For office use only	Team Control Number	For office use only
T1 ________________	**2016386**	F1 ________________
T2 ________________		F2 ________________
T3 ________________	Problem Chosen	F3 ________________
T4 ________________	**D**	F4 ________________

2020
MCM/ICM
Summary Sheet

Teaming Strategies v6.2.2

Summary

In this paper, we proposed a passing network which is composed of adjacent matrix and topological matrix, and use it to recognize patterns appeared in the team through a match.
Next, we provide several structural indicators calculated from above passing network and puts forward some analysis of indicators' influence on the team's performance.
Later a model of neural network is utilized. The network is trained by structural indicators of team Huskies and its opponents , which work as the main feature of the team's adjacent matrix. Thus, the network can work out the possibility of the Huskies to win the match.
What's more, in order to find out the best passing network structures, we combined the model of neural network(NN) with evolution algorithm(EA). Neural network helps evaluate the fitness of adjacent matrix in evolution algorithm, making comparisons between parents and children. Evolution algorithm works out the best adjacent matrix through the iteration of variation, combination and decision.
In the model, adjacent matrix of children and parents is regarded as two inputs of neural network. With neural network working out the win rate of child over it's parent, we can readily judge which one is better. Thus, we could make use of both models' advantages and produce the best team structures for Huskies and help them make improvements.
Last but not least, with flexibility and adaptability, the model can also be used in various areas to find out the most efficient team structures.

Keywords: passing network; differential evolution algorithm; neural network; pattern recognition.

Contents

1 Introduction

1.1 Restatement of the problem

As science continually developing, many works, especially academic tasks which need many people's enrollment become more and more complex. As a result, teamwork become more and more important nowadays. In order to understand what makes teamwork more effective, we need to search deeply for the accurate role a person act and more importantly, the connection between members.
Today, with the advancing technology of artificial intelligence, automatic control system and mathematical modeling, methods of analyzing relation network have been applied to various industry and events.
Therefore, to understand factors which lead to better teamwork and the co-relation between members of a team, the usage of network pattern would have guiding significance.
If given enough data of the passing events of a soccer team, a passing network could be built up along with many aspects of analysis. In this paper, using the given data of soccer team Huskies in last season, our team build a passing network, work out a detailed analysis of strength and weakness of the team, then provide instructions for better performance and potential improvements.
Provided detailed data about team Huskies of last season, explore the interactions among the players and flexibility of the team, then work out specific strategies to help improve performance and teamwork.

The problems need to solve in this paper are:

1. Create a network for the ball passing between players, and use passing network to identify network patterns.

2. Identify structural indicators and network properties that reflect successful team work across the games.

3. Use your teamwork model to infer structural strategies and possible improvement for the coach of Huskies.

4. Generalize findings about how to design more effective teams and consider other aspects of teamwork which can develop generalized models of team performance.

1.2 Notations

1.2.1 Acronyms

EA: Evolution Algorithm
NN: Neural Network
BP: Back Propagation

1.2.2 Symbol definitions

symbols	definitions
w_{ij}	The number of passes from player i to player j
l_{ij}	The topological path from player i to player j
p_{ij}	The topological shortest path from player i to player j
N	The number of outfield player
$w[(i,j),k]$	The number of passes from player i to player j until k_{th} pass
$d_k(i)$	predicted value
$y_k(i)$	true value

2 Assumptions and Justifications

- Given that J. M. Buldú's conclusion[1] on the importance of passing data towards the result of a match, we mainly use the passing characteristics, such as total passing volume or density to represent the competitiveness of a team.

- We consider shots as the result of a successful offense for the reason that goals are partly occurrent events which is much more rare and highly fluctuating.

- In this model, we assume that all 30 players in team Huskies have the chance to pass the ball during a long enough time, which means that adjacent matrix is 30*30.

- According to the real game's content, the sum of all passing events in one single game is no more than 800.

3 Model Theories

3.1 Model Overview

Firstly, with the given data of soccer team Huskies in last season, a weighted adjacency matrix can be acquired. Using the matrix and the data files given, several performance indicators can be derived from the weighted adjacency matrix and passing network of players is created. From passing network we can readily infer the role of each player and their impact to the team.

Secondly, those performance indicators are used as the input of neural network. In this way, given enough data, neural network can be well trained to assess the performance of team and thus give reliable prediction of win rate.

Thirdly, with a reliable model of neural network which can predict win rate if given performance indicators, we can use it as a tool for comparing. In our research, differential evolution algorithm(EA) is invoked and neural network work as a selector during evolution process. Using EA we can finally work out a set of optimized strategies for the soccer team and therefore offer possible suggestions for the coach.

3.2 Construction of passing network

3.2.1 Utilization of data set

A single pass event includes person pass the ball, person get the ball, original and final location of football and time that event happened. A short piece of data is shown below:

With this collection of passing events, we can create adjacent matrix according to matches, times, or passes. An adjacent matrix show how many times of passes is delivered from player A to player B during a match, in a period or within a certain passes.

3.2.2 explanation and demonstration

With adjacent matrix and the average location of players a passing network could be therefore constructed. Figure x shows the passing network of team Huskies, including network of whole season, single match and a certain 50 passes:
(here are passing network images)
In a passing network, nodes represent players and arrows represent passes. A larger node indicates more passes to and from the player, a bigger arrow repre-

sents more passes from player A to player B, and the location of node shows the average coordinate of player.

3.2.3 network analysis

- With passing networks, the importance of players and connections between players is concrete. We can infer that the squad array which Huskies most often used is '541' for the reason that the passing network of season shows that among 11 of 30 players who passed ball most frequently, there are 5 defense backs, 4 middle fields and 1 forward. It indicates Huskies is a defensive team.
- Comparing 3 different passing networks, we can readily tell their variations, which means they have multitude strategies toward different cases and denotes that Huskies has promising dynamics to a certain extent.
- With such a squad array, Huskies need a very vigorous and powerful forward who can run fast and cover a wide range of offence and defence. Meanwhile, the strategy of Huskies should focus on defence and counter attack.
- However, Huskies' strategy put to much pressure on players M1, M3, D1, D3 and F2. Moreover, consider F2 is the only lead forward in Huskies, there lies huge risks if F2 once get injured or is absent.

3.3 Recognizing the configuration and formation

After constructing the passing network, we can use it to recognize the passing configuration and team formation. First off, we identify the team formation of team Huskies through each player's average location where he passes the ball. After that, we evaluate the closeness of the connection between two players for the recognition of team Huskies' configurations. We use "connection score"CS[(i,j)] which inspects the passes in a macro perspective to represent the connection closeness between player i and player j in a match. Considering the passes volume between two players differs greatly, we use a threshold value to evaluate the effective connection between two players. If the sum of weight in the dual ring is larger than the threshold, we suppose that it build a effective connection. In this case, we set it as the average sum of weight of all dual rings. Besides, the connection point is also related to the duration they build a effective connection as well as the distribution of connection points of other dual rings.

Moreover, since the the timeline of a match is quite long, those dual rings which reaches a threshold at a very early time will maintain its effectiveness during a quite long period or even until the end of game and get a much higher score.

However, we know that a pass's influence to the result of a game can only last for relatively short time. Hence we cut the passing data to many pieces, each of which contains 50 passes, reconstruct the passing network in every pieces, calculate the score of the dual ring in every pieces of passing network, and sum them up for the final connection score CS[(i,j)]. Consequently, we get that:

$$CS[(i,j)] = \Sigma_{pieces}\Sigma_{k=1}^{49}\frac{w[(i,j),k] - \overline{w[k]}}{\sigma[k]} * T[k] \tag{1}$$

Since (i,j) and (j,i) correspond to the same dual ring, we suppose i<j by default. $w[(i,j),k]$ represents the sum of its related dual ring weight, if builds or changes, when the kth pass is added to the passing network. $\overline{w[k]}$ represents the average sum of weight in all dual rings until when the kth pass of a piece of the passing network is added to the adjacency matrix. $\sigma[k]$ represents the standard deviation of sum of weight in those dual rings. $T[k]$ represents the duration between the kth and (k-1)th pass. Note that because of the presence of only 49 intervals in 50 passes, k only range from 1 to 49.

In addition, we use "largest weight"LW[(i,j)] which inspects the passes in a micro perspective to represent the connection between player i and player j during a relatively short period of time. Similarly, we cut the passing network into pieces which contains 50 passes respectively. In this case, we calculate every two-tuple's maximum sum of weight in all pieces of passing network. Therefore, we get that:

$$LW[(i,j)] = max_{pieces}w[(i,j),49] \tag{2}$$

w[(i,j)] is the sum of weight in every dual ring after every passes in one piece of passing network is added to the adjacency matrix.

3.4 Set Structural indicators

Our team mainly study 4 different structural indicators, all of which are derived from the adjacent matrix. To maintain its unity, we calculate them in every match. Besides, to learn more about the relation between the variation trend of these 4 factors and successful offense which usually ends with a shot or goal, we construct the passing network and calculate these factors every 50 passes. The description of all indicators are as follows.

3.4.1 Clustering coefficient

In graph theory, a clustering coefficient is a measure of the degree describing the tendency for which nodes in a graph cluster together. Considering that, we measure player i weighted clustering coefficient Cw(i) in:

$$C_w[i] = \frac{\Sigma_{j,k}w_{ij}w_{jk}w_{ik}}{\Sigma_{j,k}w_{ij}w_{ik}} \tag{3}$$

where j and k are other 2 players in the team and w_{jk} refers to the number of passes between player j and player k. Finally, the output cluster coefficient of a team is the average cluster coefficient over all players in this team:

$$C = \frac{1}{N}\Sigma_{i=1}^{N} C_w[i] \tag{4}$$

As shown above, C is greater when link weight grows larger when two players interact more frequently, which means that two players interact more frequently.

3.4.2 Shortest path length

In graph theory, the shortest path length is the minimum sum of weight from the original node to the destine node. However, since the passing network is weighted as the number of passes, we have to use the topological length $l_{ij} = 1/w_{ij}$ as the link weight from player i to player j of this directed graph. Using Dijkstra's algorithm to calculate the shortest path length between every two players in the team, we define the average weight of every shortest path between two nodes as:

$$d = \frac{1}{N(N-1)}\Sigma_{i,j,i\neq j} p_{ij} \tag{5}$$

where p_{ij} is the shortest path length from player i to player j and N is the number of outfield player in a team. It is obvious that as the team connect more closely when d decrease.

3.4.3 Largest eigenvalue of the adjacency matrix

The largest eigenvalue λ_1 of the weighted adjacency matrix W, which contains the information of passing network, is a measure of the network strength. The larger λ_1 indicates the firmer relation between players.

3.4.4 Algebraic connectivity

The algebraic connectivity λ_2 refers to the second smallest eigenvalue of the Laplacian matrix L which extracted from the adjacency matrix W using the formula:

$$\widetilde{L} = S - W \tag{6}$$

where W is the weighted adjacency matrix and S is a diagonal matrix whose i-elements are the sum of the passes sent by player i. With λ_2 growing, the overall connectivity of the whole graph improves.

3.4.5 Opponent factors

Since a win in a game not only depends on the players and teamwork of your team, but also depends on your opponent's performance. Therefore, we also calculate the 4 factors above of team Huskies' opponent using the same formula and restrictions.

3.5 Neural Network

Neural network here serves as the role of making judgements to a given network structure of Huskies and opponents , and output possibility of winning the game. The possibility ranges from 0 to 1. Thus, the problem here can be regarded as a binary classification problem.

The binary classification neural network model is constructed as follows(see figure 2) to acquire the possibility of beating the opponent:

- **Network**

  The network is defined as follows. We firstly use the input adjacent matrix to get the structural indicators as the input vector x.

$$x = [factor_1....factor_4, opfactor_1....opfactor_4] \tag{7}$$

  Next, labels(outputs) is defined as 0/1, where where 1 represents that Huskies wins, 0 represents loss.

$$y = 1/0 \tag{8}$$

  Now, according to figure1, the network's output is calculated as:

$$h_i^{(l)} = f(net_i^{(l)}) \tag{9}$$

$$net_i^{(l)} = \Sigma_{j=1}^{s_{l-1}} W_{ij}^{(l)} h_j^{(l-1)} + b_i^{(l)} \tag{10}$$

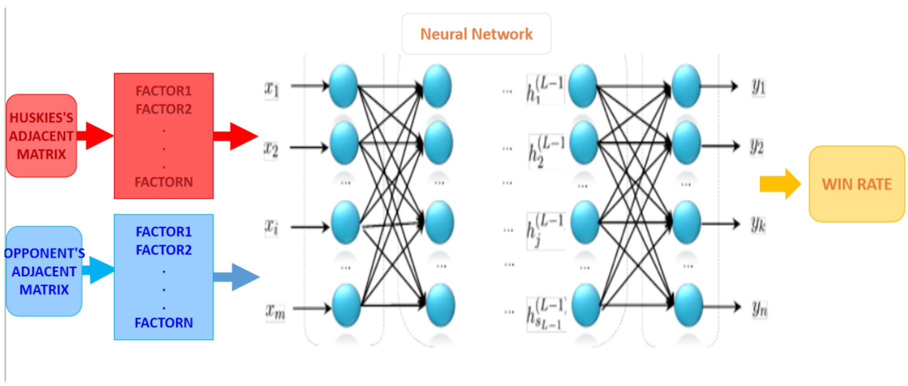


Figure 1: Neural Network Model

- **Activate Function**

  We choose Sigmoid as the activate function.

$$f(x) = \frac{1}{1+e^{-x}} \tag{11}$$

- **Back Propagation(BP)**

  We define Loss function as

$$E = \frac{1}{2m}\Sigma_{i=1}^{m}\Sigma_{k=1}^{n}(d_k(i) - y_k(i))^2 \tag{12}$$

  Using BP algorithm, we update the weight matrix W and bias b by each iteration.

$$W_{ij}^{(l)} = W_{ij}^{(l)} - \alpha * \frac{\partial_E}{\partial_{W_{ij}^{(l)}}} \tag{13}$$

  The BP process for output- hidden layer is now described as:

$$\delta_k^{(L)} = -(d_k(i) - y_k(i))f(x)'|_{x=net_k^{(L)}} \tag{14}$$

$$\frac{\partial E(i)}{\partial W_{kj}^{(L)}} = \delta_k^{(L)} h_j^{(L)} \tag{15}$$

$$\frac{\partial E(i)}{\partial b_k^{(L)}} = \delta_k^{(L)} \tag{16}$$

  Similarly,BP process for hidden-input layers is:

$$\delta_j^{(l)} = \Sigma_{k=1}^{s_{l+1}} W_{kj}^{(l+1)} \delta_k^{(l+1)} f(x)'|_{x=net_j^{(l)}} \tag{17}$$

$$\frac{\partial E(i)}{\partial W_{ji}^{(l)}} = \delta_j^{(l)} h_i^{(l-1)} \tag{18}$$

$$\frac{\partial E(i)}{\partial b_j^{(l)}} = \delta_j^{(l)} \tag{19}$$

  Therefore, weight matrix W and bias b will be updated in each iterations and the neural network can finally learn to make judgements of the given adjacent matrix.

3.6 Evolution Algorithm

In our model, differential evolution algorithm is evoked to help find a ideal model of adjacent matrix.
Differential evolution algorithm(EA) is a heuristic approach for minimizing(maximizing) possibly nonlinear and non-differential continuous space functions. The main process of the algorithm is shown below:

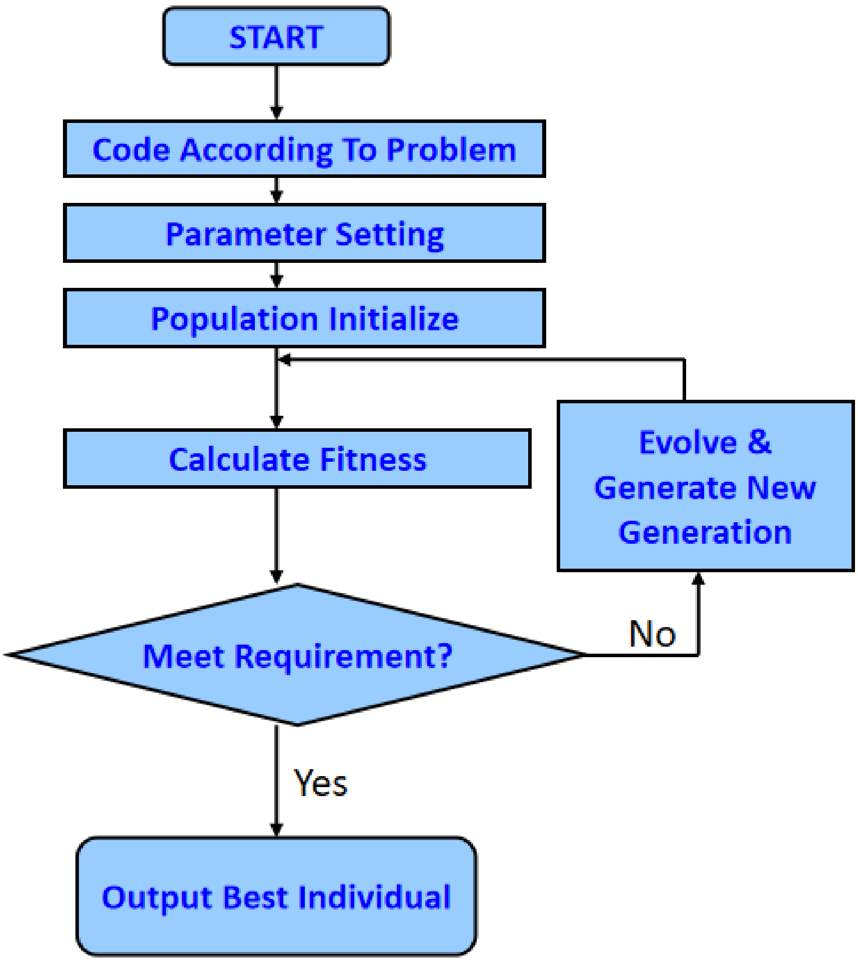


Figure 2: Main process of differential Evolution Algorithm

1. A set of models' parameter should be generated initially.
2. Methods of variation, recombination and selection are applied to generate variant parameters.
3. New set of parameters are compared with former ones, and only the better remains.
4. The process will repeat until parameters meet requirement or generation is maximum.

.

4 The Model Implementation and Results

4.1 Pattern recognition

Using the equation in 4.3, we can evaluate the players' connection in both micro and macro perspective.

Figure 3 show the passing network of team Huskies during the whole season. Every node not only represents a player, but also mark the mean position of this player in a match. These nodes' size are proportional to the total passes its representing player sent. The arrows means the player represented by the arrow's original node passes ball to the player represented by the arrow's destine node. The thickness of the arrows are proportional to the amount of passes. From this figure we can learn that team Huskies uses "541" formation as their fundamental formation, which can shorten the distance of passes at a cost of a requirement of more closeness to the midfield and farther distance to the opponent's goal.

Though share the same meaning of nodes and arrow edges as figure 3, figure 4 shows the temporary passing state as well as the team formation of team Huskies in a match. From this figure we can conclude that team Huskies used a "433" formation in this match. We can also know that during this match, player D1, D2, D3, M1 and F2 have a relatively strong connection between them, which indicates that in this game, team Huskies has a greater tendency attacking the right side of the opponent.

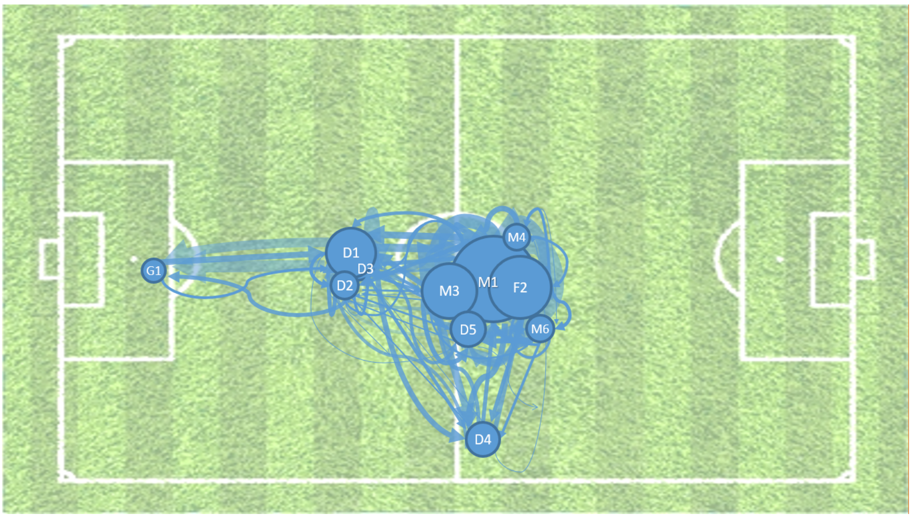


Figure 3: Passing Network of Starting 11 in a Season

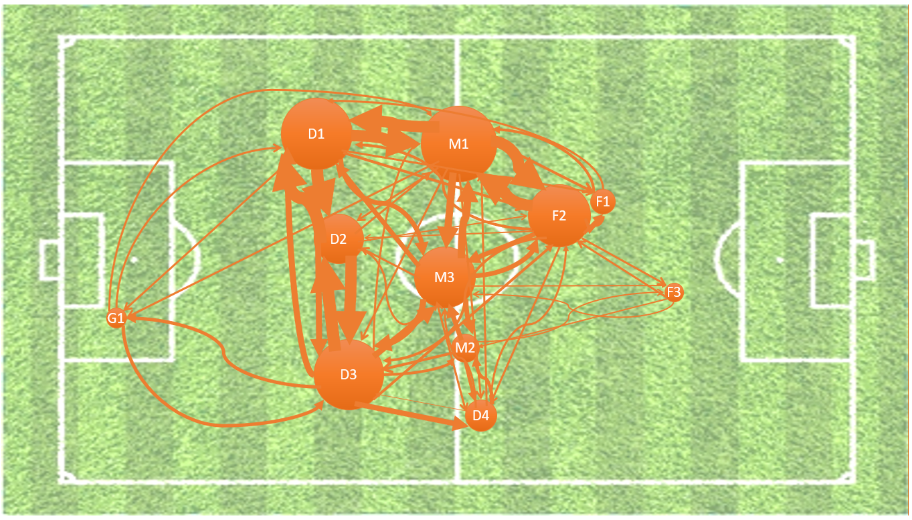


Figure 4: Passing Network of Starting 11 in a Match

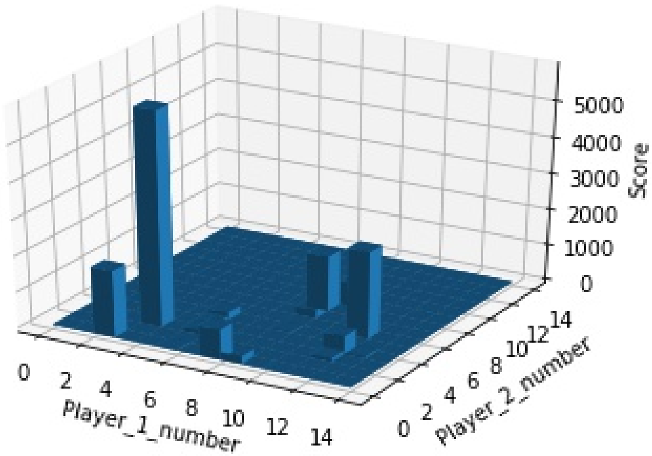


Figure 5: The Connection Score in Dyadic Configuration

0	1	2	3	4	5	6
Huskies_D1	Huskies_F1	Huskies_M1	Huskies_F2	Huskies_M2	Huskies_M3	Huskies_G1
7	8	9	10	11	12	13
Huskies_D2	Huskies_D3	Huskies_D4	Huskies_F3	Huskies_D5	Huskies_M4	Huskies_M5

Figure 6: Neural Network Model

In figure 5, the x axis and y axis represents players' numbers which show the corresponding players' name based on figure 6, and the z axis shows the connection score mentioned in section 4.3. From figure 5 we can recognize the dyadic configurations among players. Based on the fact that $(x, y) = (2, 0)$, $(x, y) = (3, 2)$, $(x, y) = (8, 7)$ and $(x, y) = (11, 5)$ have relatively larger connection score. From that perspective we we can draw the conclusion that player M1 and D1, player F2 and M1, player D3 and D2 as well as player M3 and D5 have built a solid dyadic configuration from the point of a whole match's view.

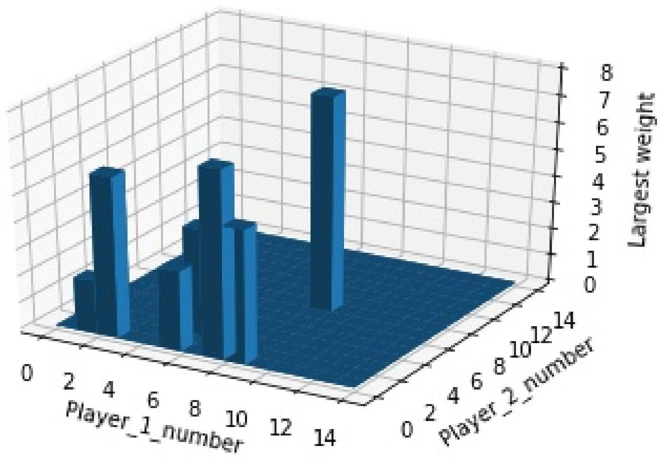


Figure 7: The Maximum Sum of Weight in Dyadic Configuration

Figure 7 shares the same meaning of both x axis and y axis of figure 6. However, figure 7's z axis represents the largest sum of weight in all dual rings of the passing network in the same match as figure 5. It indicates the instantaneous passing intensity among the players so that we can use it to analyse the dyadic configurations in a short period of time. As figure 7 shows in this match, player D2 and D3, D1 and D2, D1 and M1 as well as D1 and D3 have a relatively big instantaneous passing intensity than other dual rings and these dual rings can be regarded as forming a strong dyadic configurations in a short time.

4.2 Structural indicators

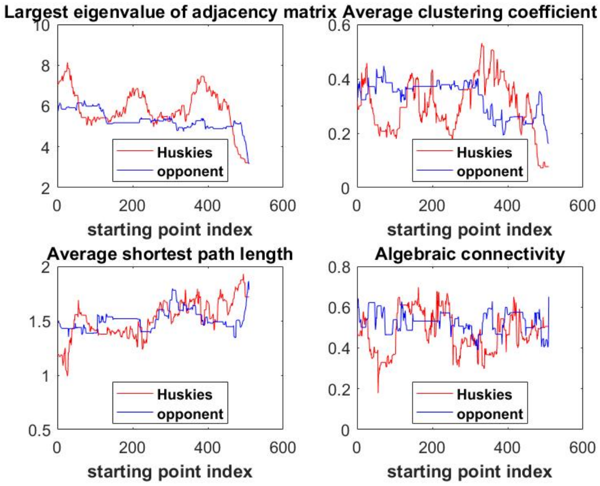


Figure 8: The Trend of 4 Factors During a Game

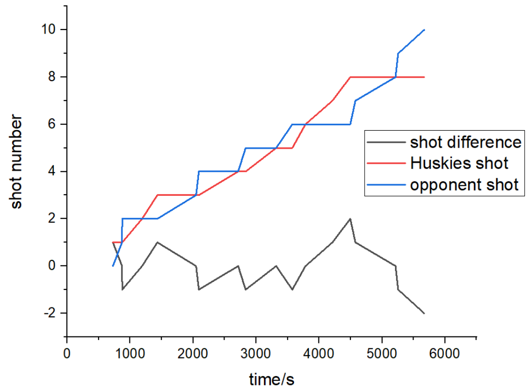


Figure 9: The Shot Trend During a Game

Figure 8 shows both team Huskies and its opponent's variation of these 4 factors towards the index of starting pass among the 50 passes we examined, while figure 9 shows the variation of both teams' shot number towards time. Based on the assumptions, figure 9 show the results of offense, or we can say the performance of players during the match. Comparing shot number's variation tendency with those 4 factors', we can draw several conclusions. Firstly, the higher largest eigenvalue of adjacency matrix showed the better performance players can give. In addition, the greater average clustering coefficient indicates the better performance of a team. Besides, the lower average shortest path length, the team performs better. Moreover, though the algebraic connectivity between 2 teams are quite close, we can also conclude from the beginning of the game at around 1000s that the larger value of algebraic connectivity indicates the better performance.

4.3 Neural Network Evolution Algorithm

In order to make improvements to Huskies team structure, we will implement the combined method of Neural Network and Evolution Algorithm. (see figure 9)

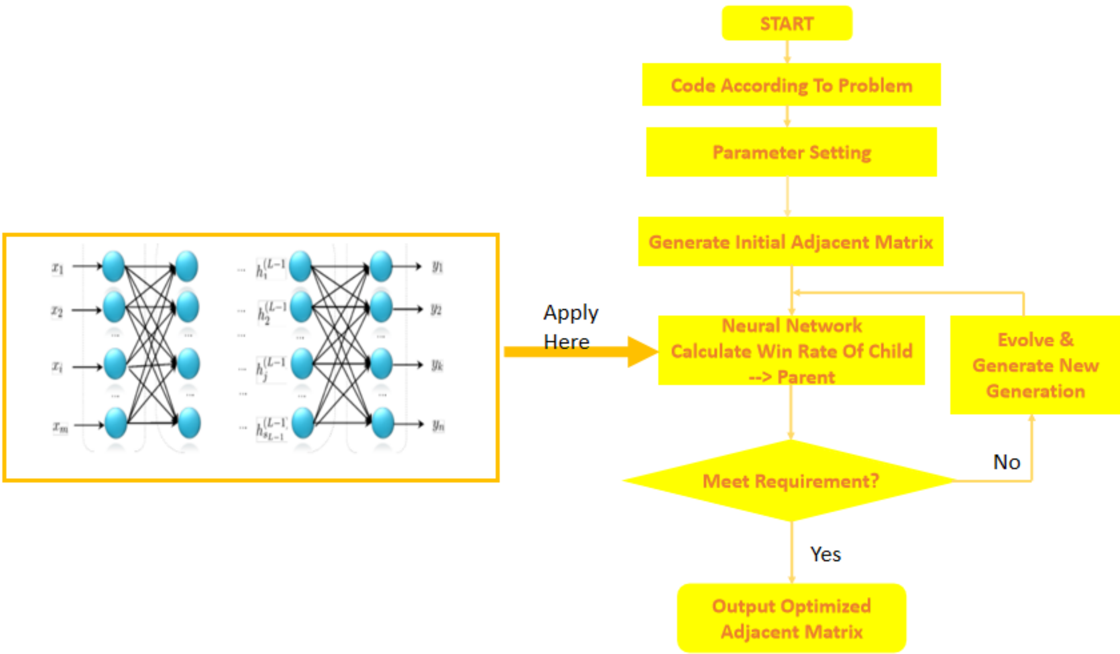


Figure 10: Neural Network + Evolution Algorithm

Here EA is utilized to generate a set of optimal adjacent matrix which can generate the best patterns of passing network and structural indicators. Firstly, EA would generate a set of adjacent matrix for Huskies randomly. Then through variation and combination, EA generate new sets of adjacent matrix. During the process of selection, the well trained neural network is evoked and works as judge of fitness. The neural network will help select better adjacent matrix according to structural indicators for the next generation. As a result, the best set of adjacent matrix could be acquired after hundreds of iteration.

4.3.1 Implementation

- **Training of neural network**
  - **dataset**

Here, dataset is mainly obtained from matches.csv and passingevents.csv. We calculate the adjacent matrix for each team and save them as 30*30 matrix. Now, we can obtain the 4 structural indicators for both teams as features of their adjacent matrices.

adjacent matrix	x	y
38*30*30	38*8	38 * 1

In data set, x represents 8 indicators for 38 matches,and y presents the result of the match (win / loss)

- **training** The network is constructed using Pytorch.nn, and the network's structure is as follow.

$n_features$	n_hidden	n_output
8	10	2

- **training result**

Train acc	valid acc
0.93	0.89

- **Evolution Algorithm**

Population Size	200
Max Generation	300

Population size represents the number of adjacent matrix in one generation. Max generation represents the maximum times of iteration.

4.3.2 Results

Using the Neural network + evolution algorithm model, we will get 200 best individuals as the best generation of Huskies team . Then we let these individuals compete with 38 opponent teams and get each pairs' adjacent matrix, thus we can feed the matrix into neural network and predict win rate of each individual.

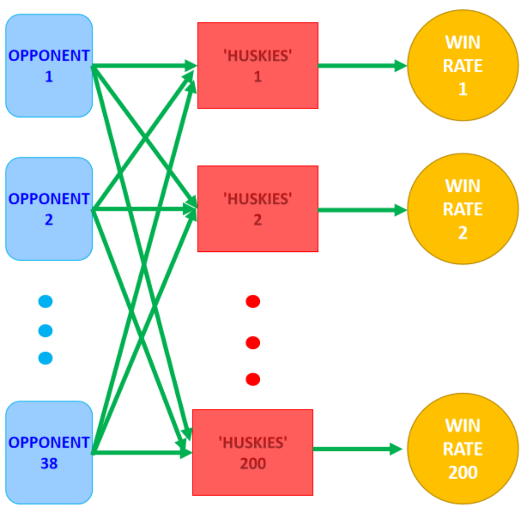
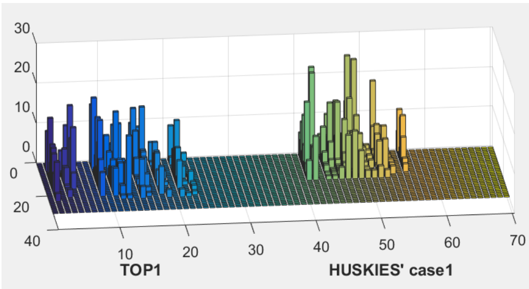
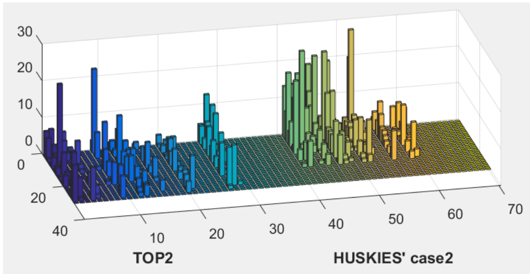


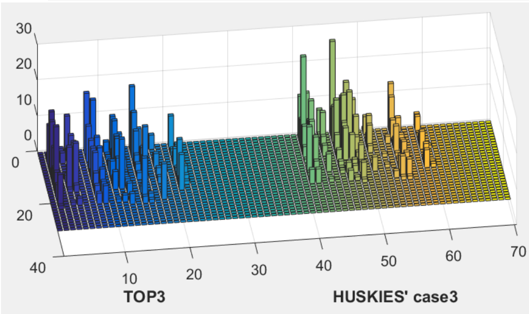
Figure 11: Find out the best structure for Huskies

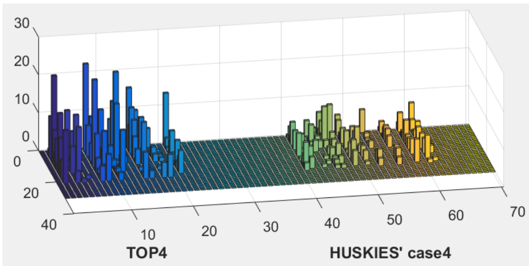
Now, we can use the model to find out the best network structure of Huskies to beat the most opponents. In this way, we will be able to observe the most successful team structure and learn to make improvements for Huskies.

Here, we will show 4 of the best adjacent matrix structure and the comparison case of the Huskies team. Since these best matrix structure widely get the highest winning rate among all the matches, the team structures are **universally effective**.









The top - 4 adjacent matrix structures for Huskies team here are colored in blue, and the comparative matrix structure used in Huskies' matches are painted yellow, which are considered relatively poor structures.Here x axis and y axis represents the number of the team member, thus the adjacent matrix is 30*30. The win rate of the top 4 predicted by neural network is: [0.92, 0.89, 0.86, 0.83], which means they defeat most of the opponents.

Now, we can make analysis of various factors' influence on team success.

By comparing indicators and appearance of the best and poor matrices, we can draw conclusions that

- A single player's passing number does not contribute much to the team's success. However, the total passing number of the whole team does contributes to teams' success.

  In figure 1,2,3, all three cases of Huskies' poor matrices indicates several large passing numbers, but the team's performance is much lower than those without outstanding personal performance.

  However, a team's total passing number is important. In case 4, the poor total passing number leads to the structure's poor performance.

- Variance of a team's passing number's is negative correlated to a team's success. Poor matrices in figure 2,3,4 indicate huge variances among team members, while the best three matrices only have small variances. That means a good team needs everyone to contribute similar effort, but not only some of them contributes the most effort.

- The Connectivity between members is also vital. Figure 1,2,3,4 all indicates that members in the top - 4 matrices are highly connected, which means the passing choices for members are various. This would help them to build up a stronger passing network and therefore contribute to team's success.

5 Strengths and weaknesses

5.1 Strengths

- **Applies widely**
  If more other structural indicators are applied, our model can readily adapted to analyse other problems, even solving multiple problems simultaneously.
- **various structural indicators**
  In our study, various structural indicators are implemented to build up the modal, which means various aspects of the data are taken into consideration.
- **Flexibility**
  With neural network & evolution algorithm and independent calculation of structural indicators, our modal can utilize various kinds of adjacent matrix.

5.2 Weakness

- **Lack of accuracy**
  Neural network lacks data to be better trained. The accuracy of the neural network can be even better improved if given hundred times of data.
- **Strategy Utility**
  Evolution Algorithm mainly focus on optimizing adjacent matrix. However it's problematic for algorithm to create adjacent matrix or simulate real behavior of passing network.
- **Abstraction of structural indicators**
  The indicators we use can not link to win rate directly, which means win rate is only analysed in macroscopic perspective. And our team didn't find the connection between win rate and a single indicator.

6 Further Discussion

Teamwork is now being more and more important and common. Although each team's task could be totally different, there are still a lot features of teamwork in common.

- **Connectivity**

  The connection between team members exists in all teamwork.For football match, it exists as passing event; for group project, it could be qualified as the number of email between group members. According to our model, connection between members does great contribute to team success,and therefore an efficient team should consist of members who are more connected to each other.

- **Variance**

  Members in team does different amount of work. However, controlling the variance of the efforts could be important. Our model's result tells that even though some member contributes a large amount of effort, the whole team's variance is so large that the team is totally imbalanced. Thus, to make a team more efficient, the team should balance each members' work load and reduce the variance of teamwork.

- **Centralization** Last but not least, centralization also contributes to a team's success. Our model points out that when a team tend to centralize, its maximum eigenvalue would increase and the performance of team would also be better. That means a team needs a powerful leader. If the leader could be the center of the team to make arrangements to other group members, the efficiency of team would improve a lot.

References

[1] Buldú, J.M., Busquets, J., Echegoyen, I. et al. (2019). Defining a historic football team: Using Network Science to analyze Guardiola's F.C. Barcelona. Sci Rep, 9, 13602.

[2] Rainer Storn, Kenneph Price, 1997. Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces.Journal of Global Optimization 11: 341–359, 1997. 341.

[3] Cintia, P., Giannotti, F., Pappalardo, L., Pedreschi, D., Malvaldi, M. (2015). The harsh rule of the goals: Data-driven performance indicators for football teams. 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA), 1-10, 7344823.

[4] Duch J., Waitzman J.S., Amaral L.A.N. (2010). Quantifying the performance of individual players in a team activity. PLoS ONE, 5: e10937.

[5] GÜRSAKAL, N., YILMAZ, F., ÇOBANOĞLU, H., ÇAĞLIYOR, S. (2018). Network Motifs in Football. Turkish Journal of Sport and Exercise, 20 (3), 263-272.

Appendices