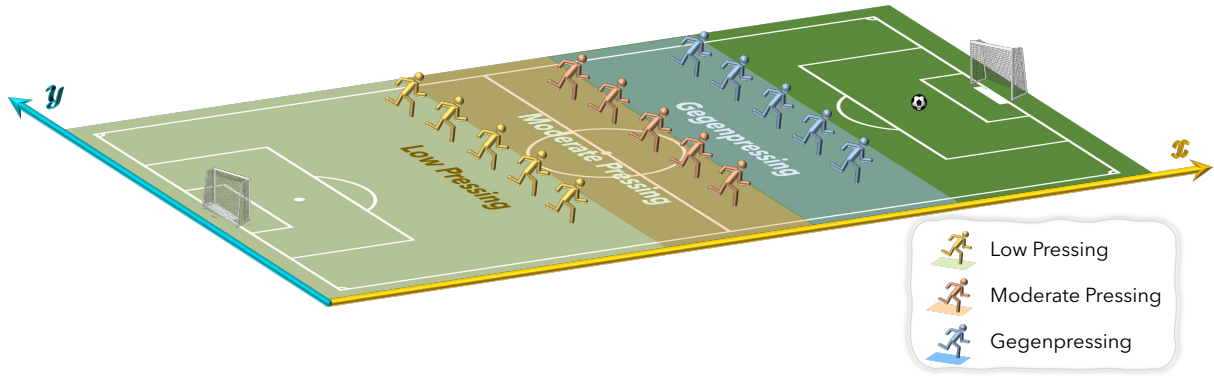## 4.2 Evaluation model based on Adversarial Regression

We describe the performance of a team S during a football game by five network measures extracted from its passing network: x_1 (*Coordination among players γ*), x_2 (*Distribution of contributions φ*), x_3 (*Tempo t_b*), x_4 (*Flexibility σ_y*), x_5 (*Pressing $\bar{x}$*). We compute the *five* network measures of teams for every game and observe a correlation among the proposed network measures and the success of a team during the competition, and therefore set up a simulation experiment to validate our approach.

We simulate the result of the match i of the Huskies (or the opponent) through the following five steps:

***Step* 1:** For each of the two teams, we compute the score S of previous performances of the team separately. (S_1 for the Huskies, and S_2 for the opponent)

$$S_i = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 \quad (i = 1, 2) \tag{13}$$

***Step* 2:** Compare the predicted measures of the two teams. If the Huskies win, $S_1 > S_2$; if they lose, $S_1 < S_2$. When the two teams draw, there should be $|S_1 - S_2| < 1$. If not, turn to ***Step* 1**.

***Step* 3:** Through the former 2 steps, we can obtain all the possible solutions. Set the standard deviation of Huskies' all matches as error value `LowestError`.

***Step* 4:** We choose the weights with `LowestError` as the regression equation.

So far, we have established our *Evaluation model based on Adversarial Regression*. The main reason is to introduce the opponent's confrontation factors, which makes the model more fair in evaluation and accords with the actual situation.

4.3 Application of Evaluation Model

- **Data collection**

In order to calculate the *weight β_i* in equation (13), we take the first **33** football matches of this season as an example, as our *training dataset*, and the last **5** matches as our *test dataset*, as shown in **Table 2**.

In our statistics of 33 matches, we cleaned and denoised the data. Among them, the 31st and 33rd matched had certain problems. We artificially adjusted the parameter

Table 2: Identification of performance indicators in 33 matches

ID	Coordination		Distribution		Tempo		Flexibilty		Pressing		Result
	H[***]	O	H	O	H	O	H	O	H	O	
1	7.61	4.00	62.25	12.60	539.50	1476.57	32.42	28.58	42.37	51.23	win
2	2.88	8.25	28.71	11.17	1019.95	692.49	29.30	27.34	46.36	49.81	tie
3	5.10	9.00	8.43	34.17	634.78	596.98	27.77	27.03	40.43	52.02	loss
4	6.01	5.66	6.14	62.40	760.01	934.64	32.74	31.94	44.14	51.62	loss
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
30	5.61	3.00	62.25	7.13	707.96	2018.52	30.94	30.18	50.75	58.75	win
31	6.58^*	4.21	20.10	13.87	986.98	1014.37	32.53	26.66	44.20	50.13	win
32[**]	/	/	/	/	/	/	/	/	/	/	/
33	3.75^*	3.35	39.83	21.90	1159.30	479.62	29.43	27.61	51.34	51.20	tie

[*] We artificially adjusted the parameter value which the value is *italic*.
[**] For too large deviation, we discard the 32nd match's data.
[***] 'H' for the Huskies, and 'O' for the opponent.

value of *Coordination among players* γ. The deviation of the 32nd match's data was too large, and we rounded it off. And then we normalize the remaining data so that its distribution satisfies *zero-mean* and *unit variance*.

- **Training algorithm**

For the data of 32 matches (*rounded off the 32nd match*), we use a *Greedy Algorithm* to experiment and calculate each possible weight β_i in the model. Then we put the possible weights into the model, and put the data of 33 matches of Huskies into the evaluation system to score, calculate the standard deviation sum of 33 games, and use it as the `LowestError` of the evaluation model.
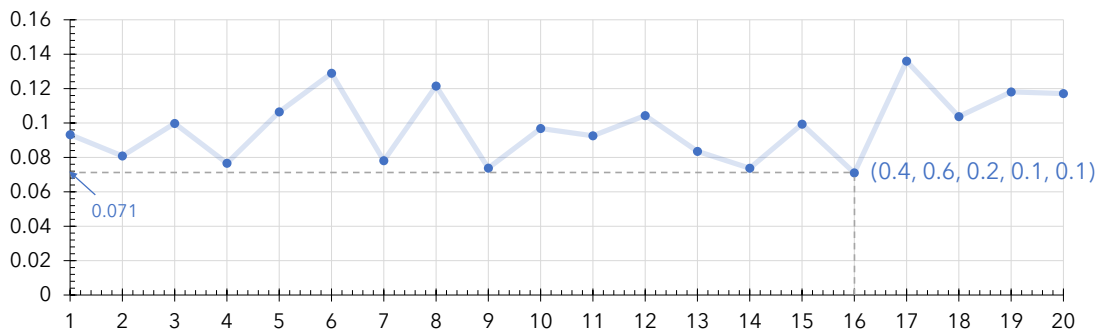

Figure 9: `LowestError` of each possible weight

According to the training of the *Greedy Algorithm*, there are 20 groups of weight coefficient combinations that meet the conditions. We calculate their `lowestError` respectively. As shown in **Figure 9**, the group with the lowest `lowestError` is selected as the final simulation result weight value $\hat{\beta}_i$.

- **Simulation results and accuracy test**

According to our supervised observations of the model, the model can keep one decimal place for the weight value of each indicator. Finally, we put the last 5 matches into the test, and we can conclude that the model is valid. The model considers the

regression formula for successful teamwork to be:

$$S = 0.4x_1 + 0.6x_2 + 0.2x_3 + 0.1x_4 + 0.1x_5 + C \tag{14}$$

i.e.

$$S = 0.4\gamma + 0.6\varphi + 0.2t_b + 0.1\sigma_y + 0.1\bar{x} + C \tag{15}$$

Where the minimum error in the optimal weight `lowestError` $= 0.071$, is valid for both data fitting and model regression. Through this model, we can score the team for each or teamwork for the entire season, and compare it with opponents, thus making the model adaptable.

From the regression equation (15) we get, we can have a general idea about the importance between the five indicators.

- The most significant factor is the *Distribution of contributions* as predicted, because it is linked with the final result directly. There is no doubt that a team, where every teammate contributes to the teamwork in his own position efficiently, has the strongest energy.

- Next is the *Coordination among players*, which is not difficult to understand.

- Then comes the *flexibility*, *tempo* and *pressing*, which has relatively small weight, however can't be ignored.

4.4 Discussion

Depending on the model we constructed above, we can evaluate the success level of teamwork. The higher the value of performance indicators are, the higher the score S is. However, in the model discussed above, we haven't taken the opponent into account. It's of great significance to clarify whether strategies are universally effective or dependent on opponents' counter-strategies.

From the Evaluation Model we built in this section, the weights of different indicators are ascertained. It seems that as long as Huskies insists its strategy, which can be embodied by the five indicators, the score remains the same. However, this thought hasn't taken opponents' counter-strategies into consideration. We take the pressing as an example and draw the following **Figure 10**.
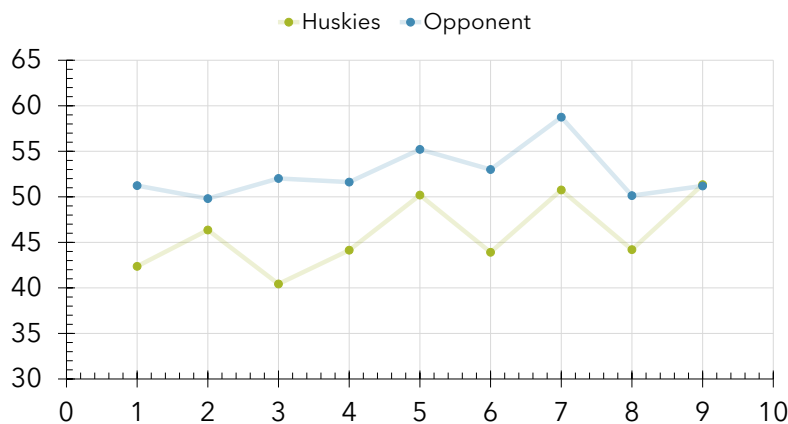

Figure 10: Comparison of Pressing

From the *Evaluation Model* we built in this section, the weights of different indicators are ascertained. It seems that as long as Huskies insists its strategy, which can be

embodied by the five indicators, the score remains the same. However, this thought hasn't taken opponents' counter-strategies into consideration. For example, if the opponent's strategy makes the degree of pressing high, which means raise the $\bar{x}_{\text{opponent}}$, the $\bar{x}_{\text{Huskies}}$ of Huskies will decrease accordingly. That's to say, the value of indicators will be affected by opponents' strategy, thus influencing the score. Therefore, the strategy is not universally effective. Only by adjusting the strategy based on the opponents' strategy, trying to make the total score higher or repress the score of opponents, can one team make the likelihood to win higher.

5　TASK 3: Advice on Improving Team Success

5.1　Structural Strategy Analysis

After statistics of this season, the Huskies won **13** matches, lost **15** matches and tied **10** matches in 38 games. For the team, the year-end data of this season is not ideal. In the next season, the team needs to make a huge change.

We have conducted a detailed analysis of the team data this season. In terms of *team formation*, the formations used by different coaches are similar. From the data of 38 games, the main formation 4-3-3, 4-4-2, 5-3-2, each game basically changes between these three formations. In general, we use 4-3-3 to represent the team's offense more, 4-4-2 to represent a conservative formation, and 5-3-2 to represent a more defensive formation.

Below we analyze the five performance indicators mentioned in **TASK 2** to determine the tactical system suitable for the current Huskies team. From the five indicators, we selected *Coordination among players* (an indicator of teammate teamwork), *Flexibility* (an indicator of offensive line diversity), and *Pressing* (an indicator of team offense and defense status) to explore the three formations. Fitness for the team. We use indicators to determine the fitness of the three formations for the team separately.

Table 3: Fitness of the three formations for the Huskies

Formation	Coordination	Flexibility	Pressing
4-3-3	5.1043	32.42	42.37
4-4-2	6.0104	40.88	43.58
5-3-2	7.5032	23.50	49.67

As is shown in **Table 3**, the most suitable tactical formation for Huskies last season was 5-3-2.

- In this defense-based formation, the team's teamwork and offensive oppression are the highest, and it is the most effective strategy for now.

- However, this form of attack is not flexible enough, and the oppression is also very poor. It is almost oppressed by its opponents in its own half.

So for Huskies at present, it is urgent to improve the offensive ability to win the match.

In summary, the Huskies need to give play to the characteristics of *good defense*, and strengthen the team's offensive capabilities. There are many classic football strategies,

such as *Tiki-Taka* (Barcelona), and *Total Football* (Netherlands). Considering the actual situation of the Huskies, the most effective strategy is *Defensive counterattack*.

5.2 Comparison with *Italy National Team*

There are many teams that use defensive counterattacks in football, such as Real Madrid and Atlético Madrid. Here we choose Italy National Team which play the defensive counterattack to the extreme peak, and selected the classic battle - *2012 European Cup semi-final* Italy v.s. Germany match data[7], and compared with average match data of the Huskies in this season.

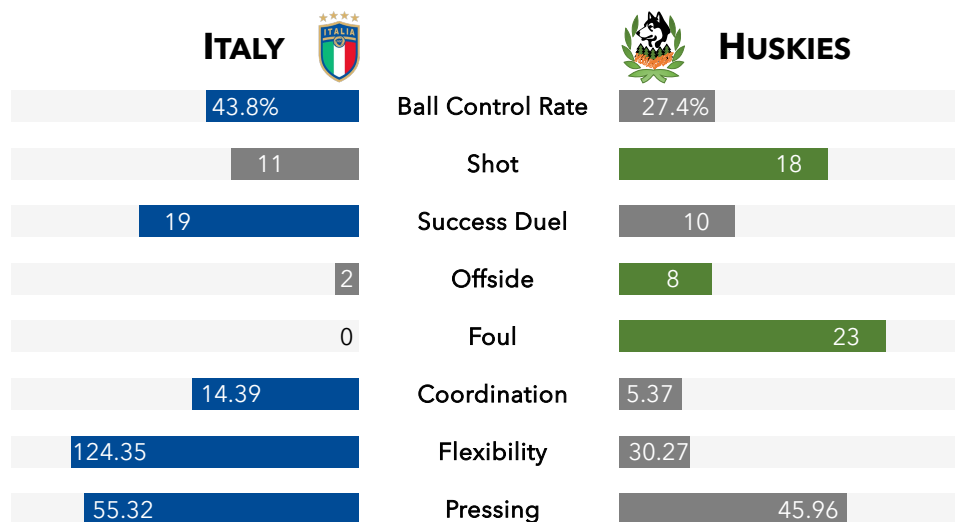| ITALY		HUSKIES
43.8%	Ball Control Rate	27.4%
11	Shot	18
19	Success Duel	10
2	Offside	8
0	Foul	23
14.39	Coordination	5.37
124.35	Flexibility	30.27
55.32	Pressing	45.96

Figure 11: Comparison with Italy National Team

From the data in **Figure 11**, we can find that when facing opponents that are stronger than themselves theoretically, such as Italy (FIFA World Ranking 14th) v.s. Germany (FIFA World Ranking 1st), whether it is in the case of *Ball control rate*, or *shot* is not dominant, still won the match 2-1. The winning experience of the Italy is the *flexible change* of formation and the effective *counterattack strategy*[8].

Below we comment on Huskies' performance in this season.

- The value of Coordination has been low, the team's cooperation lacking, resulting in a large number of *Offside* and *Foul*, unable to play a threatening cooperation.

- The number of Flexibility is not high enough. The team's offensive line is actually more simplistic. The tactics of the midfield straight out of the penalty zone cannot guarantee the goal rate and attack efficiency. The Huskies must choose to start from a wide open area.

- The Pressing value shows that the team is too conservative in defense. Most of the time it has been active in its own half, staying away from the opponent's goal, it is difficult to pose a threat, which reflects the lack of offensive organization ability of the team.

In summary, in the change of next season, we should mainly consider the three indicators of *Coordination*, *Flexibility* and *Pressing*, improve the diversification of the team's offensive methods, and defensive counterattack with efficient and threatening organizations.

5.3   Advise on next Season

After analyzing our team's network and strategy this season, in order to achieve successful teamwork and better team performance next season, the changes that the team needs to make now are:

Change 1: *Adjusted the formation to a more flexible formation*

From the data of this season, we can see that the team is good at *defending* this work, but the cooperation is not enough. So we will choose 4-1-3-2 as the team's main formation next season. It is flexible and can be evolved into 5-3-2 and 4-3-3.

Change 2: *Highlight the importance of core players*

In team cooperation, we must make the best use of every character in the team to maximize their value. From the situation of last season, M2 and F1, F2 should be the core points of the offense. When M2 holds the ball, it should The deformed formation is 5-3-2 or 4-3-3 to protect M2 players from holding the ball. F1 and F2 should find more neutral and running positions in the frontcourt to respond to the midfield pass.

Change 3: *Coherence of possession and defense*

The key to the 4-1-3-2 formation of ball control is not multi-ball control, but to control the ball on the center line of the court as far as possible, where Advance attack, retreat defend.
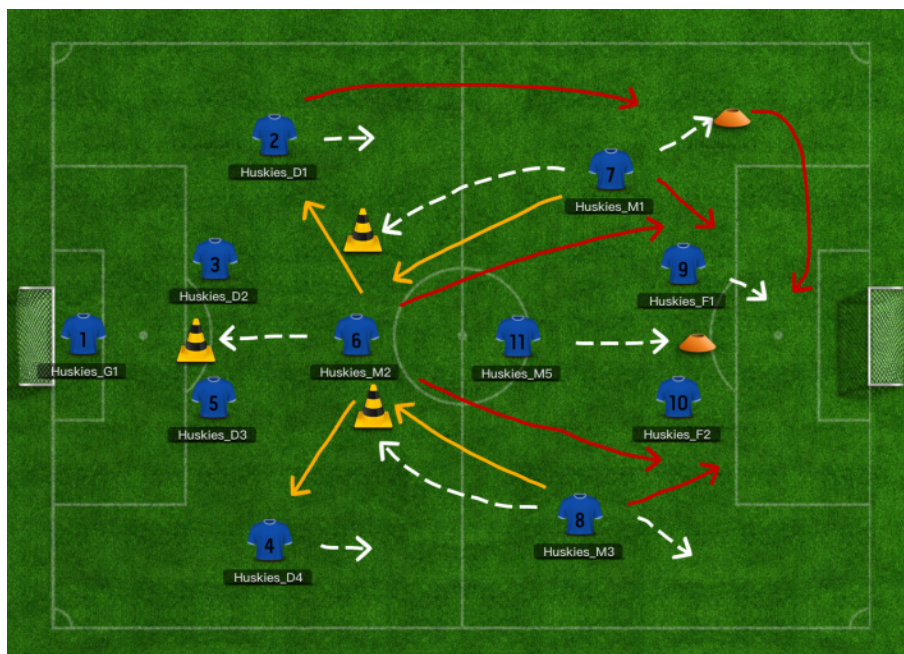

Figure 12: A schematic diagram of *Defensive counterattack*

As is shown in **Figure 12**, the *white dotted line* represents the player's moving route. Adapt to the player's moving route, the 4-1-3-2 formation can be changed to 5-3-2, 4-3-3, 4-3-1-2 and other formations according to the situation on the court or the opponent's strategy. The *red solid line* represents the team's pass line when attacking. The initiator of the attack is usually organized by M2. At this time, M1 and M3 need to back

up and protect `M2` to control the ball, or insert to create an offensive opportunity. The *yellow solid line* represents the passing route when defensive is needed or when there is no good offensive neutral, increases the ball control rate, and stabilizes the ball right near the center axis. The *yellow Warning sign barrel* represents the position that the returning player needs to return. The *orange flag bucket* represents where the offensive player can launch a threatening attack.

6 TASK 4: How to Design more Effective Teams

6.1 Generalization of our Model

In a gesture to design more effective teams, we can generalize our model to other practical teamwork. It is of great significance to understanding the complex set of factors that make some groups perform better than others. In this chapter, we make several analogies to the indicators we used in the team sport model.
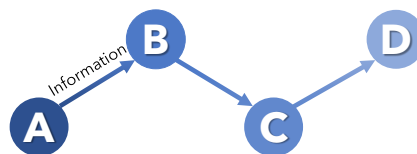
- **Coordination among teammates**

Figure 13: Coordination among teammates

Similar to the connection between players, the closeness of connection between colleagues is also an important indicator of teamwork. In terms of the coordination among teammates, we can use the informational passing distance d to quantify this parameter, which means the number of the teammates the instruction get through when it is being passed between the team. Just as **Figure 13** shows.

The distance between A and B is 1, while A is 3 away from C. Then we can get a similar network matrix like passing matrix. In this way, *cluster coefficient* can also be used here to quantify the coordination among teammates.

- **Distribution of contributions**

For teamwork, a reasonable and appropriate division of labor is also an important indicator of cooperation efficiency. In teamwork, everyone should perform his or her own duties to promote strengths and avoid weaknesses.

Unlike football games where contribution can be easily divided into two aspects, offence and defense, the contribution in other teamwork is more complex. Each teammate has his or her own position and performance evaluation criteria. Therefore, for people in different position the quantification rules vary from each other.

- **Flexibility**

In teamwork, it is inevitable that we will encounter difficulties and setbacks. At this time, we need to be flexible to make changes and try boldly. The Huskies performed poorly with the 4-3-3 formation in the 9th match, but later quickly changed to reverse the situation.

With reference to the variety of attack strategy, we can make an analogy to the

standard deviation of x-coordinate value where each passing event happens. For each team, the final target may be unique, but the plan isn't limited. Consequently, we can use the standard deviation of the number of people dispatched to different plans to quantify the flexibility.

6.2 Other Aspects

Considering the complexity in other practical teamwork, we need to take more aspects into consideration to develop generalized models of team performance.

- **Ideal leadership style**

Each team has one leader, who help leverage the skills of all their teammates and make overall plan. A good leader can make teamwork more efficient. Therefore, it is important to evaluate the leadership and leadership style.

There are many leadership styles. A leader will lead several people directly or indirectly. We can use the distance to quantify the leader's leadership intensity I as the formula shows:

$$l = \sum_{i=1}^{N} \frac{1}{d_N}. \tag{16}$$

Where N is the total number of people the leader leads, and d_N is the distance between the person and the leader. The management scientists make out a leadership intensity standard[9]. When this leadership strength is exceeded, the leader will be *tired* in the management process. On the contrary, when the management intensity is too low, the managerial talents are not fully applied, which causes the idleness of human capital. Therefore, building a leadership model and choosing the ideal leadership style also play a dominant role.

- **Effective coordination over time**

The effective coordination can be expressed as follow:

$$E = \frac{\sum\limits_{i=1}^{N} t_{\text{work}}}{t_{\text{total}}}. \tag{17}$$

Where t_{work} is the total time one person working, and t_{total} is the total duration of the entire cooperation.

For example in football game, as one player steals the ball from an opponent, another player is poised for offense. Their total working time is definitely larger than the time the whole event takes, which means they have higher effectivity. In comparison, we can also use this method to quantify the effective coordination over time.

7 Analysis on Model's Sensitivity

7.1 Evaluation Model sensitivity to *Draw* situations

When studying the Evaluation model in **TASK 2**, we dealt with the situation when the two teams were tied by absolute difference. When the difference in absolute value

of evaluation score S is less than 1, we regard this situation as the indicator passed. Next, we will study the influence of the constraints on the *difference between the scores of the two teams (error limit)* on the *weight value β_i* in the model.

We let $|S_1 - S_2| < \varepsilon$, where error limit $\varepsilon \in [0, 2]$. We substitute the error limit into the model algorithm, calculate the value of each weight β_i under different error limit ε, and calculate the `lowestError`. Finally, we draw a visual smooth curve like **Figure 14**.
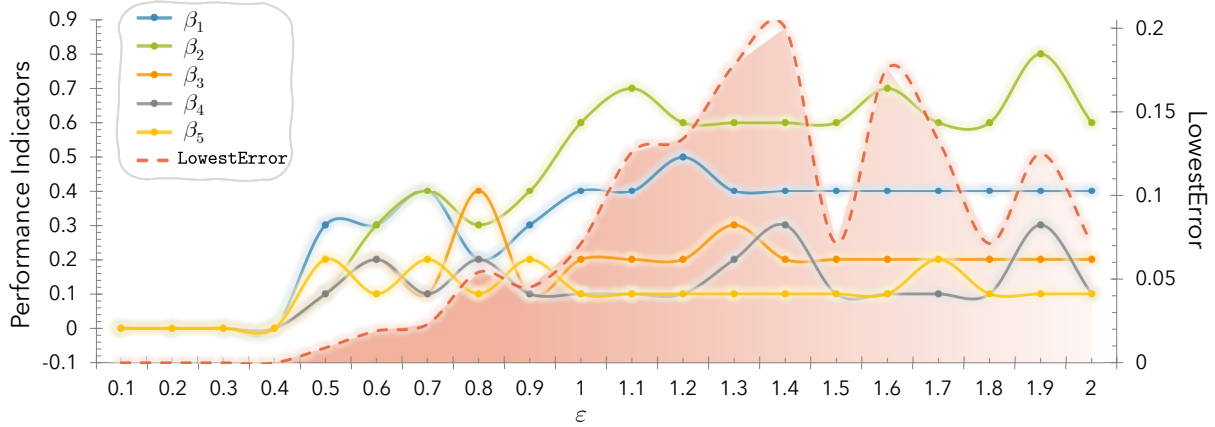

Figure 14: Five performance indicators and `lowestError` with different error limit ε

In **Figure 14**, we can see that when $\varepsilon < 1$, the `lowestError` has been kept at a low level. However, at this time the weights of the five indicators of the model have been in a chaotic state and cannot reflect the reality and the importance of the indicators. When $\varepsilon > 1$, the weight of each indicator is relatively stable. At this time, we must also consider the value of `lowestError`, and the deviation should not be too large. It can be seen that when $\varepsilon = 1, 1.5, 1.8$, it is more reasonable as a tie. So our model is more sensitive to the situation of a draw.

7.2   Analysis on performance indicator *Tempo*

In **Task 2**, the quantitative indicator for reflecting the Tempo on the football field is the time it takes for a player to pass 50 goals. Here we consider whether the time when players pass 10, 30, 50, 70, 90 balls can appropriately reflect the changing rhythm on the court. In order to make the time of different foot counts analyzable, we choose the same game time to prevent the situation where tempo cannot be compared due to different time intervals.

After our analysis of the data, we chose 46 s - 1305 s in the first half of the match. And we removed the *10-pass* and *90-pass* time because their time is too short or too long to analyze the sensitivity. Through calculation, we obtain the time curve of ball passing time per 30, 50, 70 passes during this period.

As can be seen from **Figure 15**, when calculating the *30-pass time*, the curve shows more jagged parts, that is, the tempo on the football field has a clear turning point. This is well understood according to the actual situation of the football field. When dividing the passing time, the tempo displayed on the court will not change much (i.e., it is difficult to change the defensive or offensive rhythm within 30 feet). When calculating the *70-pass time*, there is almost no change in tempo on the field, which is obviously not reasonable. The tempo indicator in our model is sensitive to the number
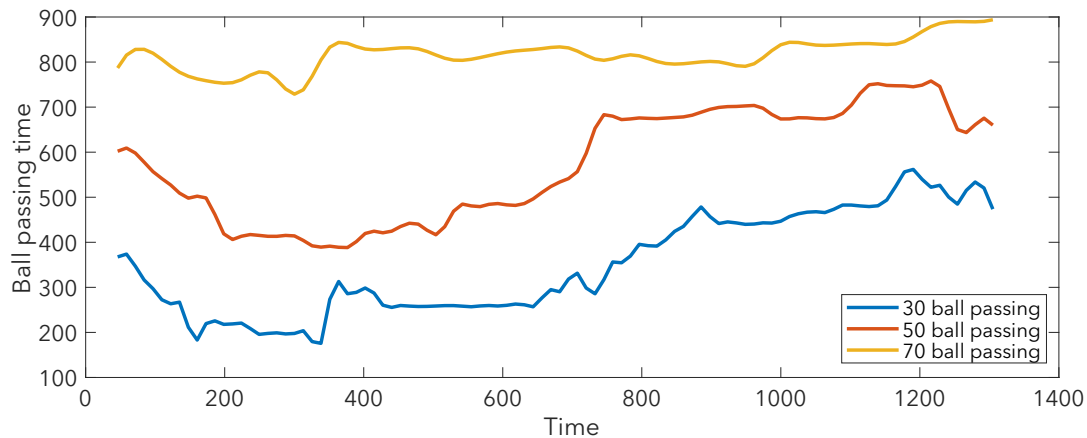
Figure 15: Ball passing time per 30, 50, 70 passes

of passes hence. This is why the *50-pass time* is the most scientific and reasonable to calculate tempo.

8 Strengths and Weaknesses

8.1 Strengths

- The selection of the network parameters of the passing network is scientific and reasonable. The network model is highly applicable and can reflect the characteristics of various aspects of the team and the evaluation scale.
- The evaluation model also introduces the opponent into the evaluation system. In order to be able to compare against the Huskies team, the weight coefficients of the five indicators are returned to be more scientific and reasonable.
- Our model analyzes the Huskies team's data thoroughly, finds more effective strategies for the team, proposes constructive changes, and visualizes our strategies.

8.2 Weaknesses

- Our model did not have enough analysis and processing for the eventsubtype part, and did not take out each team member for individual analysis.
- The indicators of the popularization model cannot be quantified, resulting in a low persuasibility of the model.

References

[1] Leighton, F. Thomson. *Introduction to parallel algorithms and architectures: array, trees, hypercubes.* 2014.

[2] Clemente, Filipe Manuel, et al. "General network analysis of national soccer teams in FIFA World Cup 2014." *International Journal of Performance Analysis in Sport* 15.1 (2015): 80-96.

[3] Dijkstra, Edsger Wybe. "A Note on Two Problems in Connexion With Graphs." *Numerische Mathematik* 1(1959):269-271.

[4] Ahnert, Sebastian E., et al. "Ensemble approach to the analysis of weighted networks.." *Physical Review E* 76.1 (2007).

[5] Wong, J. A. Hartiganm. A. . "Algorithm AS 136: A K-Means Clustering Algorithm." *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28.1(1979):100-108.

[6] Buldu, J. M., et al. "Defining a historic football team: Using Network Science to analyze Guardiolas F.C. Barcelona." *Scientific Reports* 9.1 (2019): 1-14.

[7] *Balotelli sends Italy past Germany.* (2012). Retrieved December 10, 2014, from `https://www.uefa.com/uefaeuro/season=2012/matches/round=15174/match=2003379/index.html`

[8] Sigari, Mohamad Hoseyn, et al. "Counterattack detection in broadcast soccer videos using camera motion estimation." *international symposium on artificial intelligence* (2015): 101-106.

[9] Abdelmahmoud Hassan Elsheikh. *Effect of Leadership Intensity on Integrating Some Formal and Informal Organizational Efforts for Community Development in Khartoum Province.* 2016.

Appendix: Code

TASK 1 - **matrix.m**

```matlab
clear,clc
w=zeros(11,11);
data=xlsread('MatchID1.xlsx');
[m,n]=size(data);

%% Calculate the number of passing balls between each player
for i=1:1:m
    w(data(i,1),data(i,2))=w(data(i,1),data(i,2))+data(i,3);
    w(data(i,2),data(i,1))=w(data(i,2),data(i,1))+data(i,3);
end
for i=1:1:11
    w(i,i)=0;
end
w
%% Calculate the Total Link
L=0;
for i=1:1:11
    for j=1:1:11
        L=L+w(i,j);
    end
end
L=L/2
%% Calculate the density of match
pho=2*L/(n*(n-1))
%% Calculate the topological distance
d=zeros(11,11)-1;
for i=1:1:11
    for j=1:1:11
        if w(i,j)~=0
            d(i,j)=1./w(i,j);
        end
    end
end
d
%% Calculate the matrix of distance
D=0;
for i=1:1:11
    for j=1:1:11
        if d(i,j)>D
            D=d(i,j);
        end
    end
end
D
%% Calculate the cluster coeffient of each player
Cw=zeros(11,1);
for i=1:1:11
    s=0;
    for j=1:1:11
        for k=1:1:11
            if i~=j && j~=k && i~=k
                Cw(i)=Cw(i)+w(i,j)*w(j,k)*w(i,k);
            end
            if i~=j && i~=k
                s=s+w(i,j)*w(i,k);
            end
        end
```

```
        end
     end
   Cw(i)=Cw(i)/s;
end
Cw
%% Calculate the cluster coeffient
C=0;
for i=1:1:11
    C=C+Cw(i);
end
C=C/11
```

TASK 1 - **passingtime.m**

```
clear,clc
time=xlsread('passingtime.xlsx');
t=[];t_s=[];
i=1;k=0;

%% Calculate the time of 50 passing ball
[m,n]=size(time);
while time(i,1)==1
    k=k+1;
    t(k)=time(i+49,2)-time(i,2);
    t_s(k,1)=time(i,2);t_s(k,2)=time(i+49,2);
    i=i+50;
    if time(i+49,1)~=1
        break
    end
end
% Judge the half part of the match
while time(i,1)==1
    i=i+1;
end
while time(i,1)==2
    k=k+1;
    t(k)=time(i+49,2)-time(i,2);
    t_s(k,1)=time(i,2);t_s(k,2)=time(i+49,2);
    i=i+50;
    if i+49>m
        break
    end
end

%% Calculate the number of passing balls in 10 minutes of the overall match
% [n,q]=size(time);
% t=10;
% i=1;p=[];k=0;
% while time(i,1)==1
%     if time(i,2)-time(1,2)>(k+1)*t*60
%         k=k+1;
%         if k==1
%             p(k)=i;
%         else
%             p(k)=i-sum(p);
%         end
%     end
%     i=i+1;
% end
% j=i;m=k;
% while time(i,1)==2
```

```
%      if time(i,2)-time(j,2)>(k-m+1)*60*t
%          k=k+1;
%          p(k)=i-sum(p);
%      end
%      i=i+1;
%      if i>n
%          break
%      end
% end
```

TASK 2 - **AdversarialRegression.py**

```python
import numpy as np
import pandas as pd

def writeTxt(result1, result2, fileName):
    f = open(fileName,'a')
    for ele in result1:
        f.write(str(ele)+' ')
    f.write(str(result2))
    f.write('\n')

def read_csv(filename):
    read_file = pd.read_csv(filename, header = None)
    read_file = np.array(read_file)
    return read_file

def panduan(w,re,vec_one,vec_two,m):
    flag=1
    for i in range(m):
        if re[i]==1:
            y1=np.dot(vec_one[i],w)
            y2=np.dot(vec_two[i],w)
            if y1<y2:
                flag=0
                break
        elif re[i]==-1:
            y1=np.dot(vec_one[i],w)
            y2=np.dot(vec_two[i],w)
            if y1>y2:
                flag=0
                break
        else:
            y1=np.dot(vec_one[i],w)
            y2=np.dot(vec_two[i],w)
            if abs(y1-y2)>1:
                flag=0
                break
    return flag

def lowestError(w,vec_one,vec_two,m):
    y=[]
    for i in range(m):
        y1=np.dot(vec_one[i],w)
        y.append(y1)
    return np.var(y)


vec_one=read_csv('Huskies.csv')
vec_two=read_csv('Opponent.csv')
re=read_csv('result.csv')
```

```python
fileName = "var_y.txt"
m=vec_one.shape[0]
k=0
for w1 in range(1,10):
        for w2 in range(1,10):
            for w3 in range(1,10):
                for w4 in range(1,10):
                    for w5 in range(1,10):
                        w=np.array([w1,w2,w3,w4,w5])*0.1
                        print(w)
                        if panduan(w,re,vec_one,vec_two,m):
                            s=lowestError(w,vec_one,vec_two,m)
                            k=k+1
                            writeTxt(w,s,fileName)
                        else:
                            w=[]

# for j in range(20):
#     for w1 in range(1,10):
#         for w2 in range(1,10):
#             for w3 in range(1,10):
#                 for w4 in range(1,10):
#                     for w5 in range(1,10):
#                         w=np.array([w1,w2,w3,w4,w5])*0.1
#                         print(w)
#                         if panduan(w,re,vec_one,vec_two,m,j):
#                             s=lowestError(w,vec_one,vec_two,m)
#                             k=k+1
#                             writeTxt(w,s,fileName)
#                         else:
#                             w=[]

print("sum of result:")
print(k)
```