Modeling Optimal Defensive Coverage in Football Corner Kicks Using Bipartite Graph Matching and Hall's Theorem on FIFA World Cup 2022 Data

Danesh Rasyad Damotino – 13525079
Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jalan Ganesha 10 Bandung
E-mail: drdamotino@gmail.com , 13525079@std.stei.itb.ac.id

***Abstract*—Set pieces account for a significant share of goals in elite football, yet the defensive side of a corner kick is usually discussed qualitatively rather than analyzed formally. This paper models the assignment of defenders to attackers on a corner kick as a matching problem on a weighted bipartite graph, with attacking and defending players as the two vertex sets and edge weights expressing marking quality as a function of distance. Hall's Marriage Theorem decides whether every attacker can be covered, while the Hungarian algorithm computes the optimal marking. Using StatsBomb open data from the 2022 FIFA World Cup, 300 corner kicks are analyzed. At a realistic cover radius of 3 metres, 72.3% of corners produce a Hall violation, a configuration in which at least one attacker is provably impossible to mark. The result shows that defenses are generally unable to cover every attacker on a contested corner, although this positional uncoverability does not translate directly into the number of shots or goals a team produces.**

***Keywords*—bipartite graph, matching, Hall's theorem, Hungarian algorithm, corner kick**

I. Introduction

If you watched the 2022 FIFA World Cup, you know that set pieces heavily influenced the tournament's outcome. Corner kicks are some of the most dangerous dead-ball situations in the sport. Whenever a team wins a corner, the defense has to quickly solve a complex problem: how do we assign our players to neutralize their aerial threats in the box? Analysts and coaches usually debate this using concepts like man-to-man or zonal marking, but it is rarely broken down mathematically.

From a mathematical perspective, assigning defenders to attackers is a classic matching problem on a bipartite graph, a core concept in Discrete Mathematics. When we ask, "Can the defense cover every single attacker?" we are really asking, "Does a matching exist that saturates the attackers?". Hall's Marriage Theorem gives us the exact answer to that question. Furthermore, if we want to find the best possible defensive assignments based on player positioning, we can use the Hungarian algorithm.

This paper applies these discrete math concepts to actual corner kick data from the 2022 FIFA World Cup, utilizing StatsBomb's open dataset to map exactly where players were standing the moment the ball was kicked. The goal here is twofold: first, to calculate exactly how often a defense is mathematically beaten before the ball even arrives; and second, to see if teams that successfully create these "unmarkable" attackers actually score more goals from corners.

II. Theoritical Foundations

A. Bipartite Graphs

A graph $G = (V, E)$ is made up of a set of vertices V and edges E that connect them. We call a graph bipartite if its vertices can be split into two separate groups, X and Y, where edges only ever connect a vertex in X to a vertex in Y. There are no connections within the same group. Mathematically, $G = (X \cup Y, E)$ is bipartite if $X \cap Y = \emptyset$, and every edge $(u, v) \in E$ means $u \in X$ and $v \in Y$.

These graphs are perfect for modeling situations where two different groups need to be paired up. In our case, attacking players need to be marked by defenders. To make the model more realistic, we use a weighted bipartite graph. Every edge (u, v) gets a specific number or weight $w(u, v)$ that represents the strength of that connection. For this paper, the weight represents how well a defender is positioned to contest an attacker.

B. Matching

A matching $M \subseteq E$ is just a set of edges where no two edges share the same vertex. In our bipartite graph $G = (X \cup Y, E)$, a matching pairs attackers in X with defenders in Y so that nobody is double-booked. If a vertex v is part of this matching, we say it is saturated. If every single vertex in X is saturated, we have a complete matching. A maximum matching contains the highest possible number of edges, and in a weighted scenario, a maximum-weight matching gives us the highest possible total weight, represented as $\sum_{(u,v)\in M} w(u, v)$.

In football terms, X represents the attackers in the box, and Y represents the defenders. A complete matching means every single attacker has a dedicated defender marking them. The main question we are trying to answer is whether the players' positions allow for this complete matching to happen.

C. Hall's Marriage Theorem

To figure out if a complete matching is even possible, we use Hall's Marriage Theorem (P. Hall, 1935). If we look at a specific group of attackers $S \subseteq X$, their neighborhood $N(S)$ is the group of all defenders in Y that are close enough to mark at least one of them:

$$N(S) = \{\, y \in Y : (x, y) \in E \text{ for some } x \in S \,\}$$

Theorem (Hall, 1935). A bipartite graph $G = (X \cup Y, E)$ has a complete matching saturating X if and only if $|N(S)| \geq |S|$ for every subset $S \subseteq X$.

Simply put, Hall's condition means that any given group of attackers must have at least an equal number of defenders close enough to mark them.

If we find a group of attackers S where $|N(S)| < |S|$, they have violated Hall's condition. This means there simply aren't enough defenders nearby to cover them all, leaving at least one attacker completely free. The smallest group that causes this is known as a Hall violator. Finding these violators is the core metric of this study. König's Theorem also ties into this by showing that the size of a maximum matching is equal to the size of a minimum vertex cover, giving us another way to verify how many attackers are actually covered.

D. Hungarian Algorithm

While Hall's theorem tells us if a perfect defense is possible, it doesn't tell us the best way to set it up. To solve this weighted assignment problem optimally, we use the Hungarian algorithm (Kuhn, 1955). By feeding it a weight matrix W, where W_{ij} represents how well defender i covers attacker j, the algorithm calculates the matchups that maximize $\sum_i W_{i,\sigma(i)}$. It runs in $O(n^3)$ time, which is fast for the small number of players inside a penalty box. If there are more attackers than defenders (or vice versa), we pad the matrix with zeroes to keep it square and just ignore the fake pairings later. The total weight of this optimal matching gives us a measurable score of the defense's overall coverage.

III. Problem Modeling

A. Representing a Corner Kick as a Bipartite Graph

The corner kick is modeled as a weighted bipartite graph $G = (A \cup D, E, w)$. Here, $A = \{a_1, \dots, a_m\}$ is the group of attackers inside the box, and $D = \{d_1, \dots, d_n\}$ is the group of defenders. We pull each player's exact (x, y) pitch coordinate from a StatsBomb 360 freeze frame taken at the exact moment the ball is kicked. Because edges only connect attackers to defenders, the graph is naturally bipartite. An edge (d_i, a_j) only exists if defender d_i is physically close enough to challenge attacker a_j for the ball, which we define as:

$$\|d_i - a_j\| \leq r$$

B. Edge Weighting

Every edge gets a weight to represent marking quality. If a defender is standing right on top of an attacker, they get a perfect score of 1. If they are barely inside the marking radius, they score a 0:

$$w(d_i, a_j) = \max\left(0,\ 1 - \frac{\|d_i - a_j\|}{r}\right)$$

The variable r is the maximum distance a defender can be while still realistically contesting a header. Elite defenders usually stay within two to three meters of their man, and they can close a bit of a gap while the ball is in the air. Therefore, we set $r = 3$ meters as our baseline. We also tested $r \in \{4,5,6,7\}$ meters to see how much the results change based on this assumption.

C. Player Filtering

To make sure the graph accurately reflects the physical battle in the box, we filtered out three types of players:

1. The corner taker, since they aren't a target.
2. The goalkeeper, because they defend the goal instead of a specific player.
3. Anyone standing outside the penalty area.

D. Optimal Marking and Coverability

We ask our code two things for every corner kick. First, can the defense actually cover everyone (using Hall's theorem)? If not, we find the specific group of attackers that broke the defense. Second, what was the best possible way the defenders could have positioned themselves, and what is the maximum weight of that coverage (using the Hungarian algorithm)?

IV. Implementation and Results

A. Data Source

All data comes from the 2022 FIFA World Cup (competition id 43, season id 106) via StatsBomb's open dataset. We pulled passing events specifically tagged as "Corners" and matched them with their corresponding 360 freeze frames. This data was gathered in June 2026.

Out of the whole tournament, 300 corners fit our criteria: they were direct crosses into the box, had complete freeze-frame data, and featured players from both teams actively contesting the area. We threw out short corners because they don't fit a traditional man-marking model. On average, teams loaded the box with about seven attackers per corner.

B. Implementation

The entire pipeline was built in Python. For every valid corner, the script isolates the coordinates of the attackers and defenders, builds the distance matrix, and runs the Hungarian algorithm `(scipy.optimize.linear_sum_assignment)` to find the best matchups. It then hunts for Hall violations by checking if any group of attackers didn't have enough defenders nearby.

C. Example

To make the method concrete, we work through a single corner in full: Canada attacking against Morocco (2022 FIFA World Cup group stage). At the moment the corner is struck, the

360 freeze frame places seven Canadian attackers and ten Moroccan outfield defenders inside the penalty area. Figure 1 shows the configuration; the two attackers highlighted in red are the ones the analysis will prove cannot both be marked.

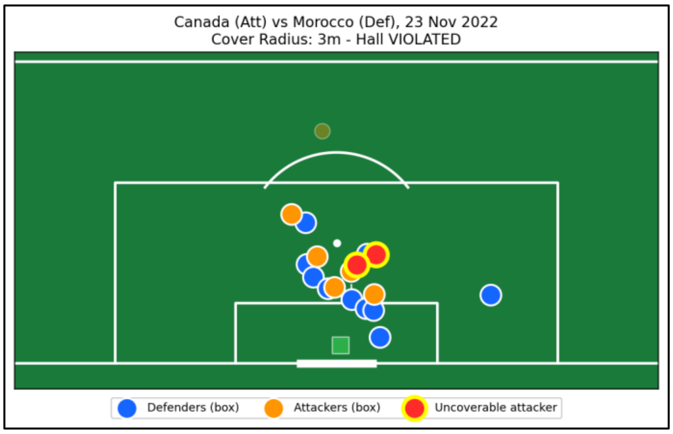

Fig. 1. Corner kick freeze frame for Canada (attacking) vs Morocco at a 3 m cover radius.

1) Player Positions

The attacker and defender pitch coordinates (x, y) extracted from the freeze frame are listed in Table I. Coordinates follow the StatsBomb convention with the attacked goal at $x = 120$.

TABLE I. PLAYER COORDINATES

Attacker	*Coordinates*	*Defender*	*Coordinates*
a_1	(105.1, 35.5)	d_1	(105.9, 36.9)
a_2	(109.2, 43.9)	d_2	(109.1, 43.0)
a_3	(109.3, 38.1)	d_3	(110.1, 37.1)
a_4	(110.1, 42.0)	d_4	(111.4, 37.7)
a_5	(110.8, 41.5)	d_5	(112.5, 39.1)
a_6	(112.4, 39.8)	d_6	(113.2, 55.3)
a_7	(113.1, 43.7)	d_7	(113.6, 41.5)
		d_8	(114.5, 42.9)
		d_9	(114.7, 43.7)
		d_{10}	(117.4, 44.3)

2) Pairings and Edge Weights

From these coordinates, we compute the distance matrix D containing the Euclidean distance between every defender i and attacker j. Applying the marking-quality function with a cover radius of $r = 3$ meters yields the weight matrix W.

Because displaying the full 10×7 matrices is impractical, Table II highlights the crucial pairings that dictate the outcome of this specific corner, showing both the physical distance and the resulting mathematical edge weight $w(d_i, a_j)$.

TABLE II. KEY PAIRINGS, DISTANCES, AND EDGE WEIGHTS

Defender to Attacker	*Distance (D)*	*Weight (W)*	*Status*
$d_1 \to a_1$	1.62 m	0.46	Optimal Match
$d_2 \to a_2$	0.88 m	0.71	Optimal Match
$d_3 \to a_3$	1.29 m	0.57	Optimal Match
$d_5 \to a_6$	0.66 m	0.78	Optimal Match
$d_7 \to a_5$	2.83 m	0.06	Optimal Match
$d_9 \to a_7$	1.24 m	0.59	Optimal Match
$d_2 \to a_4$	1.48 m	0.51	Hall Violator

3) Bipartite Graph

The non-zero entries of W define the edges of the bipartite graph $G = (A \cup D, E)$.

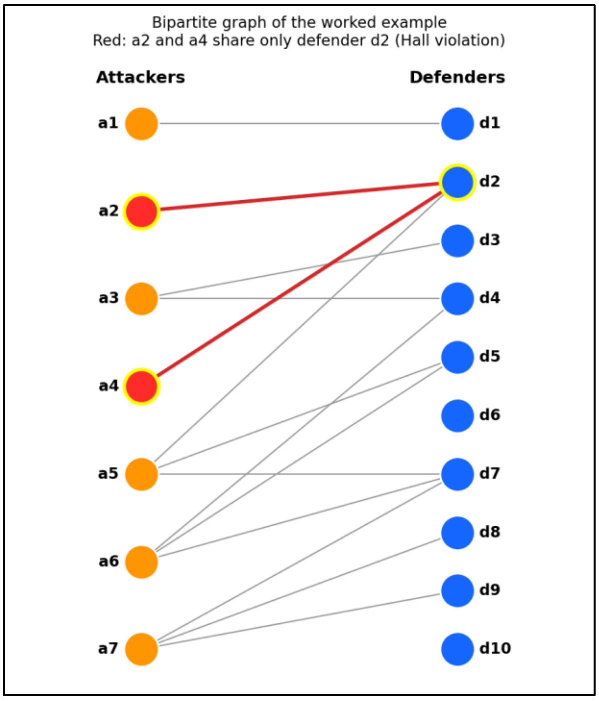

Fig. 2. Bipartite graph of the example.

4) Checking Hall's Condition

With the graph populated, we look for coverability. Consider the subset of two specific attackers, $S = \{a_2, a_4\}$. We find their neighborhood $N(S)$, which is the set of all defenders located within the 3-meter cover radius.

Scanning the distances, only defender d_2 is within 3 meters of either of them (d_2 is 0.88 m from a_2, and 1.48 m from a_4). Therefore:

$$N(a_2) = \{d_2\}$$
$$N(a_4) = \{d_2\}$$

Combining these gives us:

$$N(S) = N(a_2) \cup N(a_4) = \{d_2\}$$

Because the size of the neighborhood is smaller than the size of the attacker subset ($1 < 2$), Hall's condition is violated:

$$|N(S)| < |S|$$

In football terms, attackers a_2 and a_4 are each within marking range of only one defender, d_2. Since a single defender cannot simultaneously mark two players, at least one attacker is necessarily left free.

5) Optimal Marking

Running the Hungarian algorithm on the full weight matrix W confirms this mathematical reality. The maximum-weight matching is assigned as follows:

$$d_1 \to a_1,\ d_2 \to a_2,\ d_3 \to a_3,\ d_5 \to a_6,\ d_7 \to a_5,\ d_9 \to a_7$$

This configuration yields a total defensive coverage weight of 3.046. It leaves attacker a_4 unmatched, no alternative defensive assignment can yield a better outcome. By Hall's theorem, the unmatched attacker is an unavoidable reality of the spatial layout, not an artifact of the chosen matching algorithm. This single corner illustrates, the uncoverability phenomenon that is widespread across the tournament.

D. Hall Violation Rate

The standout finding happens at our realistic $r = 3$ meter radius: 72.3% of the 300 corners had a Hall violation. This means that roughly seven out of ten times a corner was taken, there was literally no mathematical way for the defense to mark every single attacker. At least one offensive player was always free. Naturally, if we pretend defenders can cover more ground, the violation rate drops. Table III outlines this trend.

TABLE III. HALL VIOLATION RATE VS. COVER RADIUS

Cover Radius	*Hall Violation Rate*
3m	72.3%
4m	55.3%
5m	44.3%
6m	38.0%
7m	34.0%

Even if we, for some reason, assume a defender can cover a man from 7 meters away, 34% of corners still result in a free attacker.

E. Comparison Across Teams

When we broke the data down by team (looking at squads with at least four analyzed corners), things got interesting. Teams like Denmark, Japan, Spain, and the Netherlands hit a 1.0 violation rate. Every single corner they took featured at least one unmarkable attacker. This proves that overwhelming the defense with numbers is a standard international tactic, not a fluke.

However, when we compared these perfectly executed setups to actual match outcomes, the correlation fell apart. Generating a free man didn't mean a team actually scored, or even get a shot. Some teams with a 1.0 violation rate barely registered any shots from corners, while teams with average violation rates ended up being the most dangerous in the tournament.

TABLE IV. COMPARISON BETWEEN TEAMS

Team	*Corners*	*Violation Rate*	*Corner Shoots*	*Goals*
Denmark	10	1	11	1
Japan	11	1	6	1
Spain	6	1	7	1
Netherlands	10	1	4	0
Croatia	16	0.88	7	1
Portugal	12	0.83	8	1
South Korea	18	0.83	17	1
Brazil	17	0.82	14	1
France	14	0.71	24	2
Argentina	14	0.5	15	1

Table IV makes the disconnect explicit. France, despite a mid-table violation rate of 0.71, led the entire tournament with 24 shots and 2 goals from corners, more than any of the four teams that achieved a perfect 1.00 violation rate. The Netherlands, at 1.00, managed only 4 shots and no goals, and Spain at 1.00 produced 7 shots and a single goal. South Korea (0.83) and Brazil (0.82) generated 17 and 14 shots respectively, again outperforming the maximum-violation teams.

In other words, the teams that most reliably created uncoverable attackers were not the teams that most reliably threatened the goal. This is direct evidence that a Hall violation is necessary but not sufficient. Scheming a free man into the box is only the first step; delivery quality, timing of runs, and finishing, none of which the positional model captures, ultimately decide whether that advantage becomes a shot or a goal.

F. Discussion

These findings highlight a tough reality in football: freeing up an attacker in the box is necessary to threaten the goal, but it is nowhere near enough to guarantee a score. Our model only looks at the exact moment the ball is struck. It doesn't factor in a poor cross, a mistimed run, a great save by the keeper, or a totally botched header. You can scheme a player open every single time, but if the delivery is bad, the math doesn't matter.

There are two main limitations to keep in mind. First, the freeze frames only show where players start, not where they end up when the ball arrives. Second, a team's violation rate is heavily impacted by the quality of the defense they happen to be playing against. Still, these limitations don't change the broader conclusion: on a crowded corner, defenses simply cannot cover everybody.

V. CONCLUSION

We successfully modeled corner kicks as a matching problem on a bipartite graph, using Hall's theorem to test defensive coverage and the Hungarian algorithm to find optimal matchups. Looking at 300 corners from the 2022 World Cup, we found that at a realistic 3-meter marking radius, defenses suffer a Hall violation 72.3% of the time.

While our cross-team analysis proved that elite teams regularly scheme attackers open, this positional advantage doesn't directly translate into shots or goals. Creating an unmarkable attacker is just the first step; the actual execution of the play dictates the result. Future iterations of this work could model zonal marking, track the ball's flight time, or integrate player heights into the edge weights for an even deeper analysis.

ACKNOWLEDGMENT

Puji syukur kepada Tuhan Yang Maha Esa atas kelancaran pengerjaan makalah ini. Saya juga ingin mengucapkan terima kasih yang sebesar-besarnya kepada dosen pengampu mata kuliah IF1220 Matematika Diskrit, Bapak Rinaldi, atas bimbingan dan ilmunya selama semester ini. Tidak lupa, terima kasih kepada keluarga dan teman-teman yang selalu memberikan dukungan dan semangat.

REFERENCES

[1] P. Hall, "On Representatives of Subsets," J. London Math. Soc., vol. s1-10, no. 1, pp. 26–30, 1935.

[2] H. W. Kuhn, "The Hungarian method for the assignment problem," Naval Research Logistics Quarterly, vol. 2, no. 1–2, pp. 83–97, 1955.

[3] K. H. Rosen, Discrete Mathematics and Its Applications, 8th ed. New York: McGraw-Hill, 2019.

[4] R. Munir, "Graf," lecture notes for IF1220 Matematika Diskrit, Institut Teknologi Bandung. [Online]. Available: https://informatika.stei.itb.ac.id/~rinaldi.munir/Matdis/ [Accessed: Jun. 2026].

[5] StatsBomb, "StatsBomb Open Data," GitHub repository. [Online]. Available: https://github.com/statsbomb/open-data [Accessed: Jun. 2026].

PERNYATAAN

Dengan ini saya menyatakan bahwa makalah yang saya tulis ini adalah tulisan saya sendiri, bukan saduran, atau terjemahan dari makalah orang lain, dan bukan plagiasi.

Bandung, 19 Juni 2026

Danesh Rasyad Damotino – 13525079