Problem Chosen	2020 MCM/ICM Summary Sheet	Team Control Number
D		2013711

Forza Huskies —— You Will Never Walk Alone

As society evolves, team cooperation has become a major format of working nowadays. Understanding how teammates of various skills cooperate with one other is of crucial importance. This will lead to the success of a team in diverse and even complex situations. In this paper, we build model to identify and evaluate the cooperative performance of the Huskies players, and discuss how they lead to a winning in the game.

First, we construct the Passing Network Model based on knowledge of Social Network. The model transfers the passing data to a directed weighted complex network which can reveal specific configurations and formations of the team. The graph of the network can visually reflect certain **Dyadic** and **Triadic Configurations** through link patterns, and **formations** can be revealed by the position of the nodes. Various network metrics including $Centrality$, $(Average)Clustering$, $AverageShortestPathLength$, $CentroidCoordinate$,and $MovementPositionRatio$ are used to confirm the patterns discovered as well as digging out strategies of the team.

We use passing data for the Huskies in different scales to construct different networks. In the **Season Passing Network**, we find Dyadic Configurations from D1 to D3, between M1 and F2, M1 and M3; Triadic Configuration: M1 to M3 to F2, and D4 to M1 to F2. This reflects the classic defender-midfielder-forward attacking strategy. The high centrality for M1, D1 and D3 marks their role as key midfielder and defenders respectively. In the **Single Game Passing Network**, we analysis the first game of the Huskies in the season. We find Dyadic Configurations from D1 to M1, between M1 and F2; Triadic Configuration from D1 to D2 to D3. The formation used by the Huskies is 4-3-3, and the Strategy applied is Defensive Counterattack. Defense triangle is formed among D1,D2,D3, and the attacking line is formed with the attacking midfielder F2 being the fulcrum. In the **Minute-to-Minute Passing Network**, we construct a series of networks, in which we find that the network metrics fluctuate greatly for the Huskies in Game 1 as the game proceeds. The Huskies is in defense status overall, while gradually increase attacking as time goes on.

Second, we focus on both team and player level performance to evaluate successful teamwork. In **team-level performance**, we identify the influence of the home away effect, familiarity with different formations, ball movement tracks, and other positional statistics to respectively quantify the team's **adaptability, flexibility, tempo, and attack positions**. Using these indicators, we analyze team performance of the Huskies, their opponents and real clubs from the Premier League. We find that the Huskies have a relative low adaptability in goal types and a low flexibility. Meanwhile, position analysis reveals some of the Huskies' specific attacking patterns such as attacking down the right side and attempting crosses. In **player level performance**, we use AHP to analyze different players' contributions to the team. We group the players by their position, value the importance of passes, shots, and duels differently based on the characteristic of the position, and build a rating system to estimate their contribution to the whole team. Moreover, we compare their ratings with players from other teams in the league to better understand the ability of the Huskies' players in terms of cooperation.

Third, based on the passing network and teamwork performance, we conclude the strengths and weaknesses of the Huskies. We also identify some successful strategies, such as playing offside traps, attempting high pass and long ball. Combining these findings and analysis of the Huskies' players distribution of contribution, we recommend the Huskies coach with two first elevens: 4-1-4-1 and 4-3-3, which we believe could be effective for them in the future.

Finally, we analyze that our Network Model and Team Performance Model can be generalized to evaluate the cooperation among team members in backgrounds other than soccer game.

Keywords: Social Network; Team Level Process; AHP

Contents

1 Introduction

1.1 Background

Teamwork has become a major format of working in modern society. Understanding how teammates of various skills cooperate with one other is of crucial importance, leading to the success of the team in diverse and even complex situations. Given the detailed available data, competitive sports, football in particular, has provided a good setting for teamwork analysis.

1.2 Problem Restatement

In this paper, we are going to analysis the dynamics of the soccer team Huskies based on the data of one season, examining possible cooperation among players that influences the results of the game. It is important to identify the dynamics and strategies used by Huskies, concluding advantages and disadvantages, thus giving suggestions to the coach for how to improve teamwork in next season.

We will proceed our work as follows:

- Construct ball passing network, in which the nodes represent players, and the weighted links represent frequency of ball passing between players.
- Find specific network patterns to identify team formation in various scales, including interactions from two players to the whole team, and from short-time to the whole season.
- Construct indicators reflecting successful team level processes and player performances. Develop a teamwork model using above indicators to evaluate Huskies' teamwork in terms of structure, configuration, and dynamics.
- Find advantages in Huskies' teamwork based on the above model, while pointing out disadvantages that need to be improved in the next season based on the network analysis. Offer suggestions to the Huskies players and coaches.
- Generalize above model and analysis to general teamwork in the society. Give suggestions on designing more effective teams, and list out factors that need to be considered when constructing the generalized models.

2 Preparations of Model

2.1 General Assumptions

- Each player has played with full potential, regardless of injuries, fatigue and transfers or loans.
- Coaches have no influence on teamwork among players.
- Only successful passes are counted, passes that have been stole or intercepted are not counted.
- Players are only grouped into defender, midfielder, and forward when evaluated, further division of their roles, such as DM and AM are not considered.

2.2 Notations

The primary notations used in the paper is listed below:

Table 1: Notations Description

Notations	Description
n	number of nodes in the Passing Network
V	the set of nodes in the Passing Network
$d_{i,j}$	the length of the shortest path from $node\ i$ to $node\ j$
DC, CC, BC, EC	Degree Centrality, Closeness Centrality Betweeness Centrality, Eigenvctor Centrality
AI	Adaptability indicator
R	formation rating
FI	flexibility indicator
PCM	pairwise comparison matrix describing respective importance among indicators
$WeightVector$	vector having 3 elements to describe indicators' weight respectively

3 The Passing Network Model

Ball passing movement is the core of soccer, through which players establish connections among one another. It is the key to measuring cooperative performance and group dynamic of a team. Social network has provided a useful tool to examine the ball passing movement. It can help us visualize the ball passing patterns between players, and quantify the extent of coordination through various network metrics. Social network uses graph knowledge and algebra models to capture link patterns in the network, evaluates the interactions between members, and measures the influence of individuals to the whole system (1).

We use the data in the "passingevents.csv" to construct the social network for Huskies' ball passing events, which we name **Passing Network**. In the network, nodes represent players, and links represent ball passes among players. First, we construct a weighted adjacency matrix A, where a_{ij} equals the sum of passes from player i to player j within a certain period of time. Then we transfer this matrix to a directed weighted graph by using the social network analysis tool *Gephi*. Weights of link $w_{ij} = a_{ij}$. Thickness of the link in the network is proportional to the link weight. Size of the node is proportional to the sum of in-degree and out-degree of the node. Position of the node is the average coordinate of all passes initiated by corresponding player. Color of the node reveals the role of the corresponding player, which means same color representing same role. Color of $link\ ij$ is the same as the color of $node\ i$.

3.1 Definition of Network Metrics

There are several metrics that are generally used when analyzing social network such as centrality and clustering. When calculating relevant metrics, distance of shortest path between nodes are calculated. We want to achieve the result that more frequent passes from palyer$i$ to palyer$j$(bigger a_{ij}), corresponds to a shorter distance $dis_{i,j}$

$$dis_{i,j} = \begin{cases} \frac{1}{a_{ij}}, & \text{if } a_{ij} \neq 0 \\ \infty, & \text{if } a_{ij} = 0 \end{cases}$$

We use this distance as weights to find the shortest path between two nodes in the passing network. We use $d_{i,j}$ to denote the distance of the shortest path between node$i$ and node j under this definition.

Based on knowledge in graph and the new definition of distance,we use the following metrics to analyze the network. The calculation method is based on the refrence document of Python package *Networkx*. (2)

- **Degree Centrality (DC)**
  Here we use the out-degree centrality. $outDeg_v$ counts the sum of link weights that goes out from a certain node. We normalize it by dividing $n-1$, which is the maximum possible degree of a node in the network:
  $$DC_v = \frac{outDegree_(v)}{n-1}$$
  The basic idea of out-degree centrality is that, if a node is directly connected to many other nodes, then the ability for it to affect and control the network is relatively high. Thus, a player with high DC passes the ball to more teammates, indicating more direct interactions with others in general.

- **Closeness Centrality (CC)**
  Closeness centrality of $node\ v$ measures the average shortest path distance from all other nodes in the network to v. We normalize it by dividing $n-1$, and then take the inverse of it to achieve the result of higher CC corresponding to shorter distance.
  $$CC_v = \frac{n-1}{\sum_{u \neq v} d(u,v)}$$
  A node with high closeness centrality is relatively close to other nodes in the network.In the case of a directed network as here, only the incoming links to $node\ v$ contributes to the increasing of CC_v, but not outcoming links. Thus, a player who passes ball frequently but doesn't receive ball from teammates frequently will not possess a high CC. As a result, a player with high CC are the ideal receiver of passes considered by teammates, representing the ability to control other players in the team. (1).

- **Betweenness Centrality (BC)**
  Betweenness centrality of $node\ v$ is the sum of the fraction of all-pairs shortest paths that pass through v:
  $$BC_v = \sum_{s,t \neq v} \frac{n(s,t|v)}{n(s,t)}$$
  where $n(s,t)$ is the number of shortest path from node s to node t, and $n(s,t|v)$ is the number of above paths which node v lies in.
  A node with high BC lies on communication paths frequently, thus have the ability to control the network flow. A player with high BC acts as intermediary among other players, controlling the cooperation among other players.

- **Eigenvector Centrality (EC)**
  Eigenvector centrality of a node considers the importance of its neighbors reached by incoming links. If we write EC of all nodes in a vector form, it is the same as the unique eigenvector corresponding to the largest eigenvalue of the weighted adjacency matrix A.
  If equal connection is made to a node, link from important nodes contributes to a higher EC than from unimportant nodes. Thus, a player with high EC may be explained by the high frequency of either receive the ball from many other teammates, or receive the ball from important teammates.

- **Clustering**
  Clustering of a node measures the likelihood of generating directed triangles through it. A directed triangle through $node\ v$ can be formulated if (1) there exists a $link\ vw$, (2) there exists a $nodeu$ which both $link\ vu$ and $link\ uw$ exists. Taking the link weights into account, we use the geometric average of the link weights in the triangle described above to measure the number of directed triangles. Here, we first normalize the link weights by network's maximum link weight: $\hat{w}_{vu} = w_{vu}/\max(w)$; and then normalize the expression taking the total degree of v (deg_v, sum of in-degree and out-degree) into account.

$$Clustering_v = \frac{1}{deg_v(deg_v - 1))} \sum_{u,w \in V} (\hat{w}_{vu}\hat{w}_{uw}\hat{w}_{vw})^{1/3}$$

  A player with high clustering is more likely to create triangles, which means that the percentage of him using **Triadic Configurations** to his total passes is relatively high.

3.2 Season Passing Network

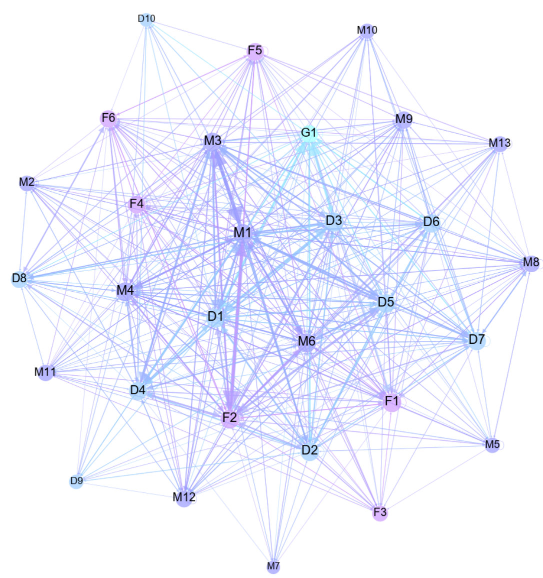


Figure 1: Season

Figure 1 shows the passing network for all Huskies passing events in the season. To show the network more clearly, the position of the nodes is not the average coordinate as described above. Instead, we put the nodes with bigger size (higher degree) in the middle. The figure helps us visually identify the ball passing patterns among players. First, M1, D6, D1, D3, D5 are placed in the middle, meaning that they receives and passes the ball more frequently, generally interacting with other teammates. Second, through the thickness of the links, we can also observe obvious **Dyadic Configurations**: ball passing between M1 to F2 (both directions), M1 and M3 (both directions), and from D1 to D3; and **Triadic Configuration**: M1 to M3 to F2, and D4 to M1 to F2. These triadic configurations from midfielder or defender, to midfielder, to forward, reflect the specific passing patterns of the Huskies.

We also calculate the network metrics, the results are shown in the appendix. Here we only list the top 3 players of each metric:

Table 2: Top 3 players of Different Indicators

Ranking	DC	CC	BC	EC	$Clustering$
1	M1	M1	M1	M1	F2
2	D1	F2	F2	F2	M1
3	D3	D1	D1	M4	M3

M1, who possesses the highest value of centrality of all form, is definitely the leader of the network, cooperating with various teammates and influencing the interactions among other players. Midfielder stands in the center position of the field, thus can easily interacts with both defense and forward, resulting to a high DC and BC. It is also the connector of defender and forward, playing the role of intermediary in the network flow, leading to a high CC. However, M1 is the only midfielder in the top ranking of DC, CC, and BC. This means that other midfielders are not playing as well as they should be, so the workload for the skillful midfielder M1 becomes extremely large.

D1 and D3 own a high degree of DC, being the key defenders in the team. The high DC for defenders also reflects the high frequency of ball passes starting from defenders, indicating that the Huskies are constantly under the defensive status, showing that the Huskies are not competitive in the league.

M1, F2, D1 rank for top3 in CC and BC, each being the key player of three positions. High CC and BC reflects their high ability to control other teammates and interactions in the network.

M1, F2 and M4 own a high EC. Since F2 and M4 don't have a high in-degree, then the high EC for them can be attributed to the high importance of their neighbors in the network.

F2 and M3 are two nodes that own a high $Clustering$ rather than M1. This is because of the **Dyadic Configurations** of ball passing from F2 to M1, and M3 to M1, and most other players interact with M1. In this way, triangles are more likely to be made through F2 and M1, or M3 and M1. This corresponds to the **Triadic Configuration** that is visually obtained in **Figure 1**.

3.3 Single Game Passing Network

Using the ball passing data in a particular game, we can construct the Single Game Passing Network. Here we show the passing network for Game 1 in Figure 2, passing network for some of the other games can be found in Figure B.

The configurations are clearly revealed in **Figure 2**:

- **Dyadic:** from D1 to M1, between M1 and F2.
- **Triadic:** from D1 to D2 to D3.

Comparing the configurations found in the season network, certain changes can be noticed. First, the configurations related to M3 is not clearly revealed in Game1. Second, the **Triadic configuration** is found among defenders, instead the offense pattern of defender-midfielder-forward.

From the position of the nodes in **Figure 2**, we make following observations. First, we can see that the **formation** of this game is 4-3-3. D1,D2,D3 forms the defender triangle; M1,M3,M2 is in charge of passes in up, middle, and down(y coordinate) side of the field respectively; F2 plays the role of attacking midfielder, being the connectors between midfielder and forward. Second, we can observe

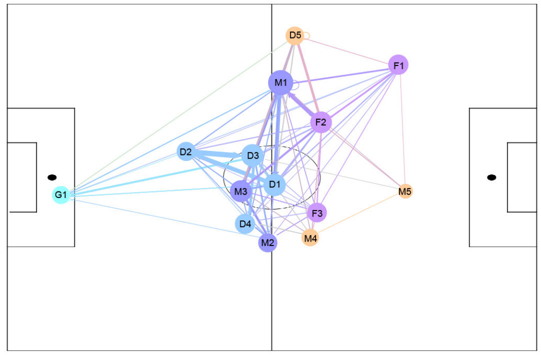


Figure 2: Passing Network for Game 1

that more passes are initiated in the left side of the field(x coordinate), indicating that the Huskies are constantly under the status of defense.

Table 3 shows relevant network metrics for Game1. The order of the player is based on the ranking of DC from high to low. We make analysis as follows:

- M1 still ranks the first for all forms of centrality, being the key player to connect other teammates and control cooperation.
- F2 ranks the second for all forms of centrality expect for DC, showing that it plays an important role in the network flow. It is worth mentioning that, in particular, the high BC reveals the role of intermediary between midfielder and forward, which also indicates his position as attacking midfielder.
- M3 possesses a high CC, EC, and $clustering$. This means that M3 still has a relative importance in the overall cooperation among teammates, forming triangles with various teammates, though no obvious configuration for M3 is observed in **Figure 2**.
- The high value of clustering for D1, D2, and D3 has confirmed the triadic configuration that we observed in **Figure 2**

Combining the discussions above, we can infer that the Huskies use the strategy of **Defensive Counterattack** in Game1:

- Defense is emphasised greatly, while attacking is relatively weak.
- Defensive line is formed by the defender triangle: D1,D2,D3.
- M1 is the most important midfielder, being the main conductor of the whole team.
- F2 is the fulcrum in the attacking line, receiving balls from midfielder and passing it to other forwards, thus the counterattack is formed.

We notice that there's substitution for the Huskies: D2 to D5 (t=46 minutes), M2 to M4 (t=60 minutes), F3 to M5 (t=77 minutes). D5, M4, M5 are marked by yellow color in **Figure 2**. First, from the coordinate of these three players, it is clear that their location are in the right half of the field, and we can see the trend: the later a player join the game, the closer he stands to our own goal. This may indicate that as time goes on, the Huskies are increasingly focused on the offense. Second, as shown

Table 3: Network Metrics for Game 1

Player	DC	CC	BC	EC	$Clustering$
M1	1.000	0.867	0.355	0.379	0.109
D1	0.923	0.722	0.109	0.271	0.134
D3	0.923	0.722	0.233	0.250	0.120
F2	0.769	0.866	0.319	0.346	0.111
D4	0.692	0.684	0.013	0.235	0.103
M3	0.692	0.813	0.160	0.318	0.141
M2	0.615	0.619	0.0	0.209	0.104
D2	0.538	0.650	0.0	0.234	0.157
G1	0.462	0.591	0.0	0.182	0.107
D5	0.462	0.650	0.0	0.246	0.147
F1	0.462	0.813	0.0	0.316	0.076
F3	0.462	0.765	0.0	0.301	0.077
M4	0.385	0.650	0.0	0.195	0.088
M5	0.154	0.591	0.0	0.154	0.076

in **Table 3**, those three players have a very low centrality and $Clustering$. This may be explained by the less time they participating in the game, thus contribute less passes and play a relative unimportant role in the Passing Network of whole game.

In order to study how the team performance has changed within one game, and to solve the problem of time difference when substitution occurs, we need to construct a Minute-to-Minute Passing Network.

3.4 Minute-to-Minute Passing Network

In order to study the temporal evolution of the team behavior within one game, Buldú et al.(2019) constructs $l - pass\ networks$ of a team. In this model, a series of ball passing networks are generated, with l sequential passes possessed by each network. Thus, the networks can be treated as dynamic entities which changes as time goes on. In Buldú et al.(2019), $l = 50$ is believed to be high enough to form each network, and low enough to record the changes of network in the scale of minutes. The metrics mentioned in Section 3.1 are designed to evaluate the connectivity of each individual nodes in a certain network. To record the evolution of the network, we use the following global metrics to describe the character of certain network (3):

- **Average Shortest Path Length($ASPL$)**
  Average shortest path length is the average of the distance of all-pairs shortest paths.

  $$ASPL = \sum_{s,t \in V} \frac{d_{s,t}}{n(n-1)}$$

  A low $ASPL$ indicates better connectivity of the network. We should notice that given equal total number of passes in a network, if the passes are more evenly distributed among players, then $ASPL$ should be lower, since we define the distance as the inverse of number of passes.

- **Average Clustering(AC)**
  Average clustering is the average of clustering of all nodes in the network. It measures the average likelihood of forming directed triangles among three players. It can be seen as an indicator to evaluate the intensity of **Triadic Configuration**. Higher AC means more interactions are made among three players, thus the **Triadic Configuration** is more robust.

- **<x> and <y> Centroid Coordinate**

$$\text{<x> Centroid Coordinate} = \frac{\sum_{all\ passings} EventOrigin_x}{number\ of\ all\ passings}$$
$$\text{<y> Centroid Coordinate} = \frac{\sum_{all\ passings} EventOrigin_y}{number\ of\ all\ passings}$$

Centoid coordinate features the average coordinate of all passing events, which can reflect the position and status of the team.

- **Movement Position Ratio** $\frac{\Delta y}{\Delta x}(MPR)$

$$\frac{\Delta y}{\Delta x} = \frac{\frac{\sum_{allpassings} |EventDestination_y - EventOrigin_y|}{\sum_{allpassings} |EventDestination_y - EventOrigin_y|}}{number\ of\ allpassings}$$

Movement Position Ratio reveals the ratio of vertical change and horizontal change, indicating the overall direction of passes. A low MPR means that the passing counted in the network traverse the field more horizontally than vertically.

Based on this model, we construct the Minute-to-Minute Passing Network Model for Game 1 in the following steps:

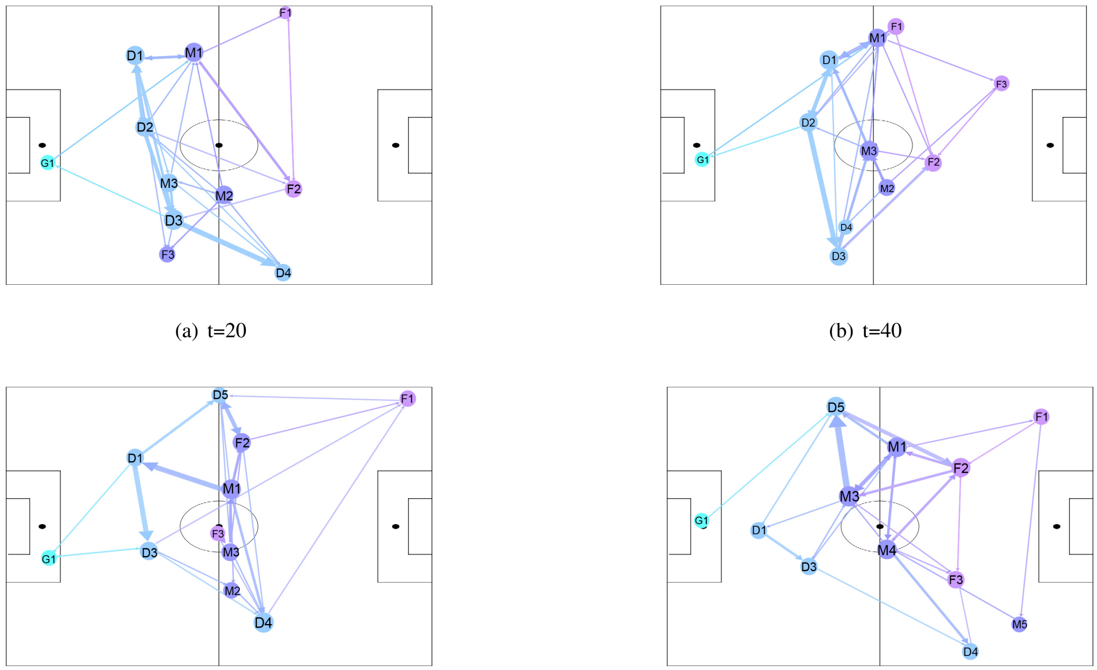
- Use the first 50 passing data of Huskies since the beginning of Game 1 to generate the first passing network, mark the time of the time of the $50-th$ passing, and calculate relevant parameters.
- Use the $2-nd$ to $51-th$ passing data to generate the second passing network, mark the time of the time of the $51-th$ passing, and calculate relevant parameters.
- Repeat the above procedure, each time sequentially discard the oldest passing and add a new passing, until there is no more passing left.
- Based on the series of networks, time, and relevant parameters, evaluate the evolution of the team behavior.

Figure 3 shows the $50-passing$ network at $t=20$, $t=40$, $t=60$, $t=80$. Relevant properties of those 4 networks are shown in **Figure 3**. To save space, we only demonstrate the $ASPL$, AC, EC here.

As we can see from **Figure 4(a)**, $ASPL$ and AC are different at these 4 timestamps. $ASPL$ is lowest at t=40, which corresponds to the fact that passes at t=40 are more evenly distributed. This can be observed in **Figure 3**, where less obvious **Dyadic Configurations** can be found at t=40 comparing with t=20, 60, 80. AC is highest at t=60 while lowest at t=80, reflecting **Triadic Configuration** is strong at t=60 while weak at t=80. From **Figure 4(b)**, we can examine how each player's centrality changes at different timestamp. Since the network reflects the passes in a minute scale, the problem of time difference incurred by substitution no longer exists. Analysis for centrality can be done as in Season and Game Passing Network, which is omitted here due to space limitation.

In order to study the position and overall strategy changes along the game, we reveals how the Centroid Coordinate and MPR evolves in **Figure 5**. For comparison, we show the result of the opponent in the figure as well.

As we can see from **Figure 5**, parameters for the Huskies fluctuates greatly as time goes on, while Opponent1 remains relatively stable. Huskies <x> Centroid Coordinate is increasing overall, from

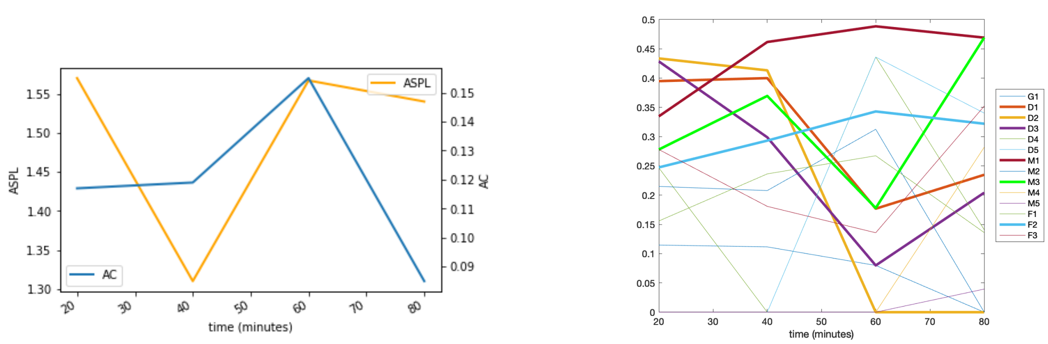


(a) t=20 (b) t=40

(c) t=60 (d) t=80

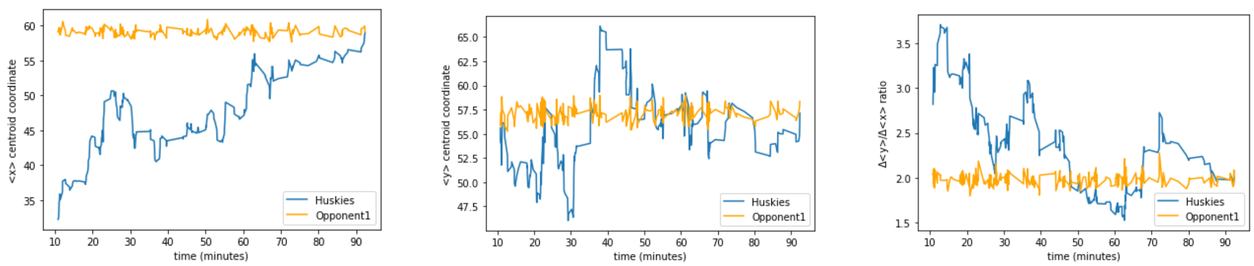
Figure 3: Minute-to-Minute Passing Network for Game 1

less than 20 to 60 in the end, but Opponent1 stays at 60 generally. This reflects that Opponent1 is more offensive than the Huskies. The Huskies mostly stay in the status of defense along the game, while attacks increase after t=60 minutes. This corresponds to the Defense Attacking strategy analyzed before. Both the Huskies and Opponent1's <y> Centroid Coordinate is above 50, meaning that they all favor left side passes. MPR for the Huskies is higher than Opponent1 in general, which reflects the vertical passing pattern of the Huskies.



(a) properties (b) Eigenvector Centrality for each player

Figure 4: Network Metrics for $l - pass$ Network at t=20, 40,60,80



(a) <x> Centroid Coordinate (b) <y> Centroid Coordinate (c) MPR

Figure 5: Centroid Coordinate and MPR in Game1

4 Teamwork Performance Model

To better understand what affect successful teamwork, we construct two models that respectively measure the performance of the whole team and individual players, using different indicators to reflect them. By analysing playing patterns among players and teams, we get the ideas of how different factors may contribute to successful teamwork.

4.1 Team Level Processes

4.1.1 Adaptability

The adaptability of a team refers to the amount of volatility it exhibits when facing different playing environments. Generally, a team play very well in familiar environment. However, when faced with poor training ground, extreme game temperature and geographic environment, away fans booing, the poorly adaptable team may be easily beaten. Therefore, through the difference between the team's performance at home and away, we can clarify its adaptability.

We firstly pick three types of data that we believe are most likely to be influenced by home away effect, namely: Goals, Passes, and Fouls. Due to the lack of information about actual goals, we assume that the number of penalties and free kicks are proportional to that type of goals.

We collected data in these types of both the Huskies and their Opponents' performance (against the Huskies) in the whole season. We define four vectors: $Husk_{home}$, $Husk_{away}$, $Oppo_{home}$, $Oppo_{away}$ to divide data into four groups:

$$Husk_{home} = (g_{11}, g_{12}, ..., g_{1i}, p_{11}, p_{12}, ..., p_{1j}, f_{11}, f_{12}, ...f_{1k})$$

$$Husk_{away} = (g_{21}, g_{22}, ..., g_{2i}, p_{21}, p_{22}, ..., p_{2j}, f_{21}, f_{22}, ...f_{2k})$$

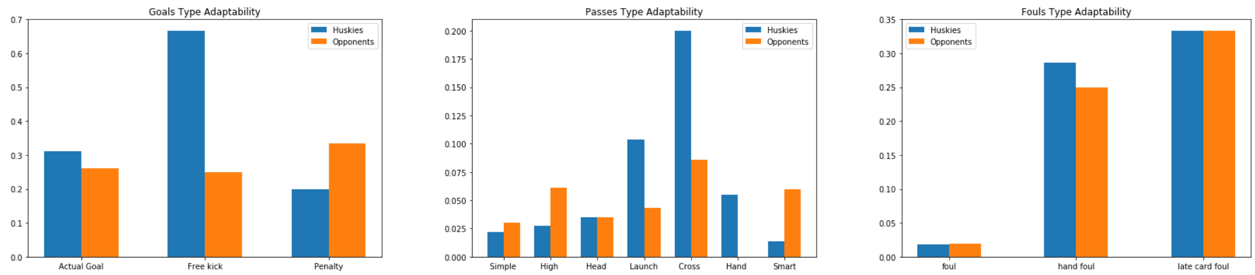
$$Oppo_{home} = (g_{31}, g_{32}, ..., g_{3i}, p_{31}, p_{32}, ..., p_{3j}, f_{31}, f_{32}, ...f_{3k})$$

$$Oppo_{away} = (g_{41}, g_{42}, ..., g_{4i}, p_{41}, p_{42}, ..., p_{4j}, f_{41}, f_{42}, ...f_{4k})$$

where g_{ij},p_{ij},f_{ij} represent the number of j^{th} type of goal, pass, and foul in the i^{th} vector.

Then, we define $Husk_{adapt}$ and $Oppo_{adapt}$ to describe how much these data types could be influenced by the home away effect. And we compare them separately in **Figure 6**.

$$Husk_{adapt} = \frac{\left|Husk_{home} - Husk_{away}\right|}{Husk_{home} + Husk_{away}}$$



(a) goals type adaptability (b) pass type adaptability (c) foul type adaptability

Figure 6: Adaptability of Different Data Type

$$Oppo_{adapt} = \frac{|Oppo_{home} - Oppo_{away}|}{Oppo_{home} + Oppo_{away}}$$

At last, we use a weight matrix to get the adaptability indicator AI of both Huskies and their opponents average level, AI is defined below:

$$\begin{bmatrix} Husk_{adapt} \\ Oppo_{adapt} \end{bmatrix} \begin{bmatrix} vec_{goal} & 0 & 0 \\ 0 & vec_{pass} & 0 \\ 0 & 0 & vec_{foul} \end{bmatrix} = \begin{bmatrix} GA_{husk} & PA_{husk} & FA_{husk} \\ GA_{oppo} & PA_{oppo} & FA_{oppo} \end{bmatrix} = \begin{bmatrix} AI_{husk} \\ AI_{oppo} \end{bmatrix}$$

where vec_{goal},vec_{pass},vec_{foul} and GA, PA, FA respectively represent the three weighted vectors and indicators of Goal, Pass, and Foul. To make it clearer, we set the AI of the opponents to be $[1, 1, 1]$ and we use the same method to process the data of Team $Liverpool$ in 2019/2020 Premier League (4).

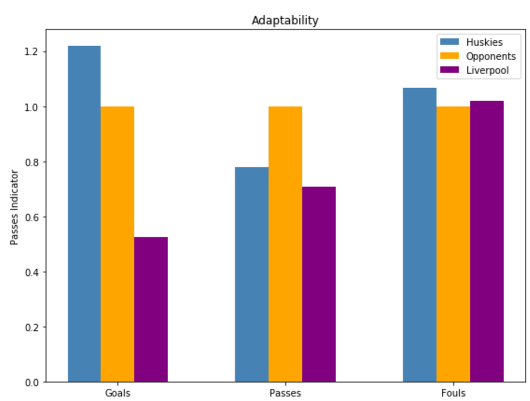


Figure 7: Adaptability Indicator

Note that AI measures the difference of performance between home and away, the higher AI is, the less adaptable the team will be. Compared with opponents, it is clear that the Huskies have a relatively higher GI and lower PI, which means their goals are much more likely to be affected by home away effect and that their passing patterns are more adaptable than their opponents. Liverpool, on the other side, possess strong Adaptability that its performance is hardly influenced by the variety of game environment.

4.1.2 Flexibility

The flexibility of a team refers to the ability of performing with different formations.Specific strategy change are made when facing various opponents. Usually, these counter-strategies can be fully reflected in the change of formations. For instance, a team may replace a defender with a midfielder in order to strengthen their control of the midfield and increase possession. The more flexible a team is, the more familiar they are with various formations.

In order to get a universal standard of formations, we choose Ppg and $GDpg$ as variables that decide whether a formation is successful. Then, we use 0-1 normalization method to get normalized value of $GPpg$ and divide Ppg by three to make it normalized as well.

$$NPpg = \frac{Pg}{3}$$

$$NGDpg = \frac{GDpg - GDpg_{min}}{GDpg_{max} - GDpg_{min}}$$

We give these two normalized variables constant weights and add them up to get a rating R, which is defined below:

$$R = \omega_p NPpg + \omega_{gd} NPgd$$

Through network analysis of every single match, we are able to get every single formation Huskies have used in the season and their ratings.

Table 4: Huskies Performance in Different Formations

Formation	Appearance	Scored	Conceded	W	D	L	R
4-2-3-1	14	17	19	6	4	4	0.514
4-3-3	9	13	11	4	2	3	0.519
3-4-2-1	4	3	11	1	0	3	0.263
4-1-4-1	3	6	0	2	1	0	0.758
3-1-4-2	2	1	3	0	1	1	0.225
4-4-2	2	2	5	0	1	1	0.213
4-4-1-1	1	0	2	0	0	1	0.075
4-2-2-2	1	0	1	0	0	1	0.1
3-4-3	1	1	5	0	0	1	0.025
3-5-2	1	1	1	0	1	0	0.375

As can be seen, the Huskies have tried 10 formations in the season, most of the time they were using $4-2-3-1$ and $4-3-3$. However, neither of them are as successful as $4-1-4-1$. In order to measure the fluctuation of a team's performance, we define the flexibility indicator as the standard deviation of formation rating R.

$$FI = \sqrt{\frac{1}{N_R}\sum(R_i - \bar{R})^2}$$

From above we can get the flexibility indicator of Huskies, $FI = 0.233$. But how do we know whether it's relatively flexible or not? We decide to pick five real clubs from 2018/2019 Premier League (5), their basic performance in that season are listed below:

Table 5: 5 Real Clubs Performance

Club	Season Ranking	W	D	L	Most Used Formation
Man City	1	32	2	4	4-3-3
Arsenal	5	21	7	10	4-2-3-1
West Ham	10	15	7	16	4-1-4-1
Burnley	15	11	7	20	4-4-2
Huddersfield	20	3	7	28	4-2-3-1

Using the same method, we get the FI of these clubs, and compare with of Huskies.

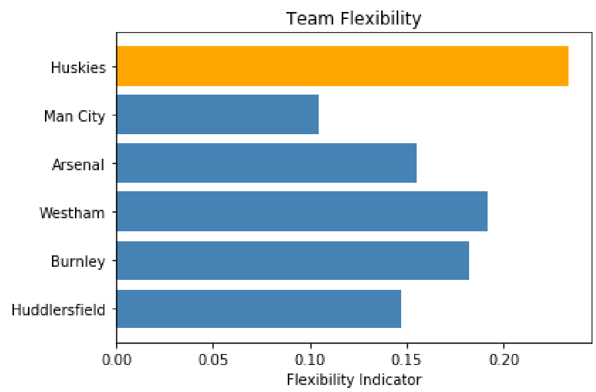


Figure 8: Flexibility of Huskies and 5 Real Clubs

Since FI indicates the fluctuation of a team's performance in different formations, the higher it is, the less flexible the team will be. In our case, we realize that Huskies is relatively inflexible. Therefore, Huskies' counter-strategy may not be as efficient their opponents', and they should focus more on the universal strategies.

4.1.3 Tempo

During soccer game, a team's tempo is reflected in the long time possession of the ball , constant pressure to the opponent's half and a great diversity of offensive approach to break though opponent's defend from time to time . What we need to do is take these aspects into account and present them to a specific standard.

To achieve this, we develop a zone-related model to track the ball movement under players feet. When the team gain the possession of the ball from its opponents (this normally come from outside throw-in, foul, steal, interception, etc.), we start to follow the movement of the ball until the opponent regain the control. To simplify the task, we equally divide the pitch into three parts: $one\ third$, $middle\ third$, $opposition\ third$. We assume that the closer the ball is towards the opponent's gate, the better the team's tempo is.

Under these assumptions, we process the data from Huskies 38 matches in the whole season, including both home and away matches shown in **Appendix C**. We summarize the time of the ball under control in the three zones, and get the proportion of the time in every 90 minutes when the ball is in these different zones.

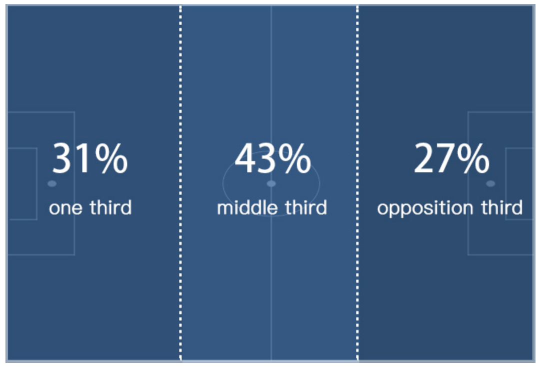


Figure 9: Huskies Tempo

These results give us suggestions on how to modify strategy to improve Huskies tempo and therefore have better match results. For example, by setting offside trap or high press, Huskies may be able to keep the ball in the opposition field for more time and be more threatening consequently.

4.1.4 Attack Position

Inspired by the method used in analysing the team tempo, we decide to focus more on the space utilization to reveal the team attack patterns. During the limited 90 minutes, it is extremely essential for teams to decide how to distribute their attack sources, attempts on different sides, shot directions and shot zones will lead to complete different results. Success of these distributions rely on the strong tacit agreements between players, if attackers do not get on well with each other, they may end up fighting for their own instead of implementing the team's tactic.

Using the similar method in analysing team tempo, we track not only the ball movement along each attacking side, but players' accelerations (especially wings) and every single shot position from the Huskies as well. **Figure 10** displays the Huskies' attack distribution along each sides, and obviously

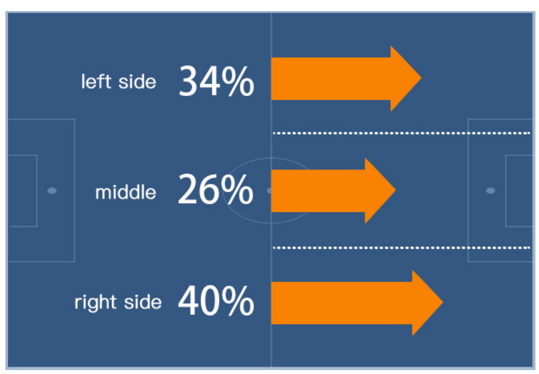
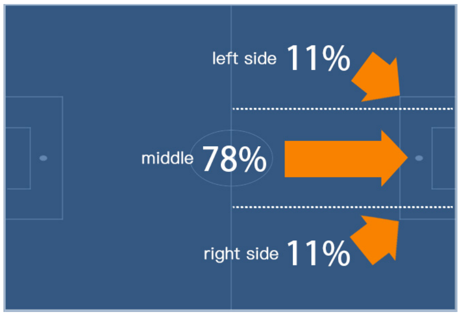


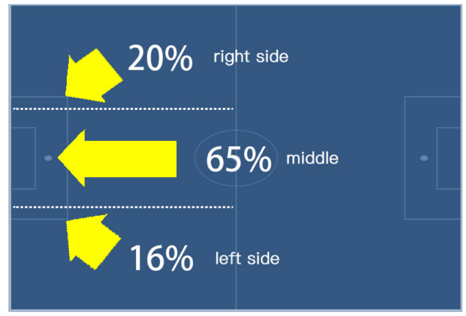
Figure 10: Attack Direction

the areas down the right side is more active than the others. This confirm our findings in the passing network, since players who often appear down the right side, such as M6, also shows great centrality in the network.

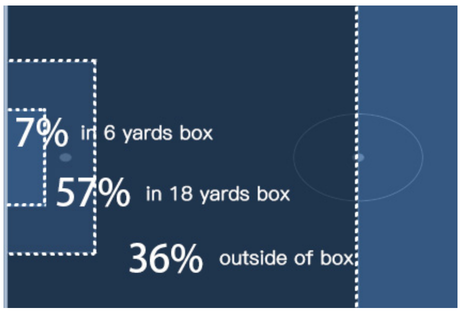
Shot directions shown in Figure 11 however, reveals that the Huskies spends more time shooting from the middle rather than from both sides. This is in fact contradict to their attack directions above, which gives us the idea that the Huskies' distribution of attacking sources may not be very effective.



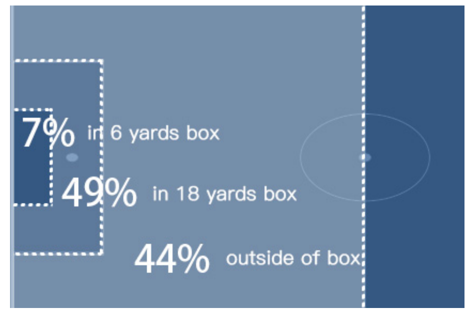
(a) Huskies Shots Direction



(b) Opponents Shots Direction



(c) Huskies Shots Zones



(d) Opponents Shots Zones

Figure 11: Shot Directions and Zones

Meanwhile, we take the shot zone into consideration to see whether the Huskies prefer to try shots from a long way out. Compared with opponents average level, we found that the Huskies prefer to shoot in the 18 yard box, which explains the contradiction we found above: the Huskies wings like to get to the byline and put some crosses in, which create lots of shooting opportunity for players in the middle.

4.2 Personal Performance

In this Part, we are going to analyze the distribution of contribution for each player. Firstly, we list events type which have positive influence on matches in all categories of events and main type of players:

Table 6: Categories of Events and Players

Positive Event Type	Main Player Type
Pass	Forward(used 'F' to represent below)
Shot	Midfielder(used 'M' to represent below)
Duel	Defender(used 'D' to represent below)

In addition, the degree of importance of each event type are always concerned differently for different player type. For example, the main task of forward is offensive so that shooting-data which is a powerful indicator of offensive should be highlighted more than dueling-data, but for defender,more dueling means they are more active in defence and shooting-data will not be so vital. Given that consideration, our team apply **Analytic hierarchy process (AHP)** to evaluate personal contribution. The flow is :

- Construction of pairwise comparison matrix(PCM) for each play role
- Consistency check & Calculate the weight of factors
- Analyze events data & Evaluate each indicator
- Calculate the total score for all players & Sort

4.2.1 Construction of Pairwise Comparison Matrix

Our group design three pairwise comparison matrix for forward, midfielder and defender,check their consistence and get $WeightVector$ finally. Indicator description order in column and row is Duel–Pass–Shot.

Table 7: Result of PCM Construction

Player Type	PCM	Consistence Check	$WeightVector$
F	$\begin{bmatrix} 1 & 1 & \frac{1}{2} \\ 1 & 1 & 1 \\ 1 & 1 & 2 \end{bmatrix}$	$CI = 0.0268$ $CR = 0.0516$	[0.2599,0.3275,0.4126]
M	$\begin{bmatrix} 1 & \frac{1}{2} & 5 \\ 2 & 1 & 7 \\ \frac{1}{5} & \frac{1}{7} & 1 \end{bmatrix}$	$CI = 0.0071$ $CR = 0.0136$	[0.3332,0.5917,0.0751]
D	$\begin{bmatrix} 1 & 4 & 9 \\ \frac{1}{4} & 1 & 3 \\ \frac{1}{9} & \frac{1}{3} & 1 \end{bmatrix}$	$CI = 0.0046$ $CR = 0.0088$	[0.7267,0.2000,0.0734]

4.2.2 Evaluate Indicator

Based on the given data package, our team define the scored method for each indicator as follow.

- **Dueling Index**:
  Duel per game $\bar{D}$ represent players activity in aspect of dueling during this match season.

- **Shooting index**:
  As we known, the closer shooting from the gate,the more threat it will make. So our team set Threat of Shooting S_{th} which relevant to inverse square of shooting distance(shooting distance is represented by distance between the shooting position and centre of gate) and the relevant parameter $Th_i = 1000m^2$.

$$S_{th} = \frac{Th_i}{(100 - EventOrigin_X)^2 + (50 - EventOrigin_Y)^2}$$

  Our team use total threat of shooting per game $\bar{S_{th}}$ to evaluate a player in aspect of offensive during this match season.

- **Passing Index**:
  Four categories of centrality are used to describe players performance in aspect of passing. Detailed definition and signification of these centrality are in the beginning of Section3-Passing Network Model. To simplify this evaluating system, equal importance are attached to four centrality. After normalizing and averaging of each centrality, we can get comprehensive passing activity P_c which means the significance of player in pass-network.

4.2.3 Calculate total score & Sort

Applying the evaluating system mentioned above we can $\bar{D}$,$\bar{S_{th}}$ of all players in this match season (included players in huskies and opponents) and P_c of every Huskies' players. In normalization of each indicator data, classification of player type should be considered. For example,in most cases, forward shot times always greater than defender. So that different indicator have no comparability among different player roles.

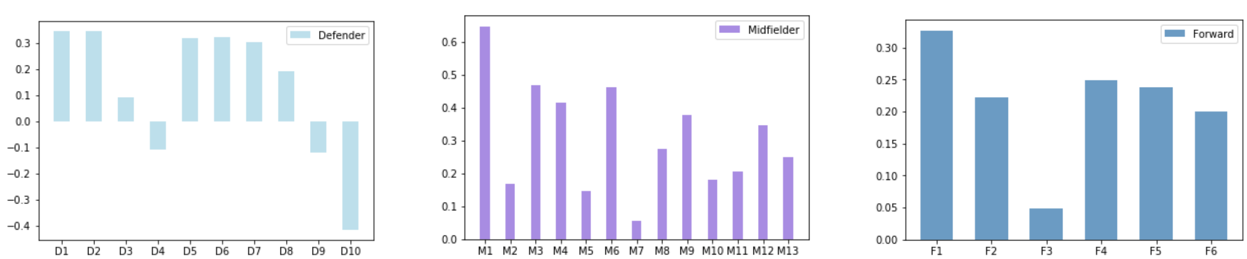
Taking shooting index of forward as an example : Shooting index sub-table of forwards in all team can be filtered from Shooting index table. Considered special point, we choose *zero-centered* to do normalization, which can eliminate the influence of extreme point. Finally, Data of players in Huskies can also be filtered from the sub-table.

$$\textit{zero-centered}\ \text{function: } Index = \frac{Point - \bar{Point}}{\sigma point}$$

Using this method, each player in Huskies will get index vector respectively and ,then, result of multiplying index vector and corresponding weight vector is total score for Huskies' players.

4.2.4 Result & Explanation

Final result are shown as picture as follow and detailed result data can be found in first appendix.



(a) Defender (b) Midfielder (c) Forward

Figure 12: Score Ranking

In **Figure 12**, zero line means average contribution of all players(Huskies and Opponents) compared with same type players. The higher of the bars, the higher distribution of contribution above average of players they have.

For defenders, there are only five members' contribution for their teams above average, four around average and one below average. Doing Horizontal contrast with midfielder and forward, level of defender are jagged and have a great difference, which will likely lead to a poor cooperation in aspect of defense and link with midfielder.

Fortunately, responding contribution of all midfielder and forward in Huskies are above average, which means the overall level of respective type players are in the middle and upper reaches so that aspects of conversion between offense and defense and seizing opportunities of goal can get positive evaluation.

5 Advice to the Huskies

5.1 Huskies Strengths and Weaknesses

With the help of Passing Network Model and Teamwork Performance Model constructed above, we are able to get the strengths and weaknesses that the Huskies have shown through all the matches. We believe it is important for the Huskies to see their advantages and disadvantages to make progress in the future.

Table 8: Huskies Strengths and Weaknesses

Strength	Weakness
Attacking down the wings	Converting chances to goals
Attacking free kicks	Individual defence ability
Creating scoring chances	Familiarity with different formations
Stealing and interception	
Air duels	

5.2 Successful Strategies

Based on the team performance analysis, particularly the evaluation of Huskies' flexibility, tempo and attack positions, we found several strategies that have been proved effective for the Huskies, as is listed in aspects of attack and defence.

Table 9: Huskies Successful Strategies

Attack	Defence
Attacking down the right wing	Offside trap
High passes and long balls	Rotate first eleven
Attempt crosses	Being aggressive
Try shots a lot	

5.3 Recommended Formations

Finally, having combined these findings and analysis of the Huskies' players distribution of contribution, we recommend the Huskies coach with two first elevens: 4-1-4-1 and 4-3-3, which we believe could be effective for them in the future.

Figure 13: Recommended 4-1-4-1

Figure 14: Recommended 4-3-3

6 Generalization and Inspiration

As what have been discussed above, we have developed Ball Passing Models and Teamwork Performance Model to evaluate the cooperation level of a soccer team, the Huskies. We do believe that similar approaches can be applied to analyze collaboration of other team sports, and even general teams in diverse backgrounds as well. For instance, we can construct the email passing network in a company to study the relationship among employers, identifying certain cooperative patterns. We only need to get the additional data of the number of email sent. Besides, we can use the method in determining the distribution of contribution among players in a soccer team to help identify contribution among researchers in a laboratory. Understanding this cooperation patterns can help to build more effective teams and thus increase teamwork efficiency.

7 Strengths & Weaknesses

7.1 Strengths

- In **Section 3**, our team analyze the passing network at both macroscopic level (Season and Single Game) and microscopic level(Minute-to-Minute). In this way, besides having an overview at the passing strategies frequently used by the Huskies, we can also get the detailed passing patterns of the team, which facilitates the analysis followed.
- In **Section 4**, indicators of cooperation are discussed in two hierarchies: Team level and Personal level.
- In **Section 4.1**, team performance is evaluated in various aspects, analysing results of the Huskies have been contrasted with real life teams.

- In **Section 4.2** which used Analytic hierarchy process (AHP), Each type player have their own evaluation parameters, and therefore the final result have higher rationality.

7.2 Weaknesses

- When our team work out the **3**, to simplify the model, We calculate average passing origin coordinate and use it to identify the passing active zone of each player. In fact, those zones may depend on passing destination coordinate, pass success rate or other data.
- In **Section 4.1**, some variables are base on rough assumptions due to the lack of specific data.
- In **Section 4.2**, ratio in pairwise comparison matrix are defined with subjective so that it will have a certain degree of deviation from reality.

References

[1] W. Cao, "Visualization analysis on passing technique of spanish tiki-taka tactics based on social network analysis method," *Journal of Chengdu University of Physical Education*, vol. 45, no. 04, pp. 65–72, 2019.

[2] `https://networkx.github.io/documentation/stable/reference/algorithms/`.

[3] I. J.M.Buldú, J.Busquets, "Defining a historic football team: Using network science to analyze guardiola's f.c. barcelona," *Sci Rep*, 2019.

[4] "Liverpool's 2019/2020 stats." `https://www.premierleague.com/clubs/10/club/stats?se=274`.

[5] "Premier league 2018/2019 table." `https://www.premierleague.com/tables?co=1&se=210&ha=-1`.

A Appendix 1 : Centrality & Clustering Data

Table 10: Centralities and Clustering for Huskies

player	DC	CC	BC	EC	$Clustering$
G1	0.897	0.763	0.050	0.158	0.00099
D1	1.000	0.935	0.122	0.212	0.00107
D2	0.966	0.879	0.000	0.202	0.00093
D3	1.000	0.935	0.118	0.215	0.00097
D4	0.828	0.879	0.110	0.201	0.00105
D5	0.931	0.906	0.074	0.217	0.00098
D6	0.931	0.853	0.004	0.199	0.00070
D7	0.862	0.879	0.101	0.208	0.00090
D8	0.690	0.763	0.011	0.179	0.00084
D9	0.448	0.604	0.000	0.090	0.00054
D10	0.310	0.580	0.000	0.070	0.00056
M1	1.034	1.000	0.467	0.228	0.00134
M2	0.586	0.674	0.000	0.138	0.00054
M3	0.966	0.935	0.091	0.211	0.00117
M4	0.966	0.935	0.004	0.219	0.00093
M5	0.517	0.644	0.000	0.120	0.00042
M6	0.966	0.935	0.054	0.221	0.00091
M7	0.310	0.569	0.000	0.064	0.00044
M8	0.724	0.707	0.000	0.154	0.00070
M9	0.724	0.853	0.000	0.197	0.00052
M10	0.483	0.659	0.000	0.125	0.00036
M11	0.655	0.725	0.000	0.155	0.00037
M12	0.655	0.763	0.000	0.177	0.00074
M13	0.586	0.690	0.000	0.146	0.00047
F1	0.931	0.935	0.000	0.222	0.00072
F2	0.931	0.935	0.172	0.219	0.00137
F3	0.552	0.725	0.000	0.153	0.00045
F4	0.724	0.906	0.000	0.217	0.00060
F5	0.759	0.829	0.000	0.198	0.00062
F6	0.690	0.829	0.000	0.187	0.00079

B Appendix 2 : Single Game Passing Network

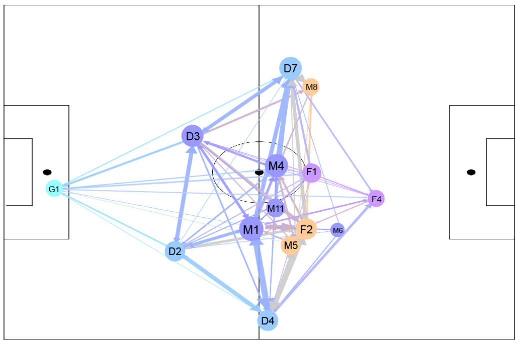


Figure 15: Passing Network for Game 10

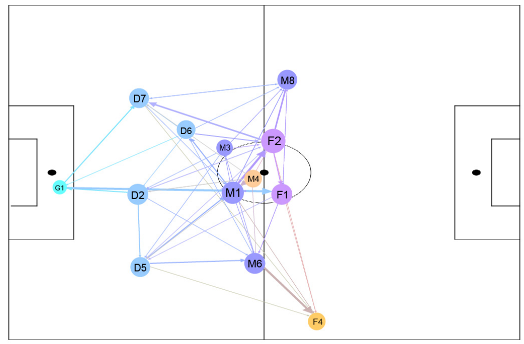


Figure 16: Passing Network for Game 16

C Appendix 3 : the Huskies Home & Away Tempo

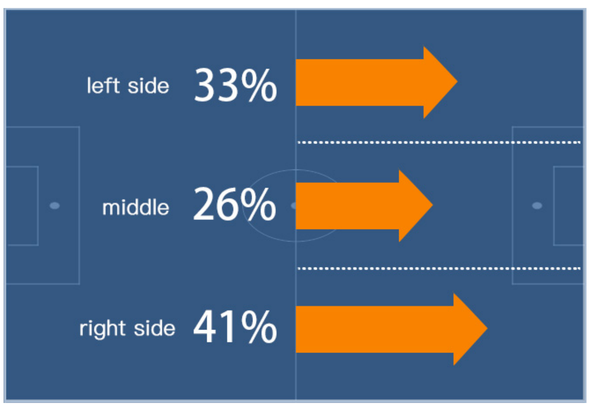


Figure 17: home

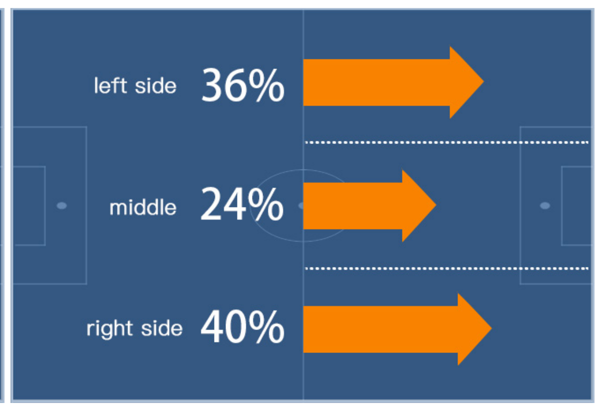


Figure 18: away

D Appendix 4 : Detailed Evaluation Data

Table 11: Detailed Evaluation Data

PlayerID	S_{th}	$\bar{D}$	$S_{th}Index$	$\bar{D}Index$	P_cindex	Sum	Ranked
F1	21.171	28.031	0.048	0.297	0.703	0.327	F1
F2	12.55	15.839	-0.008	-0.132	0.792	0.222	F4
F3	8.284	9.143	-0.036	-0.367	0.482	0.048	F5
F4	27.949	19.773	0.092	0.007	0.64	0.249	F2
F5	13.348	24	-0.003	0.156	0.608	0.238	F6
F6	10.438	21.786	-0.022	0.078	0.579	0.201	F3
M1	3.524	22.091	-0.03	0.175	1	0.648	M1
M2	1.34	6.571	-0.073	-0.301	0.462	0.167	M3
M3	15.378	17.393	0.205	0.031	0.747	0.468	M6
M4	5.246	16	0.004	-0.012	0.709	0.416	M4
M5	3.803	6.6	-0.024	-0.301	0.418	0.145	M9
M6	4.917	18.963	-0.002	0.079	0.738	0.463	M12
M7	0	6	-0.099	-0.319	0.288	0.057	M8
M8	9.693	12.562	0.092	-0.118	0.521	0.276	M13
M9	2.371	18.75	-0.053	0.072	0.605	0.378	M11
M10	3.791	10	-0.024	-0.196	0.419	0.181	M10
M11	0	7.714	-0.099	-0.266	0.51	0.206	M2
M12	11.846	17.812	0.135	0.043	0.544	0.346	M5
M13	5.155	13.5	0.003	-0.089	0.474	0.251	M7
D1	12.138	19.433	0.225	0.237	0.773	0.344	D1
D2	9.417	15.304	0.149	-0.061	0.675	0.102	D6
D3	9.769	14.72	0.159	-0.103	0.774	0.092	D5
D4	2.294	11.455	-0.049	-0.338	0.699	-0.11	D7
D5	1.924	19.524	-0.06	0.244	0.729	0.319	D8
D6	10.318	19.533	0.174	0.245	0.659	0.322	D2
D7	1.528	19.263	-0.071	0.225	0.711	0.301	D3
D8	7.686	17.5	0.101	0.098	0.56	0.191	D4
D9	4.098	12.5	0.001	-0.263	0.359	-0.119	D9
D10	5.78	7	0.048	-0.66	0.297	-0.417	D10