Problem Chosen	2020 MCM/ICM Summary Sheet	Team Control Number
D		1234567

An ICM Paper Made by Team 1234567

Summary

People around the world are complaining about the traffic jams every $day^{(1)}$. Some people may spend more than 3 hours on their way from office to home everyday on their cars, which makes them frustrated. And our problem is to investigate the methodsto optimize the design of toll plazas on highways, especially the area after merging,which is also a possible location where vehicles can be stuck in.

Before establishment of our models, we list some assumptions to make the real lifescenario easier to model.

And then we start to analyze the existing models, from which we conclude their strengths and weaknesses. By investigating their characteristics, we get the inspiration to form our two new models: Control Time Model (CTM) andWaiting Area Model(WAM). In these two new models, we introduced a way of conducting control of the departure time of the vehicles at the tollbooths. This kind of control is basically preventing the vehicles which are merging in the same lane from leaving the booths simultaneously, which may be a hidden danger for traffic accident or leads to a traffic jam. And we?ll continue to calculate the size and shape of the merging area according to our control methods and some assumptions. After that, we introduce a method for finding the optimal merging pattern based on both mathematical proof and computer simulations.

After that, we run some simulations to find out the throughput, risk and cost of different models, which are based on some statistical laws for real life situations. Then we compare these three models in all factors with the help of statistical hypothesis testing, and conclude that the CTM is the best in general. The next section is about some slight modification under different conditions such as including the self-driving car, and the different arrangements in terms of merging patterns when the proportion of different types of tollbooths vary.

For the following section, we test our model by investigating the sensitivity of construction cost and throughput (per hour) in terms of some variables included in our model, in order to justify the reliability of our model from different perspectives.

Finally, we end our report by the conclusion part followed by strengths and weaknesses analysis.

Contents

1 Introduction **2**
1.1 Background . . . 2
1.2 Restatement of the problem . . . 2
1.3 Literature review . . . 2

2 Assumptions **3**

3 Notations **4**

4 Models **4**
4.1 Passing network model model(PNM) . . . 5
4.1.1 Assumptions . . . 5
4.1.2 Network . . . 5
4.1.3 Full season analysis . . . 6
4.1.4 Individual game analysis . . . 7
4.1.5 Multiple scale analyse . . . 9
4.2 Team performance indicator model(TPIM) . . . 10
4.2.1 Factor . . . 11
4.2.2 Multiple regression analysis . . . 16

5 Suggestions and inspiration **17**
5.1 Effective structural strategies . . . 17
5.2 Players' star analysis . . . 18
5.3 Network analysis . . . 19
5.4 New season suggestions . . . 19

6 Build an excellent team **20**
6.1 Control Time Model (CTM) . . . 21
6.2 Waiting Area Model(WAM) . . . 23

7 Analysis and Results **24**
7.1 Throughput . . . 24
7.1.1 Throughput of existing model . . . 24

8 Strengths and Weaknesses **24**
8.1 Strengths . . . 24
8.2 Weaknesses . . . 24

Memorandum **25**

References **25**

Appendix A: Further on LaTeX **26**

1 Introduction

1.1 Background

The world is becoming more and more challenging, many emerging problems (from academic problems, commercial problems to global economic problems like global warming, etc.) are becoming hard for a person, an organization, or even a country to solve without cooperation. As the smallest unit of cooperation, "teams" are becoming more and essential these days, which require a group of people, but should not just be "a group of people".

To sum up, there are two key differences between a "team" and "group".

- First, a team needs a specific common target for everyone inside it (while a group does not necessarily need).
- Second, a team will have its own structure (which is not just a simple "add"of individuals) and own dynamics (which is just not a simple "mix" of personal interactions).

With different structure and dynamics, some teams will be able to complete some complex duties which is impossible for a simple addition of individuals' effort, while others may perform worse than individual. Therefore, topics about how team structure and strategies can affect a team's performance is raising more and more attention in research area.

1.2 Restatement of the problem

We are required to explore the construction of a strong team in competitive team sports, and we choose to study about football team formation and optimization. We get the data of performance of Huskies last season, based on it, not only the behavior and the interaction leading to scoring but also the team's dynamic changes throughout the season can be acquired, then we build a network model. The model will show the performance indicators of teamwork and team performance. Generally speaking, it reflects the structure, configuration, and dynamics of team work. After finishing the above work, we're able to give suggestions both on adjustment for the whole team and direction of changes for specific players. After the research on the Huskies team is completed, we will generalize the results to the more general team formation problem. We will talk about the construction of a strong team that can solve complex problems and the performance issues needing to be considered for the team, based on our model. In short, our work is just next three tasks:

- Using the network model, we build a model to reflect the structure, configuration, and dynamics of team work at multiple scales.
- Based on the results of our model, we provide strategies on adjustment of the whole team and corresponding suggestions for specific players.
- We should generalize the model, deliver our opinions on the construction of a strong team able to work out complex problems.

1.3 Literature review

The research on the quantitative analysis of football matches with models can be traced back to the early 1950s, when Charles Reep tries to manually collect data in football games to explore the key to goal and winning [1]. In late 1970s, Gould, P. and Gatrell first put forward the method of using network to study football match [2]. Although Gould P.'s work did not raise much attention at that time, thanks to the development of sensing technology and data processing ability [3], using network to study football team becomes prevalent in both research area and some professional teams.

Most of the research's network uses passing events to determine the weight of edges between nodes (which represents players). However, out of different research purposes, different method to determine the edge may be used. For instance, some networks use the total number of bi-direction passing events [4] to describe the cooperation between two players while others may use one direction passing events [5] to study the progress routine. To study teamwork in different time scale, the data source for a network's creation may also range from minutes [5] to a whole season [4].

2 Assumptions

To make our problems easier to deal with, we make the following basic assumptions, each of which is properly justified.

- **Because the amount of data is not very sufficient, we do not consider the influence of coaching factors on the game**.

- **In order to facilitate the evaluation of each player's ability value, we assume that each player has played his full strength in each game under the existing tactical structure**.

- **The impact of eliminating out-of-court factors on players' performance on the pitch. For example, the impact of life events on player performance is not considered .**.

- **We assume that there is no locker room issues and no personal grudges between players**.

3 Notations

Notation	Definition	Unit
x_{ij}	$EventOrigin_x$ of the player(ij correspond the player's field position, eg.X_{F1})	N
y_{ij}	$EventOrigin_y$ of the player(ij correspond the player's field position, eg.Y_{F1})	N
I_{ij}	The importance of the player	N
P_{ij}	The number of passes made by a player	N
A_{ij}	Agreessiveness of the player	N/A
D_{Tn}	Diversity of the team	N/A
C_{ij}	Coordinate ability of the player	N/A
S_{ij}	Skill of the player	N/A
P_{Tn}	The possession rate of the team	N/A
t_{ei}	The length of time before the conversion of the "event" subject	s
T	Total duration of a game	N/A
P_{Aij}	Agreessiveness of the team	N/A
D_{Aij}	The number of attack duel	N
D_{Dij}	The number of defend duel	N
T_{Tn}	The tempo of the team	N/A
P_c	The degree of cooperation of the team	N/A
S_{Tn}	The score without home and away factor of the team	N
S_{ci}	Team's goal difference (may be negative)	N
S_{Tnha}	The score with home and away factor of the team	N

4 Models

The flowchart above gives an overview about our whole process of model building and the relation between those models.

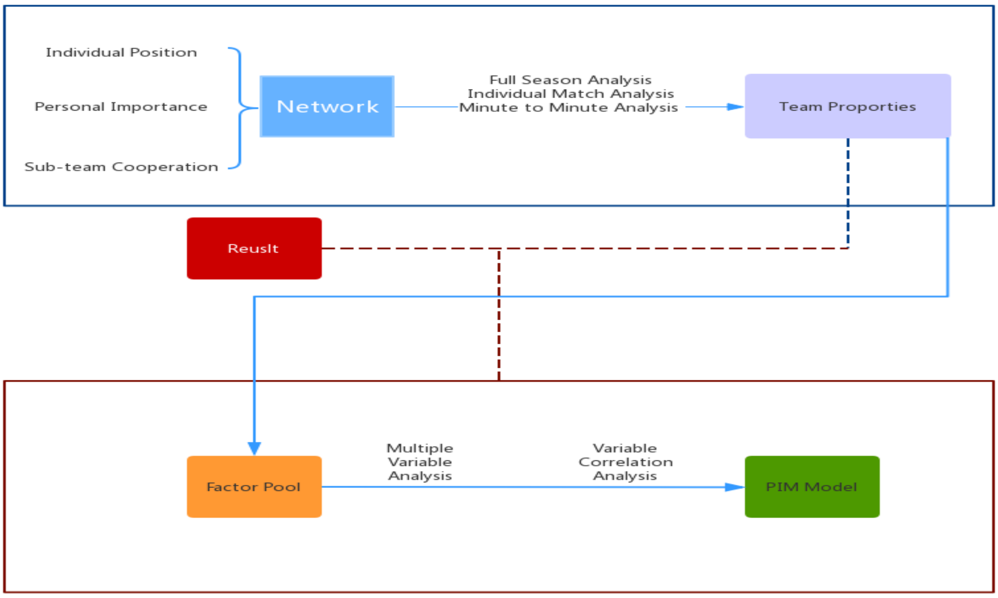


Figure 1: flowchart

First, we use the data of passing to build a Network Model, which reveals the individual position, personal importance, and sub-team cooperation for a team in each match this season. Then, based on these, we do three kinds of analysis on Huskies to "draw a portrait" for this team. The analysis integrates for three subparts, the full season analysis, the individual match analysis, and the minute's scales analysis. From the portrait of Huskies, we could know what property this team holds, but for that moment, we still could not know whether these properties could be beneficial to this team or quantize how beneficial it could be for Huskies. In order to know whether these properties of Huskies could be beneficial and quantize the benefit (or harm), we build a Team Performance Indicator Model. In this model, we create a "factor pool," which contains many factors that may be a team's performance indicator. Then we define a series of indexes to capture those factors from the data source "fullevents.csv." Every team (including Huskies and all of its opponents) will get a score for each index. Then we will do multi-variable analysis and relation analysis to give an explanation about how those factors can affect a team's performance in the soccer match.

4.1 Passing network model model(PNM)

We consider the influence of several factors and establish a Free propagation model to reveal the development principle of the AIDS under on intervention.

4.1.1 Assumptions

- **We assume that all players in this game can be replaced with the eleven players who have played the most.** Here we ignore the goalkeeper problem. By studying the cooperation between the eleven players who have played the most time to reflect the tactical arrangement of the entire team in this field
- **We won't take long-term or frequent injuries into consideration so that the data reflects exactly the normal level of every player.**
- **The range of players' activities on the court is limited and fixed, and their frequency of occurrence at other positions is lower or just zero**. In football games, players have their own responsibilities and corresponding activities in the team. For example, the position of the defender and goalkeeper is relatively backward, and the Midfielder is relatively forward. Based on this, we can use the average of the coordinates of the player's active position in all matches as the player's coordinates on the visualization.

4.1.2 Network

We used MATLAB to retrieve the data of all players in the passing event of each game and presented them with visual tools, the code will be presented in the appendix. We pick out five typical games among these thirty-eight games to analyse Team Huskie's network properties. These five games are picked out according to the result of the game, represent a big win, a big loss, a small win, a narrow loss and a draw respetively. During our seletion process, for the sake of their representativeness, when picking out the narrow loss and small win games, we try to find opponents whose team strength is close to Huskie team as a reference. Since Huskie Team will fight against with other teams twice every season, one-small-win&one-narrow-lose's opponent will become our first choice. So that the score we get from each game can better match the strength of the two teams. We also create the season's network of the team to analyse Team Huskie's network properties macroscopically.

The network includes three main elements: position, importance and cooperation degree

- **position**

  Accdrding to our assumption, the position of each player can be determianed by the average of active position in the whole game , which can be represented as:

$$(X_{ij}, Y_{ij}) = (\frac{1}{N}\sum_{t=0} x_{ij}, \frac{1}{N}\sum_{t=0} y_{ij})$$

- **importance**

  In order to simplify our model, we consider the total number of passes made by a player as the degree of his importance. The number includes all of his records in the passing event.Since there are no accurate data about the scoring, we think that the more times the player touchs the ball, the more involved for him in this game, which indicates that he has the higher tactical status. In our furtherwork, the pass rate, ballpossession and the ability to face interference will be taken into consideration. We will build a model in a weighted sum to estimate each player's importance more accurately. In this part, the importamce can be presented simply as:

$$I_{ij} = \sum_{t=0} P_{ij}$$

  and the more important the player in the game, the node of him wil become bigger accordingly.

- **coorperation degree** Since all of the statistic and the analysis are based on the player's performance on the ground, the best indicator of the cooperation degree between two players is the passing times between them. In our network, If there is only one passing movement between two people, a line will be drawn between them, and with the passing time increasing ,the line between them will become thicker, and we can evaluate the cooperaiton degreee of two players with the thickness of the line directly.

4.1.3 Full season analysis

Let's start with an analysis of the team's performance over entire season. The following Figure1 is the network of the entire season.

Accroding to the network, we can find out eleven players who were used most this season, and we can also figure out Huskies Team's offensive preferences, followings are our analysis of the network

- In terms of staffing, we can find out that Huskies Team has the abundant defender options, there are five defenders in the eleven-man roster, and we can speculate that Huskies Team is lack of good forwards. However, there exists problem that forward has important tactical status, but the node of the forward is bigger and the lines between him and other nodes are so thick that we can considere him as decisive in the game. We will discuss this problem in the following analysis of individual games to see whether the forward has great influence on the team, since different teams have different ways to win the game and Huskies is likely to rely mainly on strong defenders to achieve success.

- In terms of player positions, we can see that the strong side of the team's overall offense is the left side. Since the position of players are average results, it's hard to distinguish

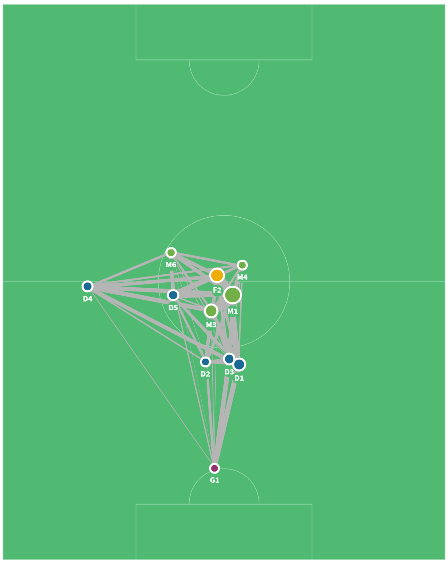


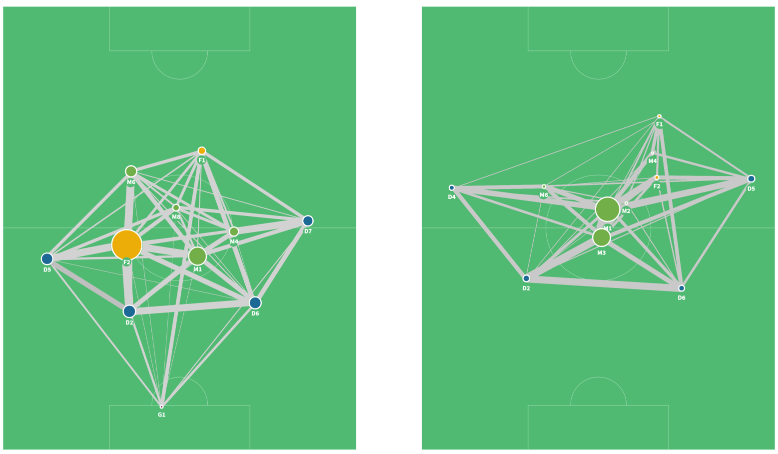
Figure 2: season analyse

whether the team is balanced in the left and right side or focused on the centerline. But there is no doubt that the left side can be their first choice when they lead off the attack especially the defenders Almost all the defenders prefer lift side attraction. There is also an interesting fact that M6, M4's $\bar{y}$ is bigger than F2, which means they are more aggressive while they are attacking, and it also reflects the forward's problem.

- In terms of runing of the ball. The main driver of the ball movement is the midfield, and it is accord with our common sense, forwards and midfielders handled more of the ball than defenders, and in general, the forwards and midfielders interact more with other positions, too.

4.1.4 Individual game analysis

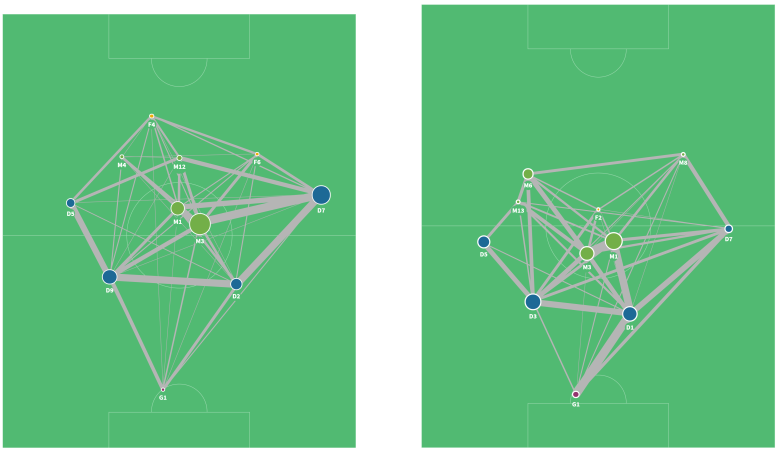
We pick out five typical game's network to identifyt the network's properties.



(a) match34:big win (b) match6:small win

Figure 3: win

These are representative wins and losses and the following is a representative tie.



(a) match26:big lose (b) match21:narrow lose

Figure 4: lose

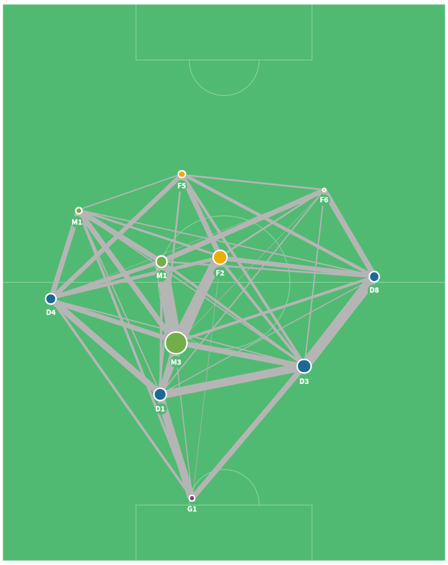


Figure 5: tie

According to the network, we can find out how the positions of the players and the team formation affect the games' result, and we can dig out their team pattern at the same time.

- Huskies's midfields have an obvious dyadic configuration. Every game we always find a pair of midfielders who cooperate with each other very closely. And whether they form this kind of cooperation on the court has little influence on the result of the game. So we tend to think that this is a kind of fixed tactics, which is specially designed for the midfield of the team. At the same time, the cooperation between the four defenders is very stable, and the relationship between the defenders is very close in each game. This can prove our previous conjecture that the commonly used four back team is used to maintain the stability of the team's performance. When we look up the data of other games, we find that D4,D2,D5 are the guards who play relatively frequently. It can be regarded as a kind of ternary structure, a fixed tactic of the team. Further contact and cooperation between players will be discussed in the next section.
- In response to the above questions about forwards, we found that in the vast majority of games, the forward's passing is not the majority, but in the big win, F2 plays a pivotal role

in the game, and a large number of passes and catches directly bring victory to the game. We have looked up the data of other victories and found that this is not an accident. When the forward plays a good game, the team's performance is not bad. But for the team, how to make the forward play better still seems to be a question. In most of the matches, we can see that the average position of forward movement is often lower than that of midfield. To some extent, it can be considered that the offensive participation is not high enough. It may be related to their own strength and attitude.

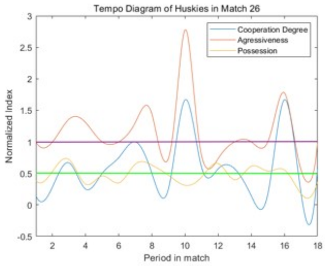
- From the perspective of the team as a whole, the main style of the team is defensive. Not only it's the personnel allocation but the specific activity scope of the person shows that this is a team that wins the game by defending and counterattack. The team is used to putting on a lineup of four defenders, and at least three midfielders are guaranteed at the same time. As for the situation of the striker, it is reletively casual, and the proportion of passing that the striker gets in the game is also very small. The team has a fixed formation in most games. This ensures the relative stability of their performance, but at the same time leads to the failure of players just to run to a fixed tactical position without timely adjustment of the form on the field. It's not hard to find out from several big wins that the layout of the team is different from that before, or the change of the core passer, or the difference of the player's position.

4.1.5 Multiple scale analyse

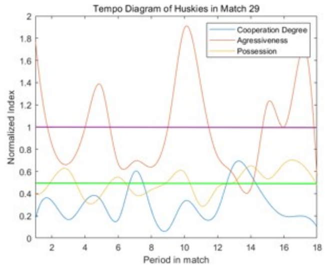
When football fans are talking about a team's performance and style in a specific match, they will often refer to a team's "tempo", which is an important time-related team property. A good tempo may indicate a team coordinate well and is in control of their team duty now.

To quantize a team's "tempo" in a specific match, we divide each 90-minute game into 18 small five-minute intervals to see the fluctuation of a team's property in each interval. We choose "five-minute" rather than smaller interval because if the time is too short, the events happen in this interval will not be enough to identify team property. Here we use three team property to portrait a team in one interval, the Cooperation degree (describe how well the team cooperate and connect), the Aggressiveness (describe the degree a team is eager to attack) and the Procession (describe the team's ability to control the ball). The mothed to calculate these three indexes will be explained in detail later.

We calculate all these three indexes for every 5-minutes interval in each match played this season and plotted them. Since 5-minute interval too large to make a smooth plot (but we could not choose smaller intervals for the reason address above), we use Lagrange's interpolation to show the tendency of the change of a team's tempo.



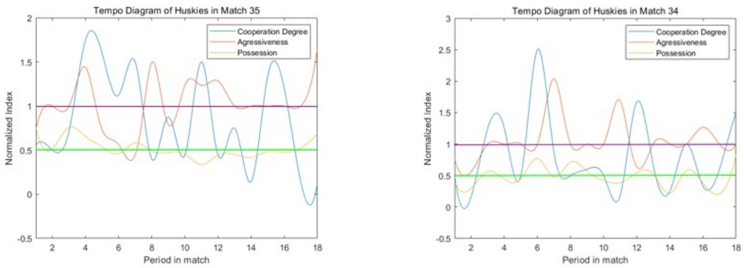
(a) figure 1 title.



(b) figure 2 title.

Figure 6: general title.

After cross comparison of these plots, some possible patterns did emerge. We find that the wining matches (figures above are two examples of winning match, others are attached the appendices) will more High Cooperation Peaks (the peaks of cooperate satisfies being larger than 1, i.e. the purple line) while the losing match(figures below are two examples of winning match, others are attached the appendices) will have less. On average, the winning matches will have 3.97 High Cooperation Peaks while losing matches only have 1.91. One explanation may be when a team is on good tempo, it can organize effective attack climax more frequently since in we definition cooperation degree equals to successive consecutive passes divided by inconsecutive passes.



(a) figure 1 title. (b) figure 2 title.

Figure 7: general title.

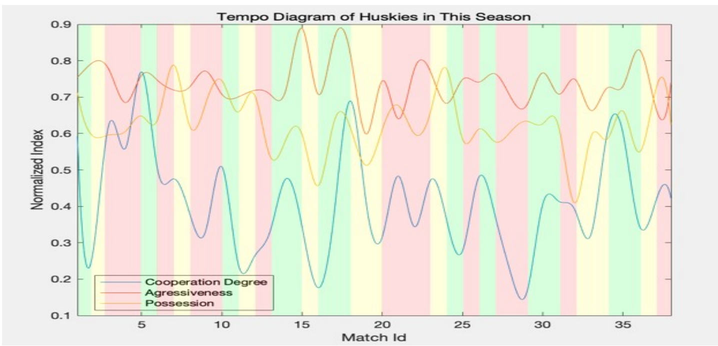


Figure 8: pic1

And in figure 7 (this figure shows the overall performance of Huskies in the whole season, the red area means loss, the green area means win, the yellow area means tie.), we can also see that high cooperation peaks of season fall more frequently in green area than in red and yellow. However, it is hard to identify obvious pattern for the tempo of aggressiveness and procession. Our conjecture is that those two factors are not strong as the cooperation degree, so they cannot obviously dominant other factors even when they are in high effective peak. Our later TPIM model will deal with this problem.

4.2 Team performance indicator model(TPIM)

In order to better study the status of the team in all aspects and put forward the most reasonable optimization plan, we set up a team performance indicator model, which quantifies the data of players on the field with specific indicators, and then obtains the areas where the team with good distance needs to be improved. We first selected some indicators that may affect the overall efficiency of the team, and then multiple regression of each factor through the method of variance analysis. Finally, we got several most critical factors. After giving a reasonable weight, we scored each team, compared the Huskies team with a better team, so as to get a clear optimization direction.

4.2.1 Factor

- **diversity**

  (...the importance of diversity)In order to identify a team's diversity, we use the diversity of players as a reference. We introduced a star system to evaluate each player's various abilities, that is, we score one to three stars for each player's abilities according to our data. The sum of variance of players' ability values in a team can reflect the diversity of a team, The greater the variance, the stronger the diversity. Considering that the number of players in each team is not the same, we have to calculate the average after getting the sum.

  First of all, we have determined three basic indicators to measure the ability of players: Aggressiveness, skill, coordination ability.

  - Aggressiveness is defined as the ratio of duels to all events. We think that players' performance in duels can directly reflect their desire to attack and judge their aggressiveness accordingly.
  - coordination ability is defined as the the ratio of passes to all events. Passing is one of the most important part of the game. And the number of passes between one player and others can directly indicate his cordination ability.
  - Skill is defined as the ratio of excellent pass to all passes, the excellent pass is the total number of passes minus the number of simple passes, and higher the ratio, the more skillful of this player.

  Our star rating standard is, firstly, get the ranking of each player's abilities in the same position player group, secondly, set three stars to the top 20 percent, one star for the last 20 percent, and the remaining 60 percent is two stars. Thirdly, Calculate the sum of variances of all ability values of all players in a team as the team's diversity indicator. In the calculation process, we weight the three indicators as the same, and the caculation can be expressed as:

$$D_{Tn} = \frac{1}{N}\sum(A_{ij}-\lambda)^2+(S_{ij}-\lambda)^2+(C_{ij}-\lambda)^2 \quad (\lambda = \frac{A_{ij}+S_{ij}+C_{ij}}{3})$$

  Based on the previous five teams, we selected another five teams based on the situation against the Huskies to make our results more accurate. The following is the sum of the variances of the indicators of the ten teams of players. Considering that the number of different teams may be different, we divide the variance and the total number of the team as the final degree of diversity.

Team	Diversity degree
Huskies	0.42962963
op14	0.365079365
op18	0.333333333
op6	0.38888889
op9	0.239316239
op11	0.259259259
op13	0.263888889
op16	0.666666667
op17	0.326797386
op7	3.555555556

- Degree of cooperation

  We chose the number of continuous passes in a game of the team as an indicator of the degree of cooperation. Considering the difference in the number of passes by different teams, we define the degree of coordination as the number of successful consecutive passes divided by the number of failed passes.Three passes without being truncated can be defined as a successful continuous pass. The following is the degree of cooperation of these teams in the game they played.

Team	Degree of cooperation
Huskies	0.529287
op14	0.68378
op18	0.69076
op6	0.31158
op9	0.99954
op17	0.3905
op7	0.38571
op11	0.50042
op16	0.865415
op13	0.4067

- **Possession**

  Our possession rate does not exactly refer to the proportion of the time that a team player takes the ball to the total length of the game. What we need is the percentage of participation by the entire team. So we calculate the length of time before the conversion of each "event" subject, and add up the length of each segment, the proportion of the total time of the standing game is the team's ball possession rate. The calculation process can be expressed as:

$$P_{Tn} = \frac{1}{T}\sum_{i=0}^{x} t_{ei}$$

  where x represents the total number of events for subject transformation Here are the ball possession rates of these teams

Team	Possession
Huskies	0.45589
op14	0.53681
op18	0.54832
op6	0.46852
op9	0.47572
op17	0.48054
op7	0.49884
op11	0.49996
op16	0.58112
op13	0.43193

- **Aggressiveness**

  Unlike metrics that measure player aggression. We choose the ratio of offensive duel to defensive duel in a team. This is used to characterize the offensive and defensive tactical configuration of a team during this period of time, so as to determine what kind of operating state it is in at the moment. In our hypothesis, offensiveness is positively related to the team's performance during this time. The expression can be described as:

$$P_{Aij} = \frac{D_{Aij}}{D_{Dij}}$$

  The following is the average of the aggressiveness of each team in each of the eighteen mini-games. We analyzed the data for each game and calculated the average.

Team	Agreessiveness
Huskies	1.004405
op14	1.0586
op18	1.0347
op6	0.99389
op9	1.0551
op17	1.0075
op7	0.96902
op11	1.0058
op16	0.995875
op13	1.0287

- Tempo

  Describe the tempo of a team is complicated. Therefore, we use a combination of multiple factors to reflect changes in team tempo. These indicators correspond to the above-mentioned possession, degree of coordination and aggressiveness respectively. We believe that these three indicators collectively represent the performance of the team at a specific time, If we want to reflect the change of the team's rhythm in the entire game, as long as we get the changes in the performance of the team in each time period, we use the sum of the three indicators to reflect the fluctuations of the team's status throughout the game. Then see the team's tempo change. We divided the game into 18 mini-games of five minutes in length, and studied the performance of the five-minute team by studying the indicators in the five minutes. The formula for calculating the tempo can be described as:

$$T_{Tn} = \sum_{t=0} Var(P_{Tn}) + Var(P_{Aij}) + Var(P_c)$$

Considering that the number of matches played by different teams may be different, we finally present the average of the T_{Tn} of each team and each game. Here are our results.

Team	Tempo
Huskies	0.529287
op14	0.86827
op18	1.22518
op6	0.71691
op9	1.20882
op17	0.74606
op7	0.63879
op11	0.69831
op16	1.29221
op13	0.95168

- Adaptability

  For a team, the ability to deal with problems in different situations reflects their adaptability. In our model, we mainly consider the impact of home and away changes on a team's record In our common sense, teams that play at home often get more encouragement and support, which will directly improve the game status of the home team. In order to measure the difference between the performance of each team at home and away, we have established two evaluation model, the first model does not consider the impact of home and away on the ten teams. The score of each team can be expressed as:

$$S_{Tn} = \sum_{i=1}^{x} S_{ci} * \delta$$

  where x is the number of matches, δ represents the winning factor, There are different values for the team's victory or defeat or tie. According to this model, we can get the ranking of the ten teams.

  In the second model we added the home and away factors and added a larger weight to them to highlight their impact on the team. The score of each team in the second model can be expressed as

$$S_{Tnha} = \sum_{i=1}^{x} S_{cih} * \delta * h + \sum_{i=1}^{y} S_{cia} * \delta * a$$

  where x y is the number of matches, h, a respectively the corresponding weight of home and away. In the new model we get a new score ranking of a ten teams. By comparing the differences in the rankings of each team in two different models, the influence of home and away factors on the team's performance can be reflected. The following is the difference between home and away performance of each team based on the two models.

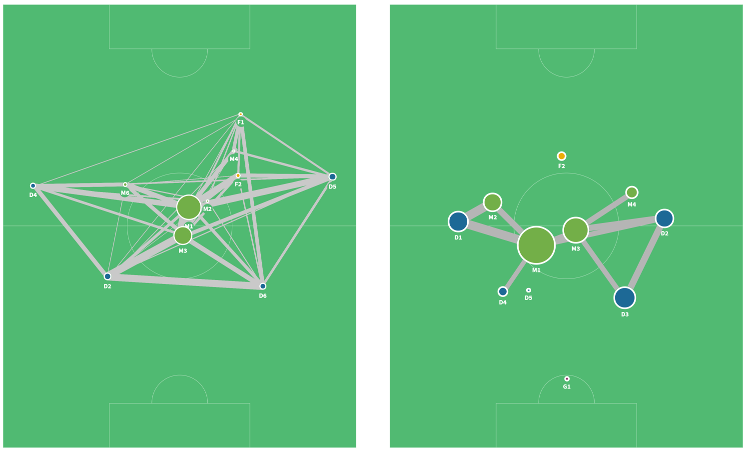
Team	rank(no h&a)	rank(h&a)	Change
*op*9	1	1	0
*op*7	2	3	-1
*op*13	3	2	1
*op*16	4	6	-2
*op*6	5	4	1
*op*11	6	5	1
Husikies	7	7	0
*op*17	8	9	-1
*op*18	9	10	-1
*op*14	10	8	2

From this we can get the extent to which the performance of these ten teams is affected by the home and away games, and then infer their adaptability.

- Strategic flexibility

  When analyzing a team's tactical strategy, we often care whether it uses a universally effective strategy or a specific one designed for a specific opponent. Obviously, targeted strategies do not have long-term significance in comparison, but whether they can change their inherent strategies based on changes in opponents can also show the flexibility of a team. In the context of a football game, we want to see two teams with similar strengths play two games and win each other. In this scenario, the analysis of the changes in the formation of the two teams can be more intuitively reflected. Due to lack of sufficient game data for teams other than Huskies, we only discuss the tactical nature of Huskies in this section, and this indicator will not be added to our final scoring model. But this can be used as another dimension to provide us with optimization suggestions that can be adopted.

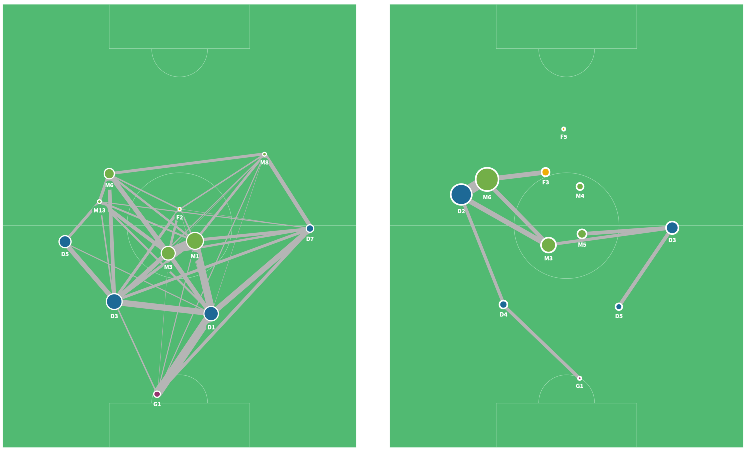
  Among these teams, we found the team with the closest record to the Huskies, opponent 6, We will present the two games played by both sides from the perspective of two teams, from which we can see the nature of the strategy used by the Huskies.



(a) Huskies. (b) op6.

Figure 9: Huskie : op6 2:1.

The following is a game where the Husky team beat opponent6. In order to make the picture clearer, we only selected the ten thickest lines to draw their teamwork when drawing the opponent's network diagram.



(a) Huskies. (b) op6.

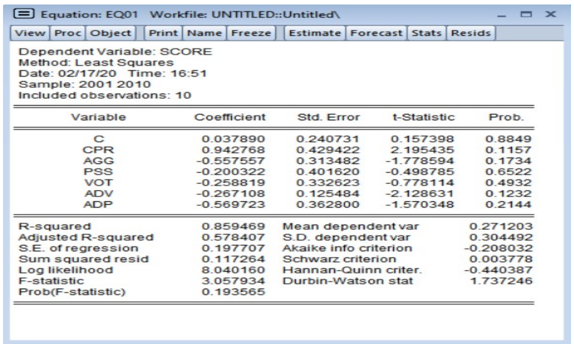
Figure 10: Huskie : op6 1:2.

In the game below, the Huskies lost to opponent6. Comparing the performance of both sides in the two games, we have the following inferences:

- From the perspective of opponent, we can see that the tactics of opponent6 in the two games have changed. Compared to the first game, they strengthened the cooperation between the defenders, weakened the connection between the midfield and other players, and allowed forward players greater participation in the game. From the position distribution of the players, the second game is obviously more aggressive than the first game. Except for the players with less participation to maintain their positions, other players are trying to impact the Huskies as much as possible. The most prominent of these is the forward, who has been spinning almost in front of the Huskies' goal.
- From the perspective of the Huskies, we can see that there is no big difference between the two games in terms of staffing and cooperation between the personnel, but in the losing game, the role of the defender was amplified and the midfielder Does not function as a hub as before. However, we also saw a big difference between the positions of the players in the two games. The players in the winning games were generally in the front and used aggressive play.
- In general we tend to think that the Huskies' strategy is relatively fixed. In other games we frequently find that they use a four-guard formation. The arrangement of the defenders is often a trapezoidal structure. In the face of the same opponent, the tactical deployment has not changed from the previous ones. For example, opponent6 here, it is obvious that the tactics of opponent6 have been adjusted and the focus of the attack has changed, but the Husky team has not made corresponding countermeasures. The players are just obsessed with running into their own positions and setting up a fixed formation, but they do not play the best role of this formation.

4.2.2 Multiple regression analysis

In order to find out the main factors attributing to team performance indicators in degree of cooperation, aggressiveness, possession, variance of tempo, diversity and adaptability, we conducted regression analysis.The regression results are as follows



Equation: EQ01 Workfile: UNTITLED::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

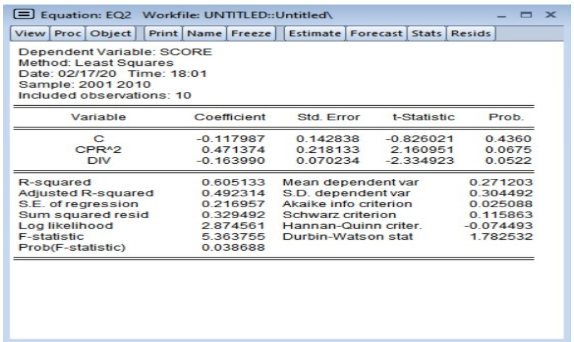
Dependent Variable: SCORE
Method: Least Squares
Date: 02/17/20 Time: 16:51
Sample: 2001 2010
Included observations: 10

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.037890	0.240731	0.157398	0.8849
CPR	0.942768	0.429422	2.195435	0.1157
AGG	-0.557557	0.313482	-1.778594	0.1734
PSS	-0.200322	0.401620	-0.498785	0.6522
VOT	-0.258819	0.332623	-0.778114	0.4932
ADV	-0.267108	0.125484	-2.128631	0.1232
ADP	-0.569723	0.362800	-1.570348	0.2144

R-squared	0.859469	Mean dependent var	0.271203
Adjusted R-squared	0.578407	S.D. dependent var	0.304492
S.E. of regression	0.197707	Akaike info criterion	-0.208032
Sum squared resid	0.117264	Schwarz criterion	0.003778
Log likelihood	8.040160	Hannan-Quinn criter.	-0.440387
F-statistic	3.057934	Durbin-Watson stat	1.737246
Prob(F-statistic)	0.193565		

Figure 11: Preliminary regression analysis

$R^2 = 85.95\%$, the goodness of fit is very high, but no variables are able to pass the t-test under 5% confidence, so there may be multiple collinearity, and we carry out stepwise regression, try to change its function form, and find out the main factors and the optimal regression equation. After multiple regression, the best fitting can be obtained by taking the degree of cooperation and diversity as explanatory variables. The results are as follows.



Equation: EQ2 Workfile: UNTITLED::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: SCORE
Method: Least Squares
Date: 02/17/20 Time: 18:01
Sample: 2001 2010
Included observations: 10

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.117987	0.142838	-0.826021	0.4360
CPR^2	0.471374	0.218133	2.160951	0.0675
DIV	-0.163990	0.070234	-2.334923	0.0522

R-squared	0.605133	Mean dependent var	0.271203
Adjusted R-squared	0.492314	S.D. dependent var	0.304492
S.E. of regression	0.216957	Akaike info criterion	0.025088
Sum squared resid	0.329492	Schwarz criterion	0.115863
Log likelihood	2.874561	Hannan-Quinn criter.	-0.074493
F-statistic	5.363755	Durbin-Watson stat	1.782532
Prob(F-statistic)	0.038688		

Figure 12: Stepwise regression analysis

So the regression equation is score = -0.117987 + 0.471374cpr^2-0.163990div, $R^2 = 60.51\%$. However, according to the practical experience, a team with multiple structure can deal with the complex game more easily, and the coefficient before the diversity variable is negative, which is not in line with the actual situation. As we think, the reason may be that the discussion on the function form is not enough, but due to the time limit, we will not make in-depth discussion, and we choose to turn to the research and judgment on the correlation coefficient matrix between variables Break. The following is the variable correlation coefficient matrix.

	CPR	AGG	PSS	VOT	DIV	ADP
CPR	1.000000	0.558017	0.493179	0.837267	0.011996	-0.130216
AGG	0.558017	1.000000	0.037592	0.468693	-0.411810	-0.121680
PSS	0.493179	0.037592	1.000000	0.430973	0.507452	0.144616
VOT	0.837267	0.468693	0.430973	1.000000	0.135261	-0.334524
DIV	0.011996	-0.411810	0.507452	0.135261	1.000000	-0.393755
ADP	-0.130216	-0.121680	0.144616	-0.334524	-0.393755	1.000000

Figure 13: Variable correlation coefficient matrix

Because in the first round of stepwise regression, the simplest single variable selected is score = f (CPR), we can take CPR as standard, and score and CPR are positive correlation, while in the matrix of variable correlation coefficient, we can notice that CPR and DIV are positive correlation, so we can infer that DIV and score are positive correlation, too. So the final regression equation is score = -0.117987 + 0.471374cpr^2 + 0.163990div

5 Suggestions and inspiration

From the passing network model and the team performance indicator model, we get the dynamics and attributes of the Huskies team. We also got the factors that a good team needs to have and the weight of these factors in team performance. In this part, we will further analyze the Husky team by combining the passing network model and the team performance indicator model, and find the current feasible strategy and the direction of the next optimization.

5.1 Effective structural strategies

Based on the team performance indicator model we have obtained, we can calculate the score of ten teams under the calculation formula provided by the model, which is presented in the table below.

Team	Score
Huskies	0.826165114
op14	0.819392075
op18	0.776906931
op6	0.709852671
op9	0.850802724
op17	0.628098337
op7	0.673978716
op11	0.414944984
op16	1.187505142
op13	0.422673634

According to the model we have established, the Huskies team ranks very high in this team. This may because we have determined only two impact indicators after linear regression, namely the degree of coordination and diversity.

The cooperation between the defender and the midfield is a highlight of the Husky team. As mentioned in the previous analysis, the Husky team has a clear binary and ternary structure, which means that it has several groups of players who cooperate very well. The transmission of the ball between the midfielders is used to direct the offense and defense. Husky's current four-guard lineup is like a net to prevent the opponent from deepening.

5.2 Players' star analysis

In the previous team diversity analysis, we rated the stars for each player of each team. Combined with the team's tactical design and offensive, defensive choices, the requirements for the ability of players in each positioning are also different. Below is the star rating of each position member of the Huskies team.

Of the most played in each game, we found that the number of forwards was the lowest. From this we can infer that the Huskies have insufficient forward reserves, at thesame time, from the star map of the players, it can be seen that the skills of the Huskies forwards are generally modest, means that tthey are generally lack advanced football skills, and the team also lacks top passers in midfield and forward. Since the operation of the Huskies ball depends mainly on the midfielder and the defender, once the team's offense and defense are organized by players with relatively weak passing ability, the operation of the entire system will appear stiff.

	dual*	pass*	sp*
H D1	3	2	1
H D10	3	2	1
H D2	2	2	1
H D3	2	2	2
H D4	1	2	1
H D5	2	1	1
H D6	3	1	2
H D7	2	1	2
H D8	2	1	2
H D9	2	3	1

(a) defender

	dual	pass	skill
H F1	3	1	1
H F2	1	3	1
H F3	1	3	1
H F4	3	1	1
H F5	2	2	1
H F6	1	3	2

(b) forward

	dual	pass	skill
H M1	2	2	1
H M10	3	1	1
H M11	2	2	1
H M12	3	1	2
H M13	2	2	2
H M2	2	2	2
H M3	2	2	2
H M4	2	1	1
H M5	2	2	2
H M6	2	1	2
H M7	2	2	1
H M8	3	1	2
H M9	3	1	2

Figure 14: midfield

So we can also see in the network diagram that those winning games are either the midfielder and the backcourt players have performed very well, or the forward alone plays the role of the offensive hub. This is instructive for the Huskies' signing up for the new season.

5.3 Network analysis

- Forward's tactical position has long been underestimated. In our network diagram, we can see almost all matches, the importance of the forwards (the size of the points) is always low, and sometimes there is little connection with other people (the thickness of the line). Some of their average position is not even deeper than some midfielders.
- With a strong guard lineup, four guards can primarily protect the backcourt. So that although the team's possession rate is not high, they can still rely on defense to win the game. However, an overly fixed lineup may lead to a single tactic. We checked the game data of 38 games. The network diagram shows that the relative positions and passing conditions of the four defenders are always relatively fixed regardless of the final record. This will also make it easier for the team to be targeted.
- The midfield plays a vital role in the game and often determines the direction of a game. According to the results of the network diagram, if the midfield plays a critical (larger node) role and has a high degree of participation (thicker lines), the Huskies can win most of their opponents. It is also essential to keep their valuable midfielder resources.

5.4 New season suggestions

Based on the above analysis, we have made the following suggestions for the preparation of the Huskies next season,

- Introduce some high-level forwards. Forward is now the weakness of the Huskies. Not only does it not have enough reserves, but the overall quality of the forward is not high enough.
- Provide more tactical options for defender and midfield. The midfield and backcourt are the relative advantages of the Huskies. Still, too simplistic tactics may also make players tired of finding their position and ignore changes in the situation.
- Prepare for more countermeasures against opponents. In the previous analysis, we can see that the Husky team's tactics tend to be single and cannot always be adjusted in time as the opponent's tactical layout changes.

6 Build an excellent team

From our previous analysis of the Huskies, we can generalize the general way of designing effective teams. The main process can be represented by the following flowchart. In this exprssion:

- δ is the probability of producing AIDS babies
- P_p is the proportion of the total population who are pregnant
- P_a indicates the rate of infection among pregnant women

To calculate the increase in the number of AIDS babies caused by mother-to-child transmission, we first obtain the Qy of the proportion of pregnant women in the population. Pregnant women are divided into pregnant women and pregnant women. The babies born by HIV-negative pregnant women are naturally also HIV-negative, while HIV-positive pregnant women are likely to give birth to HIV-positive and HIV-negative babies. Therefore, we checked the data to find that the proportion of HIV-positive babies in the two infants was δ. Therefore, the number of new AIDS babies generated by mother-to-child transmission in the population is the total number the proportion of pregnant women t the proportion of AIDS patients in pregnant women a the probability of producing AIDS babies δ. Considering that our inspection period is one year, multiple births or multiple births in one year are not considered.

the infected member reduced because death

We count deaths between susceptible populations and patients separately. Because of the short life span of patients after the onset of AIDS, generally in two or three years, we can consider AIDS and its induced diseases as the primary cause of patient death compared to natural death factors, and ignore the natural death factors of AIDS patients. The average survival time of AIDS patients after infection is x years, so when the patient's illness time reaches x years, we judge it as death and clear all its data. The deaths of susceptible people are determined by the average life expectancy and natural mortality. Every year, among the susceptible people, the death tolls of all ages are determined according to the proportion of deaths of all ages in South Africa. The sum of the two is the total number of deaths.

population growth

Our population growth model is estimated by a logistic retarded growth model,which can be expressed as

$$\frac{dP}{dt} = r*(a-P/K)*P$$

Among them, P is the total number of people in the area at time t, r is the maximum growth rate of the population in the area, and K is the upper limit of the population load in the area. The general solution is $P(t) = r/[ce^{-rt} + r/K]$, and c is a constant determined by the initial value. Divide P to the left of the equal sign, which is

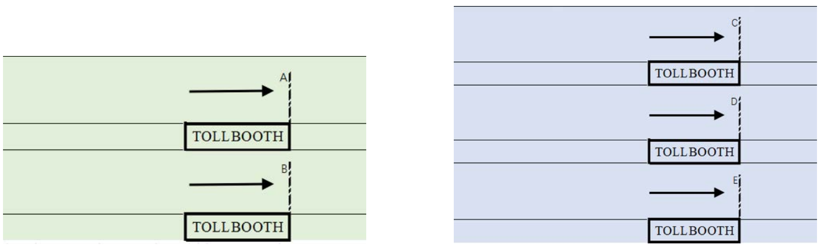
$$\frac{1}{p} * \frac{dp}{dt} = a + bp$$

where a = r, b = -r / K. We can use the method of least squares estimation to substitute South Africa's data from 2000 to 2018. Regression calculation shows that r = 0.263727 and K = 112703846, that is, the maximum annual growth rate of South Africa is 26.3727%, and the upper limit of population carrying is 112703846. At this time, c = 4.144387586816374*10^{220} can be calculated, and the annual number of people can be calculated by$P(t) = r/[ce^{-rt} + r/K]$.The process of regression analysis can be found in the appendix.

6.1 Control Time Model (CTM)

Given aforementioned deficiencies of existing merging pattern, we propose a new model, partially based on the current one. Instead of having all the vehicles moving and merging at their own discretion, control time model will control the departure time of vehicles to ensure a smooth and safe emerging process.

Specifically, for situations where two booths merge into one lane, the second vehicle will only be allowed to proceed t0 seconds after the first vehicle moves forward. The time t0 is defined as the control time. Similarly, for situations where three booths merge into one lane, the third vehicle will be allowed to proceed t0 seconds after the second vehicle moves forward, whilst the second vehicle t0 seconds after the first vehicle, as shown in Figure 1



(a) Either A or B is open. (b) Either C, D or E is open.

Figure 15

In this way, the regulated merging of the vehicles into another lane would be more efficient than the situation where vehicles are proceeding without regulation, for drivers should take time to make decision when multiple booths merge into one lane simultaneously, let alone the risk for doing so.

We model that vehicles start with constant acceleration as until reach the maximum speed vm in the straight path. They then immediately starts merging into their prescribed road in two consecutive tangent circle arcs.

Now we evaluate the "appropriate" control time. When emergency happens, one vehicle take severe brake action with acceleration ab, after response time tres, the posterior vehicle take severe brake action with the same acceleration. Consider the distance of the two vehicles (t = 0 is the time when emergencies happen):

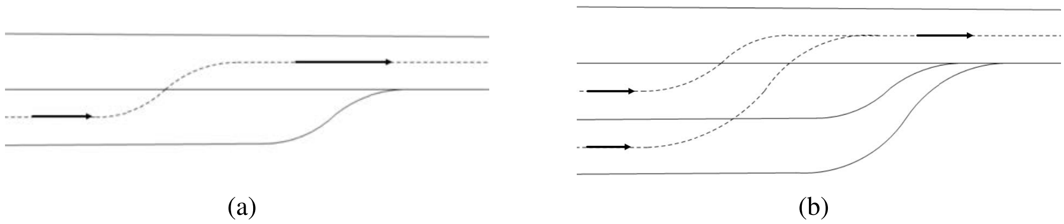


Figure 16: Two or three tollbooth egress lanes merge into one lane

$$d = v_m t_{rea} + \int_0^t v_{posterior}(t)dt - \int_0^t v_{previous}(t)dt$$

where

$$v_{previous} = \begin{cases} v_m, & t \leq t_{rea} \\ v_m - a_b(t - t_{rea}), & t_{rea} < t \leq t_{rea} + \frac{v_m}{a_b} \\ 0, & t_{rea} + \frac{v_m}{a_b} \leq t \end{cases} \tag{1}$$

Solve the equation $d \geq 0$, we get $t_0 \geq tres$. By saying "appropriate" control time, we take t_0= $1s \geq t_{res} = 0.8s$ (Lee et al, 2002).

After describing the control time model, we first calculate the throughput of the toll plaza.

Let i_k be the number of booths corresponding with the k^{th} lane. For example, if the first three booths merge into the first lane, we say $i_1 = 3$.

By computer simulation (details are included in Appendix), we have t_{ik} close to $\frac{T}{i_k}$ when $i_k \ll \frac{T}{t_0}$ (especially when the variance of service time is small), and t_{ik} increases as i_k increases, and the increase speed is quite large when $i_k \geq \frac{T}{t_0}$. where tik is the averaged timefor a car to come into k_{th} lane.

Therefore, to maximize the efficiency, we'll keep i_k as small as possible, whereas the throughput comes suboptimal if $i_k \geq \frac{T}{t_0}$ for some k. The optimal throughput is closed to $\frac{B}{T}$ car(s) per second.

By substituting B by 8, the expectation of human service time T by 15. According to our optimization method above, when we keep the i_k as small as possible, which is in other words, let them distributed evenly(e.g. 2,3,3). We will get the optimal throughput closed to $\frac{B}{T} \approx 0.533$ car per second.

Then we calculate the size of the plaza, by doing so we discuss the value of D.

Define f : $\{1, 2, \ldots, \text{B}\} \rightarrow \{1, 2, \ldots, \text{L}\}$ which is non-decreasing, and $f(i) = j$ if and only if vehicles from i_{th} booth merge into j_{th} lane. It obviously follows that $i_k = \| f^{-1}(\{\text{k}\})\|$. We call the mapping f a "merging pattern".

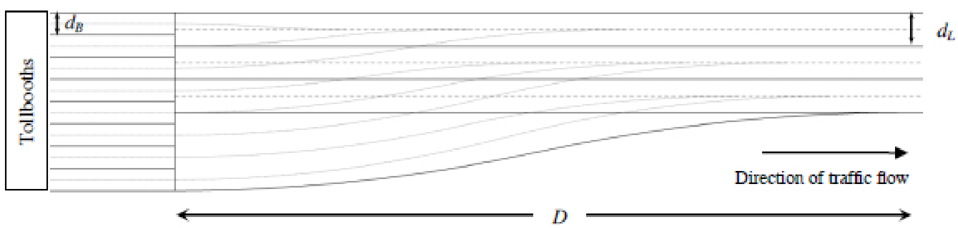


Figure 17: B = 8 tollbooth egress lanes merge into L = 3 lanes

By geometric relations, we have

$$(\frac{D-\frac{v_m^2}{2a}}{2})^2+(r-\frac{(i-\frac{1}{2})d_B-(j-\frac{1}{2})d_L}{2})^2=r^2$$

Observed from above relation, a certain (i, j) prescribes a lower bound of D, and for a certain merging pattern, the distance between the tollbooth and the end of the plaza that we eventually employ is the greatest lower bound of D. The less the greatest lower bound is, the smaller the size will be, and so will the cost of plaza. By property of quadratic function, for a certain j, the greater i is, the greater D is. We have

$$\sum_{k=1}^{j} i_k d_B - j d_L \geq (\sum_{k=1}^{j} i_k+1)d_B-(k+1)d_L$$

since $d_L \geq d_B$. That implies if j is changed into j + 1 whilst i is changed into i + 1, D will get smaller. Therefore, the global maximum must be obtained at one of the local maximum points, and then we let x be

$$x = \max_{j\in\{1,2,\ldots,L\}}\left\{\frac{(\sum_{k=1}^{j} i_k-\frac{1}{2})d_B-(j-\frac{1}{2})d_L}{2}\right\}$$

And for given B and L, the optimal choice of merging pattern f would be the case if we can minimize x to be

$$x_{min} = \min_{\{i_k\}\in\{\|f^{-1}(\{k\})\|:\forall f\}} x$$

After trying out some pairs of small (B,L), we induce that for any given small number B, L (which can be put in practical use) satisfying $B \geq L$, there exists a merging pattern f_m such that

$$x = \frac{(B-\frac{1}{2})d_B-(L-\frac{1}{2})d_l}{2} := x_0$$

Since for any ik we have the relation $f(B) = L$, which implies above expression will appear in every calculation of x for all i_k, hence $x_{min} \leq x_0$, hence the existence of x_0 guarantees ”=”, providing us with the optimal merging pattern to minimize D.

By substituting B = 8, L = 3, d_B = 2.5m, d_L = 3.75m, r = 115m, v_m = 15m/s, a =2.78m/s^2, the minimum of D is 105.5m, when (i1, i2, i3) $\in$ {(3, 3, 2), (3, 2, 3), (2, 3, 3)}.

6.2 Waiting Area Model(WAM)

After completing the above model, something seems to be deficient also appears in our model. In the control time model, vehicles may need to wait at the booths even after the service finished, while there are a lot of vehicles queuing after it. So we wonder that whether it is better to build a waiting area at a plaza (as shown in Figure 5), so that the vehicles whose service has finished can go to the waiting area, then the consecutive vehicle can be serviced. So in this model, the staff spending more time serving instead of doing nothing while the vehicles are just waiting at the booths

We will test and compare this model with control time model in the following sections,to see whether it is efficient or not.

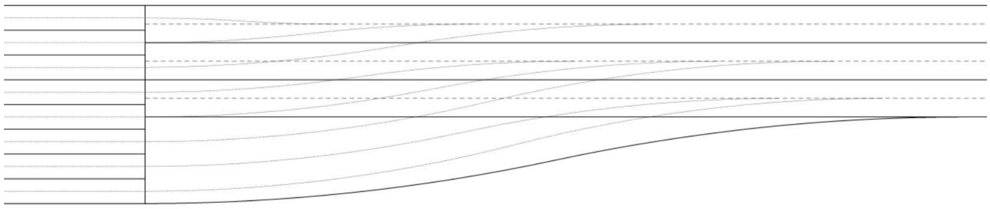


Figure 18: B = 8 tollbooth egress lanes merge into L = 3 lanes

7 Analysis and Results

7.1 Throughput

Throughput (θ) here is defined to be the number of vehicles per hour passing the point where the end of the plaza joins one of the the L outgoing traffic lanes. Here we discuss only one outgoing lane merged from two lanes in human-staffed tollbooth area in heavy traffic where vehicles come continuously during heavy traffic so that the cashier work continually, as shown in Figure 6.

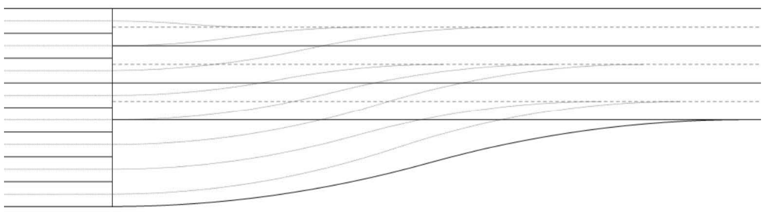


Figure 19: 8 tollbooth egress lanes merge intoL = 3 lanes in heavy traffic

7.1.1 Throughput of existing model

The toll collection service time is a random variable which can be seen as a sum of random variables including time needed by driver to search money or credit vehicled, time needed by cashier to take related record, etc. According to the central limit theorem, we can assume that the service time of ith vehicle ($t_{h,i}$) of the each lane follow normal distribution with $\mu = 15$ and $\sigma^2 = 4$ independently and identically.

8 Strengths and Weaknesses

8.1 Strengths

- First one...
- Second one ...

8.2 Weaknesses

- Only one ...

Memorandum

To: Heishan Yan
From: Team XXXXXXX
Date: October 1st, 2019
Subject: A better choice than MS Word: LaTeX

In the memo, we want to introduce you an alternate typesetting program to the prevailing MS Word: **LaTeX**. In fact, the history of LaTeX is even longer than that of MS Word. In 1970s, the famous computer scientist Donald Knuth first came out with a typesetting program, which named TeX ...

Firstly, ...

Secondly, ...

Lastly, ...

According to all those mentioned above, it is really worth to have a try on LaTeX!

References

[1] Abdel-Aty, M. A., & Radwan, A. E. (2000). *Modeling traffi c accident occurrence andinvolvement*. Accident Analysis & Prevention, 32(5), 633-642.

[2] Corwin, I., Ganatra, S., & Rozenblyum, N. (2005). *A Single-vehicle Interaction Model of Traffic for a Highway Toll Plaza*. The UMAP Journal, 26(223), 299-315.

[3] Friedman, D. A., & Waldfogel, J. (1995). *The administrative and compliance cost of manual highway toll collection: evidence from Massachusetts and New Jersey*. National Tax Journal, 217-228.

Appendix A: Further on LaTeX

To clarify the importance of using LaTeX in MCM or ICM, several points need to be covered, which are ...

To be more specific, ...

All in all, ...

Anyway, nobody **really** needs such appendix ...