

Mythbusting Set-Pieces in Soccer

Paul Power, Jennifer Hobbs, Hector Ruiz, Xinyu Wei, Patrick Lucey

STATS AI Group paul.power@stats.com

Paper Track: Other Sports (Soccer)

Paper ID: 5631

1. Introduction

In the 2016-2017 English Premier League season, 16% of all goals scored came from set-pieces (corners and free kicks). However, there exists a great disparity in those numbers as a team such as West Bromwich Albion scored 16 out of their 43 goals from set-pieces (>35% of their goals), whilst others obtain less than 7% of their goals from set-pieces (e.g., Sunderland scored 2 out of their 29 goals from set-pieces). The gap in resources between the richest and poorest teams in world football is growing wider each season as demonstrated by Paris St German spending a world record 220 million Euro on Neymar who scored 15 goals. The ability for a small market team to replicate the same goal output for the price of an effective set-piece strategy is a clear market inefficiency that can be exploited. Therefore, any new method that can help teams exploit this inefficiency would be a key advantage.

To get an indication on how valuable certain facets of set-pieces are, we first take a **mythbusting** approach to highlight key elements around set-pieces. The ten myths we analyzed and their outcomes are summarized below:

- **Myth 1:** A team is more likely to score from set-pieces compared to normal play possession
 - **Confirmed:** 1.8% chance of scoring from set-pieces vs 1.1% in open play ($p < 0.05$)
- **Myth 2:** A team is more likely to score from a free-kick cross compared to a corner
 - **Busted:** 2.1% chance of scoring from a corner vs 1.1% chance of scoring from a freekick ($p < 0.05$)
- **Myth 3:** From free-kicks, teams are more likely to score by shooting directly, rather than crossing the ball
 - **Confirmed:** 7.2% chance of scoring from shooting directly vs 1.1% chance from crossing the ball ($p < 0.05$)
- **Myth 4:** A team is more likely to concede from their own corner than score
 - **Busted:** 0.2% chance of conceding vs 2.1% of scoring a goal ($p < 0.05$)
- **Myth 5:** Teams are more likely to score from in-swinging corners compared to out-swinging or driven corners
 - **Confirmed:** 2.7% chance of scoring from in-swinging corners vs 2.2% from out-swinging ($p < 0.05$)
- **Myth 6:** In corners, a goal is more likely to occur from a shot from the second-ball rather than a shot directly
 - **Confirmed:** Teams have a 2.5% chance of scoring after winning the second ball vs 2.0% directly ($p = 0.86$)
- **Myth 7:** In corners, a team is more likely to score from a flick-on compared to directly shooting from a corner
 - **Confirmed:** Teams have a 4.8% chance of scoring from a redirected corner vs 2.0% directly ($p < 0.05$)
- **Myth 8:** When defending a corner, zonal marking is the least effective method
 - **Busted:** A hybrid defense concedes most dangerous shots (12% conversion, $p < 0.05$)
- **Myth 9:** When defending a corner, a team should have players on the post (near and/or far)
 - **Busted:** Teams with players on both posts concedes the most goals (2.7%) while a player on back post concedes the least (0.9%, $p < 0.05$)
- **Myth 10:** A set-piece specialist is more valuable to mid-to-low level teams, compared to high-level teams
 - **Confirmed:** An average set-piece taker will win a team 0.9 points while an elite set-piece taker will win 1.9 points (worth ~8% of a team's points for a bottom 6 team vs ~3.5% for a top 6 team.)

From the above list, it is clear that type of delivery (i.e., in-swinging, flick-on) and the defensive set-up (i.e., player on the post, zonal vs man-to-man) can significantly improve a team's chances of

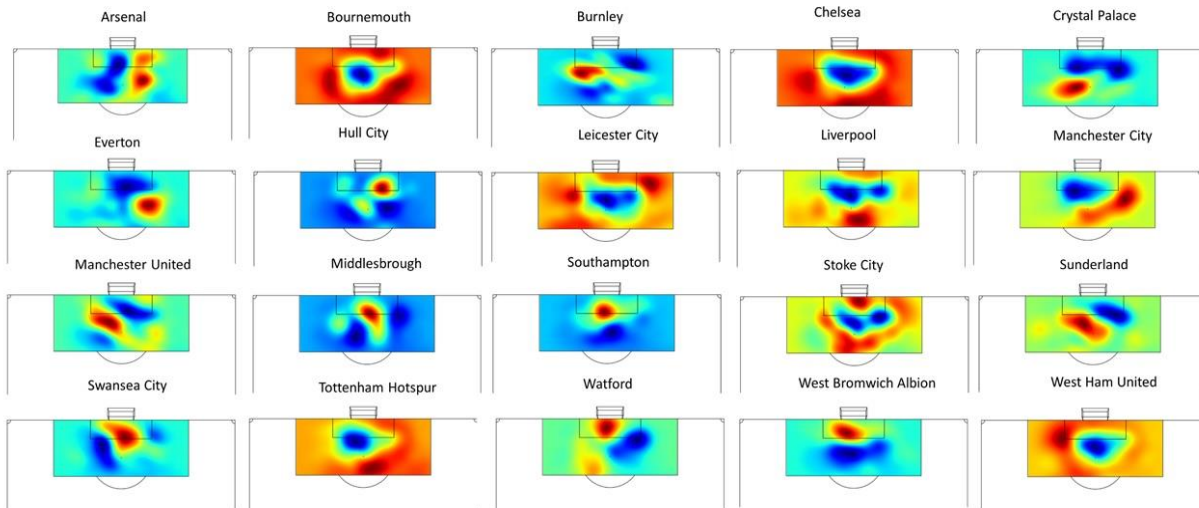


Figure 1. Probability maps of where teams create shots from corners compared to the league average. Data is modeled as a two dimensional Gaussian distribution with red zones showing where teams are more likely and blue zones where they are less likely to have a shot from.

scoring. However, it must be noted, that these are *global trends* across the league and do not take into consideration the *specific characteristics of a team* such as the location of delivery and the relative height/physique of players on a team. For example in Figure 1, we show the difference in location of where teams tend to create chances from corners compared to the league average. We can see that West Bromwich Albion tend to create chances within the 6-yard box while Sunderland tend to get chances outside of the 6-yard box.

In this paper, we present an *attribute-driven approach* to set-piece analysis, which utilizes a hybrid of deep-learning methods to detect complex attributes such as defensive marking schemes, and hand-crafted features to enable interpretability. Specifically, we employ a *Convolutional Neural Network (CNN)*, which adequately captures the defensive structure of a team around set-pieces. Additionally, we use *expected metrics* such as expected goal value (xG) to value the quality of chances that a team creates based on the location and quality of delivery in addition to the defensive attributes.

To do our analysis, we used three seasons worth of English Premier League (EPL) data, which contains both tracking data of the players and ball, in addition to event labels. Overall, this dataset consists of over 12,000 corners and 3,600 free-kicks. The rest of the paper is as follows: in the next section we describe our dataset and the hand-crafted features we extracted. In Section 3 we describe our image-based representation and CNN approach which captures player-to-player interactions. In Section 4, we do league wide analysis of set-piece performance in the EPL for the 16-17 season, and in Section 5 we present a framework to predict/scout upcoming team set-pieces.

2. Set-Piece Segmentation and Grammar

The data we used for our analysis contains the player and ball x,y locations sampled at 10 Hz for both teams as well as event location information. A current limitation within the existing body of research for set-pieces is the definition of when a set-piece stops [2, 3, 4]. This is normally

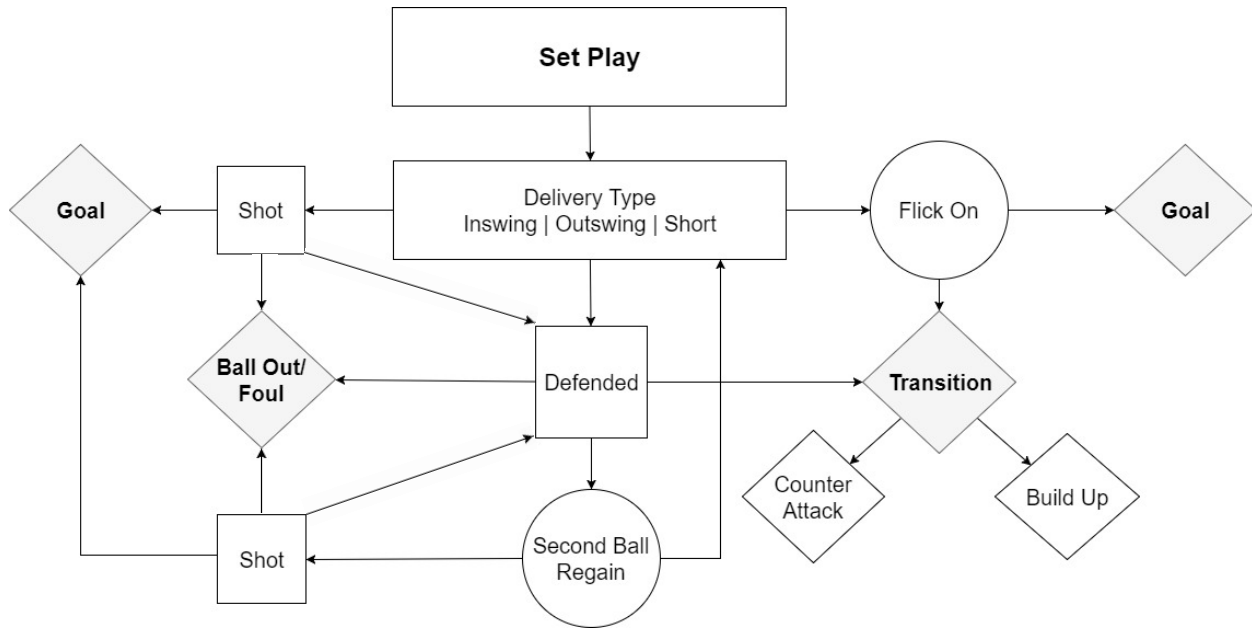


Figure 2. Set-Piece Grammar: Start - transition - end nodes. Each path is created via a rule-based system, which is defined via coaching experts.

determined to be the first contact after the set-piece is taken (e.g., a shot or clearance) [3]. However, this fails to capture second ball regains or situations where the attacking team look to redirect the path of the ball via a flick-on, for example.

Therefore, to capture the evolving dynamics of a set-piece we created a *set-piece grammar model* based on domain expertise (Figure 2). The model defines the temporal sequencing of events during example, a corner could start with an in-swinging cross, which is flicked-on at the near post for a second attacker to score. This situation has a clear beginning, middle and end. However, another example may see an attacking free-kick passed short for it to be crossed into the penalty area, which is then cleared by a defender only to be regained by an attacker and crossed back into the penalty area for the goalkeeper to claim and drop at the attacker's feet to score. This second situation is much more complex with three potential moments when the set-piece could be deemed to be finished (initial clearance, goalkeeper claim or goal). We therefore need a grammar, which enables multiple phases to occur.

In Figure 2, we showcase our defined grammar for set-pieces. The model captures the start and end of a set-piece and the specific type of events that occur in-between. We are now able to move away from just looking at the first contact after a set-piece delivery and capture which teams are able to create goals via flick-on's and second ball regains (which can be thought of as rebounds). We can also measure which defending teams are able to transition into counter attacks.

Now that we have a grammar that can decouple important phases during a set-piece, we can now incorporate the fine-grain information such as the location of the delivery.

No matter how well executed the attacking movement may be for a set-piece, if the ball does not get beyond the first defender the set-piece will never be successful. Additionally, the type of delivery has a large influence on whether a shot or goal is created. For example, in-swinging corners have a 18.61% chance of leading to a shot compared to 20.85% for out-swinging. Conversely, when a shot

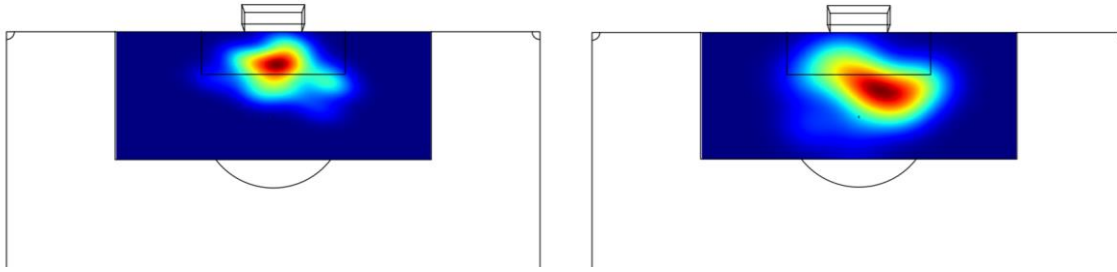


Figure 3. Distribution of where goals are scored for (left) in-swinging and (right) out-swinging corners. Corner deliveries have been standardized to come from the right hand side.

is taken, an in-swinging cross has a 10.81% chance of being scored compare to just 6.46% for an out-swinging cross.

This indicates that while it is harder to create a shot from an in-swinging delivery you have nearly double the chance of scoring. The key reason for this is due to the position from which the shot is taken. Figure 3 shows the distribution of where goals are scored based on both methods - *left* is the in-swinging and the *right* are the out-swinging (normalize to come from the right hand-side). For the in-swinging heat-map, it can be seen that the area from which the goals are scored is significantly smaller but much closer to the goal than for out-swinging corners. This simple example highlights the additional level of detailed analysis our grammar model enables us to perform.

2.2 Detecting Defensive Tactics via Convolutional Neural Networks

When analyzing the defensive elements of set-pieces, specifically corners, there are two main attributes of defensive play that are of interest: i) determining whether a team is playing man-to-man or zonal marking, and ii) determining if they have defenders on the post or not (i.e., front, back or both). For the latter, given tracking data, it is quite straight forward to determine whether a player is on the post or not using simple data association (i.e., assigning a player to the post if his distance is within a certain threshold).

For the former, however, this is very challenging. It is challenging because the current methods of representing player-tracking data are not adequate to capture the nuances of corner behavior. Previous research has used a template approach to align player positions to a role [5]. This has proved to be a highly effect method for boosting prediction tasks during open play such as predicting expected goal value of a shot [6] or modeling defensive behavior [7] where players motion paths are relatively linear with few role swaps. However, such approaches assume that all the players are important and present in the play - which is often not the case for set-pieces. For example, some attacking teams may keep 3 defenders back meaning there could be 4 attackers vs 8 defenders; this type of representation would not be able to deal with due to the varying number of 'active' players.

Another approach would be to take a bottom-up/feature-crating approach [8], where the distance and angle between players is calculated. However, this approach is limited by the immense number and complexity of the features needed to capture nuances such as "on-the-post",

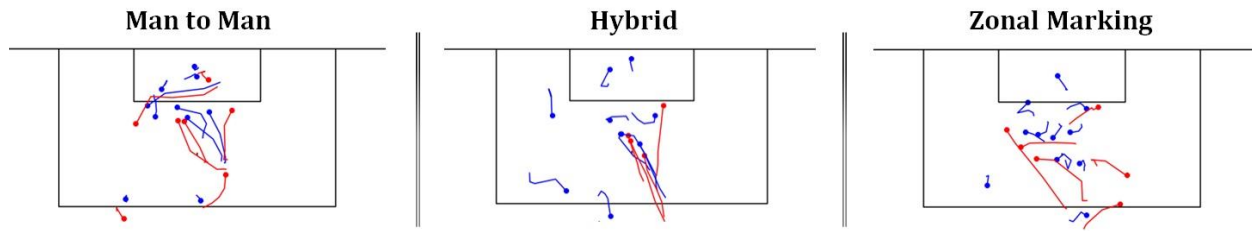


Figure 4. (Left) Man to Man marking with no players on the post. (Middle) Hybrid marking with player on front zone, (Right) Zonal Marking.

“double-teamed”, or “zonal-marking”. Additionally, as the area around the goal is rather small, with the distance closest to the goal being the most important, fine-nuanced movement may be ignored by large movements of a few players.

However, even though current approaches maybe inadequate to capture such behavior, the fine-grained space, variable number of players and different important motion patterns (short and long) are ideal for another type of representation - *images*. As can be seen in Figure 4, for a human expert, it is quite clear that we can observe the different nuances between corner play. On the right we see a zonal marking scheme, the left we have a man-to-man, and the middle we have a hybrid approach. Visually – the patterns are clear to us, which gives us a strong clue as how to represent this behavior.

Recent work in the computer vision field, has utilized a convolutional neural network (CNN) to learn the relative predictive features from images. Instead of getting humans to hand-craft features to represent images (e.g., edge detectors, wavelets), it has been shown the CNN can learn these features better given a large amount of data. The intuition is that they are able to pick up on local edge patterns and interactions, and compose them together in a hierarchical way to better represent image-data. Leveraging from the impressive body of work in the computer vision field, we have utilized a similar approach for our task of detecting defensive structures.

To do this, for each corner we extract out the image of player motion two seconds before and after a corner is taken and create a single channel grayscale image (offensive team is black and the defensive team is white). By utilizing the image, it can handle a variable number of players and also captures the interaction of players near each other without having to hand-craft the distances and angles at every frame.

The input image is of size 46 x 25 pixels and is fed through our CNN. The architecture of our CNN used two hidden layers with 80 and 50 neurons. Relu activation was applied at each layer with a softmax logit function used for the final label prediction. To train our CNN we used 1500 training examples, and treated it as a three-class problem with 80% being hybrid, 10% man to man and 10% zonal marking. Our accuracy of detecting the defending type was 82.53%.

Offensive Efficiency					Defensive Efficiency				
Ranking	Team	Goals Scored	Shots Taken	Efficiency	Ranking	Team	Goals Conceded	Shots Conceded	Efficiency
1	West Bromwich Albion	16	117	13.68%	1	Bournemouth	6	143	4.20%
2	Swansea City	13	119	10.92%	2	Chelsea	4	92	4.35%
3	Chelsea	15	144	10.42%	3	Arsenal	5	114	4.39%
4	Hull City	9	94	9.57%	4	Liverpool	5	97	5.15%
5	Tottenham Hotspur	15	165	9.09%	5	Burnley	10	188	5.32%
6	West Ham United	10	120	8.33%	6	Tottenham Hotspur	5	92	5.43%
7	Manchester City	12	161	7.45%	7	West Ham United	6	105	5.71%
8	Bournemouth	9	123	7.32%	8	Manchester City	6	97	6.19%
9	Middlesbrough	6	85	7.06%	9	Manchester United	7	107	6.54%
10	Liverpool	11	161	6.83%	10	West Bromwich Albion	8	119	6.72%
11	Crystal Palace	9	133	6.77%	11	Middlesbrough	10	146	6.85%
12	Southampton	8	124	6.45%	12	Stoke City	11	155	7.10%
13	Leicester City	6	111	5.41%	13	Sunderland	14	178	7.87%
14	Watford	7	130	5.38%	14	Everton	10	125	8.00%
15	Manchester United	7	130	5.38%	15	Watford	12	146	8.22%
16	Stoke City	6	118	5.08%	16	Swansea City	8	96	8.33%
17	Arsenal	7	142	4.93%	17	Leicester City	11	128	8.59%
18	Everton	6	125	4.80%	18	Southampton	12	129	9.30%
19	Burnley	5	137	3.65%	19	Crystal Palace	12	122	9.84%
20	Sunderland	2	83	2.41%	20	Hull City	17	143	11.89%

Table 1. League ranking for teams from the 2016/17 EPL season for their offensive and defensive goal scoring efficiency at set-pieces.

3. Set-Piece Analysis of the 2016-17 EPL Season

To show the utility of our approach, we first conducted analysis of the offensive and defensive behaviors of all teams from the 2016/17 English Premier League season. In Table 1, we show the ranking of all teams with respect to their offensive efficiency (left) and the defensive efficiency (right). With respect to the offensive efficiency, we can see that West Bromwich Albion were both the most efficient team (13.68% conversion rate) and highest scoring (16 goals) with Sunderland being the least effective scoring only 2 goals from set-pieces at a conversion rate of 2.41%. Sunderland were relegated in this season whilst West Bromwich Albion finished 10th. Interestingly though, both teams scored 18 and 19 goals from open play respectively highlighting how vital a role an effective set-piece strategy can have on a team's chances of surviving.

Defensively, Bournemouth were the most efficient team in preventing goals from being scored (4.20% conversion rate) while Hull City were the least effective conceding the highest number of goals (17) at the highest rate (11.89%). Similar to Sunderland, Hull City were relegated with the worst goal difference in the league of -43. While Hull City were the fourth most effective team at scoring from set-pieces scoring 9, their net goal difference from set-pieces was -8. If Hull City had been able to be as effective in defense as they were in attack they may have had a better chance of staying up.

While it is useful to be able to measure which teams are effective at scoring or preventing goals from set-pieces, this information alone is not enough for a coach to act upon when preparing for the next game. We therefore need a better representation of teams' set-piece style. Using the grammar we defined in Section 2 to describe offensive behavior, and methods we described in Section 3 to detect the defensive schemes, we can now perform objective analysis of set-pieces.

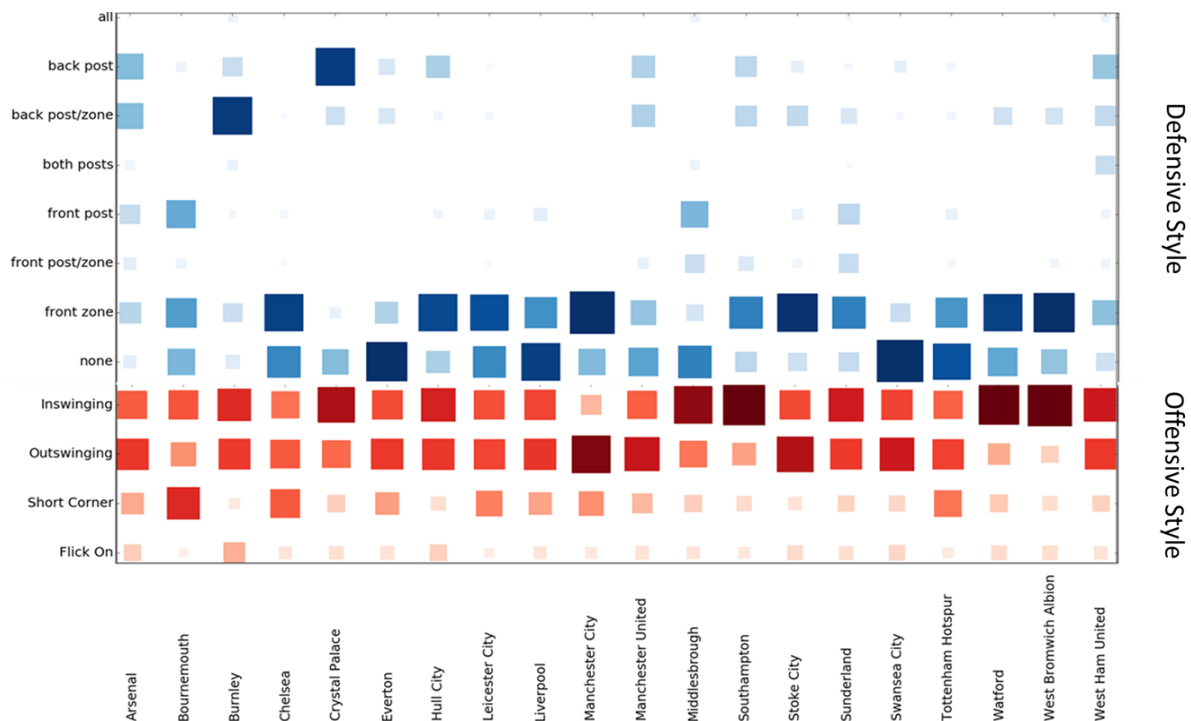


Figure 5. Hinton diagram representing a team's offensive and defensive set-piece style. The larger the size of the square the higher usage and darker color the more effective.

Figure 5 shows each team's average attacking and defending set-piece style for the 2016-17 EPL season, and their effectiveness at both scoring and preventing goals is visualized via our Hinton Diagram. We can clearly see that different teams have different strategies, for example, Chelsea and West Bromwich Albion both scored 16 goals from set-pieces (excluding direct shots from free kicks). However, the method by which they achieved this is significantly different. West Bromwich Albion's corners style was highly predictable with 73% of their corners being in-swinging while Chelsea were much more unpredictable using a mixture of in-swing, out-swing and short corners.

4. Predicting Future Opponent Set-Piece Behavior

When preparing to play an upcoming opponent, an analyst will spend a significant amount of time analyzing their recent set-piece play. At a high-level, he/she is initially interested in answering the four following questions:

- How do my opponents defend set-pieces?* (i.e., zonal vs man-marking vs hybrid)
- How do my opponents deliver set-pieces?* (i.e., short, long, flick-on's, location)
- Where are my opponents strong and weak?* (i.e., what plans do we have to put in place to stop them and exploit them?)
- How are my opponents most likely to play against my team?* (i.e., based on how we defend how will they deliver the ball?)

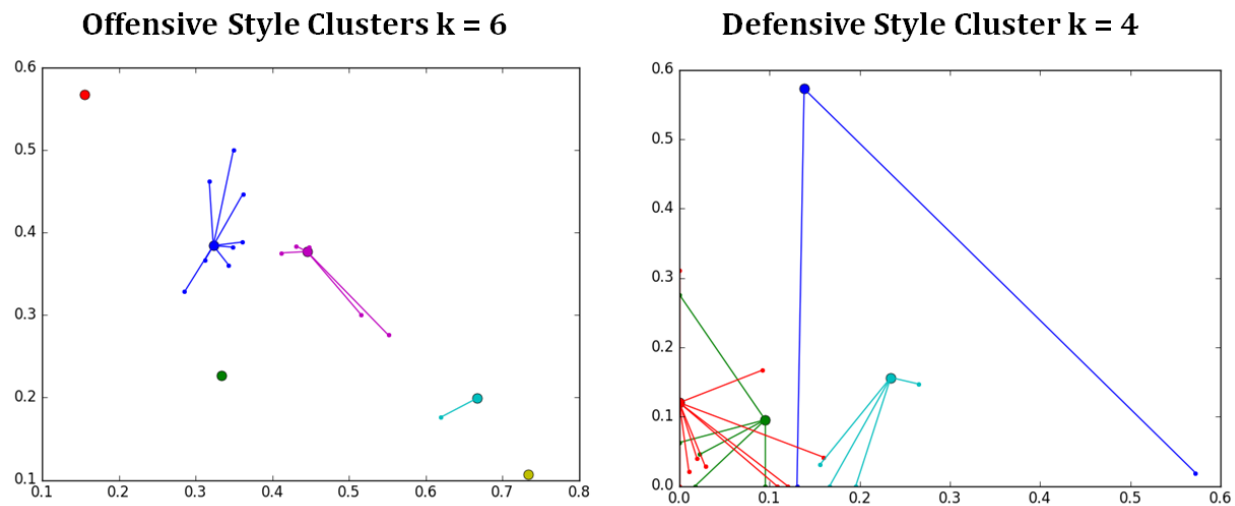


Figure 6. Affinity propagation clusters for the attacking set-piece styles (left) and defensive set-piece styles (right). Large nodes are the representative team for that cluster. In order to visualize the clusters we use two of the dimensions inswinging vs outswinging for offensive plot and back post zone vs front zone for the defensive plot.

In this section, we demonstrate how our new analysis can help analysts answer these key questions via three case studies.

4.1 How Similar are the Opposition to Teams We've Already Played?

A key issue in scouting the opposition for the upcoming game is in selecting the most relevant matches to analyse. The last matchup is a single example and not necessarily the most relevant: injuries, personnel changes, home vs. away status, relegation/promotion, and simply the time since the last matchup all impact the “relevance” of a given match in complex ways. Finding sufficient relevant examples is a highly time consuming task that our method can help alleviate.

In order to determine if a team has a unique attacking or defending style we use a method called affinity propagation to cluster the average team behavior [12]. Affinity propagation is an unsupervised clustering technique that works by finding exemplars, which can be thought of as centroids. The Euclidean squared distance is then calculated between the exemplar and all the other teams to determine how similar teams are to each other. The team that best represents that cluster is identified as the final exemplar.

Based on this method we found six clusters for offensive style and four for the defensive style (figure 6). For the offensive style we find three unique teams (West Bromwich Albion (red), Manchester City (blue) and Bournemouth (green)), a cluster of two ((Southampton and Watford (light blue))) and two larger clusters which represent teams with large variation in their delivery (blue) and teams who focus on inswing and outswing set-pieces (purple). For the defensive style, we found four cluster types, which indicates that a team's defensive set-up may be more predictable than a team's attacking style.

Using this technique an analyst could quickly see if the opponent has played a team within their cluster and see how they perform against that style both offensively and defensively.

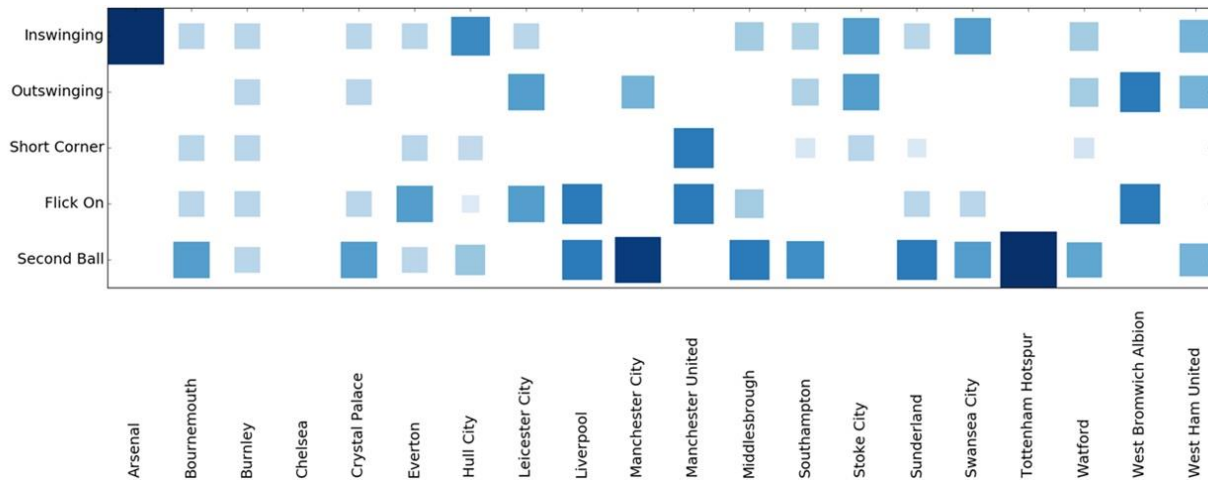


Figure 7. Hinton diagram representing how a team is most likely to conceded a goal from a corner. The larger the size of the square and darker color the more goals conceded.

4.2 Exploiting Defensive Set Up

An issue with analysis can be the recent bias effect [9] where a rare event might occur in consecutive matches leading to the conclusion that a team is either good or bad at that. For example, Liverpool have been highly criticized for how they defend corners. Numerous articles [10, 11] point to their use of zonal marking as the reason they conceded 2 goals in 3 games (which they happened to lose). However, as demonstrated in Table 1, Liverpool had the fourth most efficient defense in 16/17 season. By using our new method to detect the defensive tactics employed by a team, we are now able to assess a team's approach to defending a corner.

Liverpool used zonal marking 32.57% of the time, conceding 3 goals, and hybrid 67.43%, conceding 2 goals. If we look at the number of shots allowed, Contrary to popular belief, Liverpool conceded significantly fewer shots than expected when using either zonal marking (4.02%) and hybrid (4.65%) compared to the league average (9%, $p < 0.01$). However, when Liverpool did concede a shot, the average expected goal value was 20.13% when using zonal marking, compared to 8.06% using a hybrid. This indicates that while Liverpool's defensive set up was significantly better than average, the shots conceded during zonal marking were more than twice as dangerous. Furthermore, when we examine *how* the goals were conceded via our new segmentation in Figure 7, we see that Liverpool were highly susceptible to shots coming from flick-on's or second balls. This is critical and actionable information that a coach can access quickly to plan around.

4.3 Manipulating the Oppositions Delivery

To determine whether a team is predictable in their delivery, we first should look at how variable each team is. In Figure 8 (left), we rank the teams in terms of their offensive variance of delivery. To do this, we utilized the delivery attributes (i.e., in-swinging, out-swinging, short, flick-on) to

Attacking Style Variance

Ranking	Team	Variance
1	Manchester City	0.97
2	Stoke City	0.87
3	Liverpool	0.81
4	Chelsea	0.75
5	Everton	0.75
6	Tottenham Hotspur	0.74
7	Southampton	0.73
8	Burnley	0.73
9	Watford	0.72
10	West Ham United	0.72
11	Hull City	0.69
12	Arsenal	0.69
13	Swansea City	0.67
14	Crystal Palace	0.66
15	Middlesbrough	0.63
16	Sunderland	0.60
17	Bournemouth	0.60
18	West Bromwich Albion	0.53
19	Leicester City	0.52
20	Manchester United	0.42

Manchester City Delivery Style

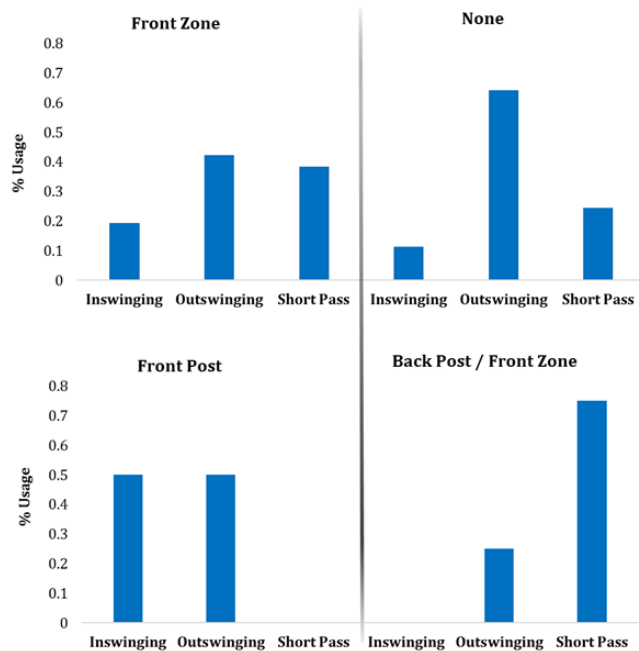


Figure 8. (Left): Show the variance in delivery. **(Right)** Manchester City delivery style.

create an offensive descriptor for a team in each match. We then normalized and scaled this descriptor across the previous 5 matches and calculated the L2 distance to the teams season average to create an average style metric. Finally, we calculated the variance of each teams average style metric for the season, to create an overall measure of set-piece variability per team. The selection of a sliding 5 game window was to mimic the behavior of scouts who normally use the previous 5 games, as it captures the most recent team form.

Figure 8 reveals that Manchester City had the highest amount of variance game-to-game whilst Manchester United showed the least amount of variance from their average style. Interestingly when we look at Manchester United's delivery style in Figure 5 we can see that they use a variation between short and long deliveries. This indicates they tend to vary their in-game rather than game-to-game strategy.

This is the inverse of what we observe with Manchester City, who at first glance, appear to be more predictive based on their average style plot (having a predominance for out-swinging corners). However, an examination of their variability suggests that they may actually change their style based on how the opposition play.

To understand what is driving Manchester City's variance we examine how their style changes based on the defensive set up. The defensive system employed (man to man, zonal or hybrid) had no significant influence on the type of delivery ($p = 0.95$), whereas how the opposition defended the posts did ($p=0.045$). Figure 8 shows how Manchester City tended to play a short corner when playing against teams with a player protecting the front zone. This is a common technique used to attract a defender from the near post to create space to deliver the ball towards. As we saw in

Figure 3, out-swinging deliveries are the most dangerous when met at the near post. Therefore, tactically it makes sense to try to dislodge a defender from this area. However, short corners were Manchester City's least effective style of scoring, creating only 12 shots from 68 corners with no goals. A coach can use this information to try to manipulate how the opposition play in order to help keep his defenders one-step ahead.

5. Summary

In this paper, we were able to bust several long held myths such as; you are less likely to conceded a goal from your own corner than score yourself and that in-swinging corners are more dangerous than out-swinging. We introduced a novel, imaged-based representation approach for analyzing set-piece situations to detect defensive structure. In addition, we were also able to demonstrate, through our new set-piece segmentation, which types of delivery are the most dangerous and how this varies by team. As a result, we are now able to provide a recommendation to a coach and analyst about how a team may play against them and how to prepare for and exploit the opposition's strengths and weaknesses.

References

- [1] <http://www.espn.com/soccer/club/slavia-mozyr/613/blog/post/3170513/neymar-from-barcelona-to-psg-for-222m-euros-timeline-as-worlds-biggest-transfer-nears?src=com>
- [2] C. Pulling, Long Corner Kicks In The English Premier League: Deliveries Into The Goal Area And Critical Area, *Kinesiology*, 47, 193-201, 2015.
- [3] C. Casal, R. Dios, T. Arda and A. Boubeta, Analysis of Corner Kick Success in Elite Football, *International Journal of Performance Analysis*, 15, 430-451, 2015
- [4] C. Anderson and D. Sulley, The Numbers Game: Why Everything You Know About Football is Wrong.
- [5] A. Bialkowski, P. Lucey, P. Carr, Y. Yue, S. Sridharan and I. Matthews. "Large-scale analysis of soccer matches using spatiotemporal tracking data". In 2014 IEEE International Conference on Data Mining (ICDM), 2014
- [6] P. Lucey, A. Bialkowski, M. Monfort, P. Carr and I. Matthews, "Quality vs Quantity: Improved Shot Prediction in Soccer using Strategic Features from Spatiotemporal Data", in MIT SSAC, 2015.
- [7] H. Le, P. Carr, Y. Yue and P. Lucey, Data-Driven Ghosting using Deep Imitation Learning, in MIT SSAC, 2016.
- [8] D. Cervone, A. D'Amour, L. Bornn and K. Goldsberry, "POINTWISE: Predicting Points and Valuing Decisions in Real Time with NBA Optical Tracking Data" at MIT SSAC, 2014.
- [9] L. Ross and C. Anderson, "Shortcomings in the attribution process: On the origins and maintenance of erroneous social assessments", in Kahneman, Daniel; Slovic, Paul; Tversky, Amos, *Judgment under uncertainty: Heuristics and biases*, Cambridge University Press, pp. 129–52, 1982.
- [10] <http://www.skysports.com/football/news/15117/10985307/liverpools-set-piece-problem-exposed-in-3-3-draw-with-watford>
- [11] <http://www.mirror.co.uk/sport/football/news/liverpool-klopp-van-dijk-setpiece-10981217>
- [12] B. Frey and D. Dueck, Clustering by Passing Messages Between Data Points, *Science*, 315, 2007.