Team Coordination: Quantify teamwork in restricted areas in hockey.

Baptiste Gorteau

January 27, 2025

Big Data Cup, team coordination focus.

Project GitHub Repository

Abstract

Growing up as a soccer enthusiast, I have always been fascinated by the ballet of passes proposed by Marcelo Bielsa and Pep Guardiola's teams, to mention only them. When I became interested in hockey, I immediately found similarities in the passing game between the two sports. The 2025 Big Data Cup [2] was for me the perfect moment to explore hockey team's teamwork and more specifically in restricted areas.

1 Introduction and approach summary

In hockey and sport in general, when we think of teamwork, two terms that come up regularly are coordination and vision. From an analytical point of view, these two terms may seem complicated to emphasize. In this work, we are going to try to make these terms analyzable in a simple and comprehensible way. The analytical route I decided to take in this project is the creation of a metric. The comprehension goal that I set for myself pushed me to choose this path rather than that of machine learning, with which I was hesitant.

For this study, we use the dataset provided by Stathletes as part of the 2025 Big Data Cup [2]. To compute our metric, we use in priority the tracking data from the dataset combined to Voronoi diagram. The Voronoi diagram is a mathematical structure that divides a space into multiple regions, where each region is associated with a specific point (called a site or generator). Every point within a region is closer to its corresponding site than to any other site. This method is used in fields like biology, mapping, video games and sport [3]. Multiple works have been made to showcase the use of Voronoi diagram in sport. We can for example cite an article from 2012 by Sofia Fonseca, João Milho, Bruno Travassos and Duarte Araújo [1]. The tracking data presented comes from match video recordings and don't show all the players on the rink at each frame. Because of this restriction, we decide to focus our analysis on teamwork in restricted areas. We define the restricted area as the area around the puck at every moment of the play. We will define more precisely this area in the first part of the study.

In a first section, we will explain how we create our base metric from Voronoi diagram. In a second section, we will first show how our metric can help us to have a better view of teamwork in restricted areas. Then, we will present how can we decline our base metric in two new ones to analyze teamwork. The first metric will focus on how well the teammates place themselves around the puck carrier, and the second one on how well the puck carrier finds his open teammates.

2 Creation of the metric

In this section, we present how we use the Voronoi diagram to create a metric allowing us to compute teamwork. We will call this metric the Close Availability Score (CAS).

2.1 Adapt Voronoi diagrams to our situation

We start by applying the Voronoi diagrams to the X and Y points of the player's tracking data. After applying the method, we see in figure 1 that lots of regions are open and not usable for further analysis. To tackle this problem, we add four new points just outside the ends of the rink. These points allow us to close most open regions, as you can see in figure 2. The final step consist in adapting the Voronoi diagrams to our restricted area. We decide to focus on an area of 60 by 60 feet, where the puck is the middle of it. If the puck is close to a border, the size of the area will be reduced. We cut the former regions by using this square area to obtain smaller regions which form a square (figure 3).

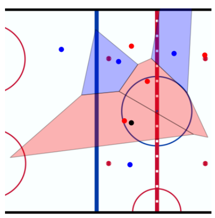

Figure 1: First application of Voronoi diagram.

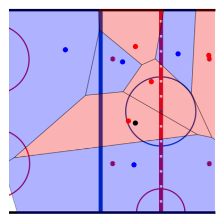

Figure 2: Voronoi diagram after adding points at the ends of the rink.

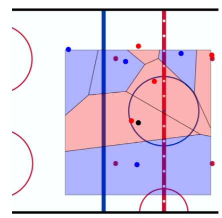

Figure 3: Adapt the Voronoi diagram to our restricted area.

2.2 Create a score from the adapted diagram data

With this new diagram data, we can compute the 'Close Availability Score' (CAS). This score will only be computed for the teammates of the puck carrier if their regions share a border with his region. The teammates need to be inside the restricted area to be considered as available. To create the CAS, we use two elements provided by our diagram, the surface S_{R_i} of teammate's i region R_i and the border $b_{i,pc}$ between R_i and the region of the puck carrier R_{pc}. We compute a score for each of these two elements. These scores will then be used to compute the CAS. We call these scores the 'Surface Score' Ss and the 'Border Score' B.

$$Ss = \frac{S_{R_i}}{180} \qquad\qquad B = \frac{b_{i,pc}}{6}$$

To create Ss, we divide region's surface S_{R_i} by 180. 180 corresponds to half of the restricted surface which is 1800 square feet (60×30) that we divided by 10 so Ss is mainly between 0 and 100. For B, we divide the border $b_{i,pc}$ by 6. 6 corresponds to the length of the restricted area's borders (60 feet) divided by 10 to keep B mainly between 0 and 10. Now that we have Ss and B, we can compute the 'Close Availability Score' CAS.

$$CAS = Ss \times B$$

We decided to compute the score that way because we wanted to give an importance to the border $b_{i,pc}$ between the puck carrier his available teammate. If the region of the available teammate i have an important surface S_{R_i} but the border $b_{i,pc}$ is narrow, it will still be "dangerous" for the puck carrier to make a pass to his teammate i. In this situation, the score will not be as big as if the border $b_{i,pc}$ had been larger. When CAS is computed, we can use and analyze it from different angles.

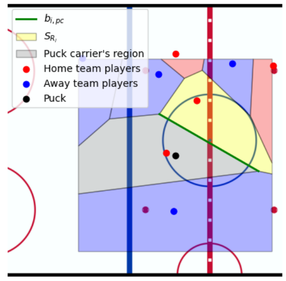

Figure 4: Visualization of $b_{i,pc}$ and S_{R_i}.

3 Different uses of the Close Availability Score CAS

Now, we are going to look at the different ways we can use our metric to visualize and analyze teamwork in restricted areas. In our analysis, we focus on plays. A play start at the moment where a player receives the puck and ends when he passes it to another player of his team.

3.1 Visualize the CAS during a play

The first use we can make of CAS to be more familiar with it is to represent it during a play. In order to do that, we represent an action with player's regions and the evolution of the score of the available teammates. The color of the score evolution on the right graph corresponds to the color of the player's region border on the left graph. In this example, we display four frames at different moments of a play. The full video sequence is available for download at this link.

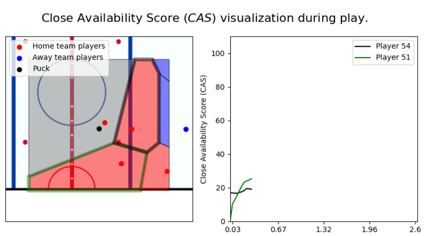

Figure 5: Begining of play

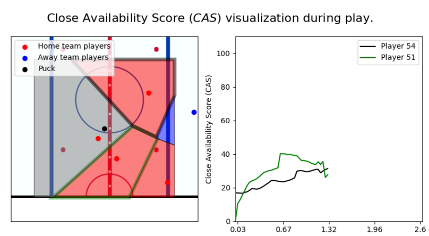

Figure 6: Midle of play

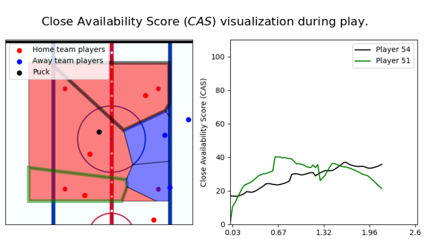

Figure 7: Start of the pass

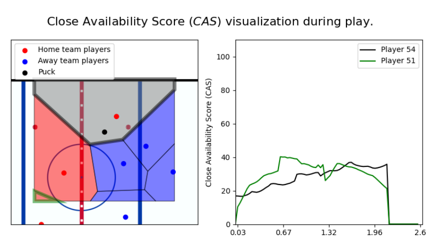

Figure 8: Pass reception

This visualization gives us a more detailed view of our score. Indeed, we can notice how quickly it can change in a very short period of time. This observation highlights the importance of the placement of each teammate but also the importance of the puck carrier's decision-making. Based on these findings, we will carry out two analyzes based on the CAS. The first will focus on the positioning of available teammates, and the second on the puck carrier's decision-making.

3.2 Which teammates are most available, and how are they found?

The first analytical usage we can make of the CAS is to compute how available each of the players makes themselves in restricted areas during the game. In order to do that, we create a new metric, the mean CAS $\overline{CAS}$. This variable represents the mean of all the CAS values over zero for a player during a play.

$$\overline{CAS} = \frac{\sum_{CAS_i>0} CAS_i}{|\{CAS_i \mid CAS_i > 0\}|}$$

Where $|\{CAS_i \mid CAS_i > 0\}|$ represents the number of CAS values over zero and $\sum_{CAS_i>0} CAS_i$ the sum of these values.

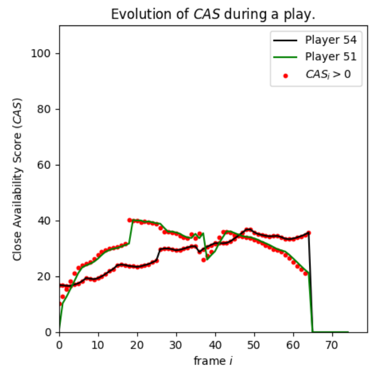

Figure 9: Example of CAS values for $\overline{CAS}$ metric.

We can now look at the average $\overline{CAS}$ of a player during a game. To get insightful information, we will compare the average $\overline{CAS}$ with the percentage of plays where the player was available, and found by the puck carrier. We also add the average uptime of the player on a play to the visualization. The uptime of a player corresponds to the time during a player is available on a play. We apply that to the players of the three games included in the data. The sample size we took for each game is 333 plays for game 1, 340 for game 2 and 391 for game 3. We only consider the players available in more than 10 plays.

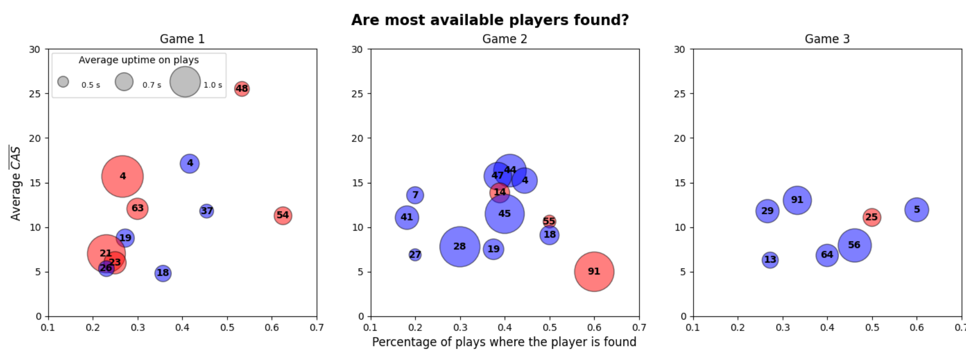

Figure 10: Are most available players found?

These three graphs offers us a significant number of interpretations. One of the interpretation we can make is that the players with the highest uptime are not always those with the highest average $\overline{CAS}$ and vice versa. For example, we can talk about player 91 during game 2. This one has the lowest average $\overline{CAS}$ of the selected players, but is also the one best found when available. The first interpretation that we could make of this case is a high capacity to master less good spaces. The second interpretation is to think that there is an important connection between this player and his teammates. The second player that catches our attention is player 48 from game 1. This player has the highest average $\overline{CAS}$ of the three games. This player also has one of the lowest average uptimes observed. We could interpret this as an ability to position himself very well and be quickly found. This hypothesis can be supported by his percentage of times where he is found by his teammates, which is very good.

3.3 Which player finds the best his available teamates?

Now, we use our metric to see how well the puck carriers finds his available teammates. The goal here is to see if at the moment where the puck carrier made the pass, the CAS of the founded teammate was one of the best of the play. To discover that, we once again create a metric from the CAS. This metric will be called 'Find top CAS' ($FTCAS$). For this metric, we base ourselves on the founded player's CAS at the launch of the pass. We name this score CAS_ℓ. Then, we look at how many CAS values are over zero and below CAS_ℓ.

$$FTCAS = \frac{|\{CAS_i \mid CAS_\ell > CAS_i > 0\}|}{|\{CAS_i \mid CAS_i > 0\}|} \times 100$$

Where $|\{CAS_i \mid CAS_\ell > CAS_i > 0\}|$ represents the number of CAS values over zero and bellow CAS_ℓ and $|\{CAS_i \mid CAS_i > 0\}|$ the number of CAS values over zero.

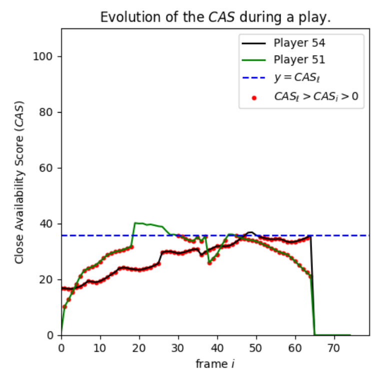

Figure 11: Example of CAS values for $FTCAS$ metric.

Now that we know how this metric works, we can look at which are the players in each game that find the best their available teammates. We look at the players with more than three plays where they made a pass to a player from the restricted area.

Figure 12: Game 1 $FTCAS$ ranking

Player	Team	average FTCAS
4	Home	91.62
48	Home	72.05
4	Away	72.04
37	Away	63.56
23	Home	59.16

Figure 13: Game 2 $FTCAS$ ranking

Player	Team	average FTCAS
41	Away	89.85
38	Home	88.90
55	Home	71.11
47	Away	70.95
44	Away	68.07

Figure 14: Game 3 $FTCAS$ ranking

Player	Team	average FTCAS
5	Away	76.65
29	Away	73.14
29	Home	70.57
78	Away	66.90

By observing the ranking of the players by their $FTCAS$ in the first game, we notice that player 4 of the home team demonstrates incredible vision of the game. His $FTCAS$ of 91.62 testifies to an ability to make the pass at almost the best moment of the play to one of his teammates present in the restricted area. We can note that player number 48 of the same team demonstrates very good game vision with a $FTCAS$ of 72.05. Note that we had already noticed this player in the previous section for his very good ability to make himself available.

In the rankings of the second game, we observe a player from each team who has a very high $FTCAS$. The three other players present in this ranking also come from both teams and have very good $FTCAS$ around 70. From these scores and the players involved, we could think that the match was able to present a good quality of play and was tight.

Finally, Game 3 didn't see many players complete more than three passes to their teammates in the restricted area. The players concerned are mainly from the away team, and their $FTCAS$ are very good without being excellent. The fact that the away team won this match 2 to 1 may lead us to look at a possible correlation between the players' vision of the game in restricted areas and the result of a game.

4 Conclusion and openings

In conclusion, we can say that this metric allowed us to fly over different important bridges of team play in restricted areas. It also allowed us to see to how Voronoi diagrams can be the source of relevant analyses in hockey. Our base metric that we have transformed has helped us to highlight a combination of exceptional coordination and vision in certain players. We can think of player number 48 from the home team in game 1. This one has the best average $\overline{CAS}$ of the three games, while having a very good $FTCAS$. The analyses that we carried out also allowed us to ask questions about possible more in-depth analyzes based on our metric. We can think, for example, of looking at whether there is a correlation between the results of a match and the coordination of the players of a team. We could also pay more attention to areas of the puck where team coordination is measured. Finally, we can also wonder what a study on a larger number of matches would give. For example, we could analyze college games and come up with underrated players that could be drafted.

References

[1] Bruno Travassos Sofia Fonseca João Milho and Duarte Araújo. "Spatial dynamics of team sports exposed by Voronoi diagrams". In: *Human Movement Science* (2012).

[2] Stathletes. *BIG DATA CUP*. 2025. URL: https://stathletes.com/big-data-cup.

[3] Wikipedia. *Voronoi diagram*. 2025. URL: https://en.wikipedia.org/wiki/Voronoi_diagram.