

Searching for a Unique Style in Soccer

Laszlo Gyarmati, Haewoon Kwak
Qatar Computing Research Institute
{lgyarmati,hkwak}@qf.org.qa

Pablo Rodriguez
Telefonica Research
pablorr@tid.es

ABSTRACT

Is it possible to have a unique, recognizable style in soccer nowadays? We address this question by proposing a method to quantify the motif characteristics of soccer teams based on their pass networks. We introduce the concept of “flow motifs” to characterize the statistically significant pass sequence patterns. It extends the idea of the network motifs, highly significant subgraphs that usually consists of three or four nodes. The analysis of the motifs in the pass networks allows us to compare and differentiate the styles of different teams. Although most teams tend to apply homogenous style, surprisingly, a unique strategy of soccer exists. Specifically, FC Barcelona’s famous tiki-taka does not consist of uncountable random passes but rather has a precise, finely constructed structure.

Categories and Subject Descriptors

G.3 [Probability and Statistics]: Probabilistic algorithms;
I.5.4 [Pattern Recognition]: Applications

General Terms

Algorithms, Theory

Keywords

soccer analytics, network motifs, pass network

1. INTRODUCTION

Although the basic rules of soccer have barely changed since the 1920s¹, they intrinsically enabled teams to develop distinctive strategies that dominated the soccer landscape for several years (e.g., the Hungarian and the Dutch team of the 50s and 70s) [12]. However, in these days the possibilities to prepare against these teams and thus ruin their

¹the introduction of the two-opponent version of the offside rule

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Copyright 20XX ACM X-XXXXX-XX-X/XX/XX ...\$15.00.

strategies was limited due to technological reasons. Nowadays, technological advancements of the last decade allow team staffs to view any first division soccer game on a short notice; hence, it seems to be challenging to have and sustain a unique playing characteristic—which is additionally successful too—in the global soccer space. Does such a unique, recognizable style exist in soccer nowadays?

The identification and understanding of the style of soccer teams have practical impacts apart from the esthetics of the game. The players of a team should obey the style (i.e., strategy) of the team to maximize the team’s chance to win. Hence, it is crucial to raise youngsters and to sign players who are capable of playing according to the style of the team. Failing to do so not only has an impact on the success but also on the profitability of the club. There are numerous examples of newly signed players who were not compatible with the style of their new clubs [11], therefore, there is a need for a quantitative analysis of team’s style to avoid these discrepancies.

The rareness of goals is the profoundest feature of soccer that distinguishes it from other team sports. Although the teams have 11 players and 90 minutes to score in each match, it is not unusual to have a goalless draw as the final score [1]. These results are not solely due to a spectacular performance of the goalkeepers but rather a consequence of the low number of scoring chances. Hence, metrics related to scoring cannot describe the style (i.e., the strategy) of a soccer team.

Passes, on the other hand, happen numerously in every game irrespective of the quality of the teams. The pass network of a soccer team consists of the players as vertices and the passes between the players as the edges. Prior art focused either on the high-level statistics of the pass networks (e.g., betweenness, shortest paths) or the strength of the connection between pairs of players [2, 8, 6, 4]. These metrics describe the static properties of a pass network, e.g., the metrics aggregate all the passes into one network, neglect the order of passes, etc. On the contrary, we focus on the dynamic aspects of the pass networks by examining the “flow motifs” of the teams.

We propose the concept of “flow motifs” to characterize the statistically significant pass sequence patterns. It extends the idea of the network motifs, highly significant subgraphs that usually consists of three or four nodes, suggested by Milo et al. [5] that mainly apply to the static complex networks (e.g., food webs, protein-structure networks, and social networks). We extend their work towards “flow motifs” to analyze pass networks that are highly dynamic and in

which the order of connections is important. Our methodology starts with the extraction of the passing sequences, i.e., the order of players whom the ball traversed. Afterwards, we determine computationally the significance of the different k-pass-long motifs in the passing style of the teams. Our flow motif profile focuses on how ball traverses within a team. We not only count the number of passes, but also check which players are involved in, and how they organize the flow of passes. Based on these computed flow motif profiles, we finally cluster the teams. To the best of our knowledge, our study is the first of its kind that investigates motifs in soccer passing sequences. Our contribution is twofold:

1. we propose a method to quantify the motif characteristics of soccer teams based on their pass networks, and
2. we identify similarities and disparities between teams and leagues using the teams’ motif fingerprints.

In the recent decade, several data-provider companies and websites have arisen to annotate soccer matches and to publish soccer datasets. For example, such initiatives include Prozone [9], OptaPro [7], Instat Football [3], and Squawka [10], among others. The prevalence of data-providers enables us to take a data-driven, quantitative approach to identify the styles of the soccer teams. We focus on the 2012/13 seasons of major European soccer leagues and analyze the passing strategies of the teams throughout the whole season.

2. METHODOLOGY

The “flow motifs” of a pass network, in which players are linked via executed passes, consist of a given number of consecutive passes, namely, an ordered list of players who were involved in the particular passes. Throughout this paper we focus on motifs consisting of three consecutive passes, however, it is straightforward to generalize our methodology to investigate motifs with fewer/more passes. Our methodology relaxes the identity of the involved players, i.e., it does not differentiate motifs based on the names of the players, rather focuses on the certain structure of the passes. There are five distinct motifs when we analyze three-pass long motifs: ABAB, ABAC, ABCA, ABCB, and ABCD. For example, the motif ABAB denotes the following pass sequence: first, player 1 passes to player 2; second, player 2 passes the ball back to player 1; and finally, player 1 passes again to player 2. If a similar pass sequence happens between player 3 and player 4 the identified motif is ABAB again (i.e., the crucial characteristic is what happened and not between whom).

Our methodology quantifies the prevalence of the flow motifs in the pass networks compared to random networks whose degree distribution is the same. To achieve this, we start with a list of passes that a team made during a match. The format of a pass record is

$$p_n = \langle \text{player}_i(n), \text{player}_j(n), t(n) \rangle$$

where $\text{player}_i(n)$ passed the ball to $\text{player}_j(n)$ in the $t(n)$ time instance. Second, we derive all the ball possessions that a team had. A ball possession $\langle p_1, p_2, \dots, p_n \rangle$ consists of such passes that fulfill two constraints:

$$\begin{aligned} \text{player}_j(m) &= \text{player}_i(m+1), \forall m \in \{1, \dots, n-1\} \\ t(m+1) - t(m) &\leq T_{\max}, \forall m \in \{1, \dots, n-1\} \end{aligned}$$

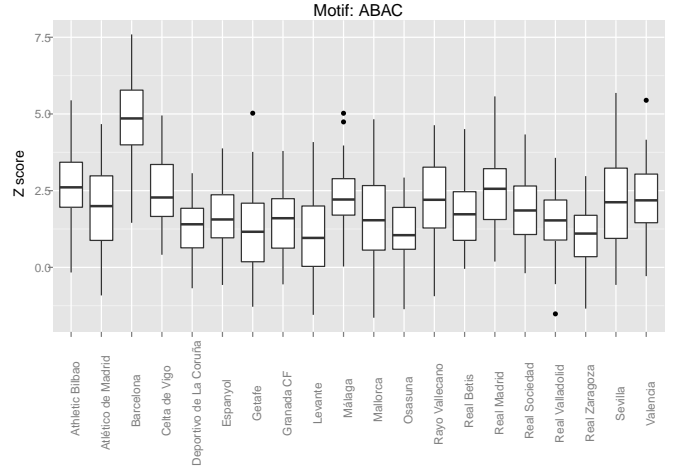


Figure 1: The prevalence of the ABAC motif in case of the teams of the Spanish first division (median, quartiles) with respect to their z-scores. FC Barcelona applies the ABAC motif much more frequently than any other team in the league.

where T_{\max} denotes the time threshold between two passes. These constraints assure that the passes are consecutive (i.e., a player receives the ball and then passes it forward) and not having major breaks. Throughout our study, we use $T_{\max} = 5\text{sec}$ to determine if two passes are belonging to the same ball possession.

Third, we extract all the three-pass long sub-possession from the ball possessions (e.g., a ball possession having n passes contains $n - 2$ motifs) and convert the player identifiers into the appropriate A, B, C, and D labels to assemble the motifs. For example, a ball possession where the ball moves between players as $2 \rightarrow 4 \rightarrow 5 \rightarrow 6 \rightarrow 4 \rightarrow 6$ translates into three motifs, namely, ABCD, ABCA, and ABCB:

$$\begin{array}{c} \overbrace{2 \rightarrow 4 \rightarrow 5 \rightarrow 6 \rightarrow 4 \rightarrow 6}^{ABCA} \\ \underbrace{\hspace{1.5cm}}_{ABCD} \end{array}$$

After having the motifs that are present in the pass network, we quantify the prevalence of the motifs by comparing the pass network of the team to random pass networks having identical properties (in particular, the number of vertices and their degree distribution). Specifically, we perturb the labels of the motifs prevalent in the original pass network randomly and such we create pseudo motif-distributions. In our data analyses, we generate 1000 random pass networks for each original pass network. Finally, we compute the z-scores (a.k.a. standard scores) of the motifs by comparing the original and the constructed random pass networks. As a result, we have a characteristic of the (passing) style of a team for every match—in terms of the z-scores of the motifs.

3. DATA ANALYSIS AND RESULTS

We use publicly accessible information on the pass networks of soccer teams. In particular, the dataset contains information from the 2012/13 seasons of the Spanish, Italian, English, French, and German first division. For example, the part of the dataset that contains information on the

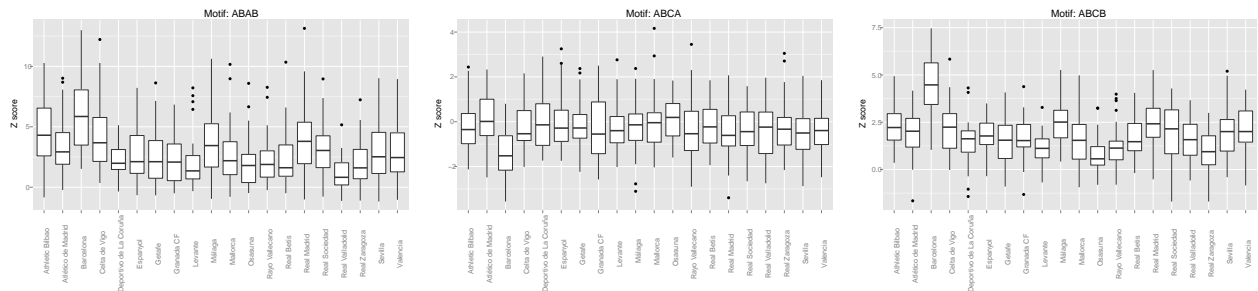


Figure 3: Z-scores of the ABAB, ABCA, and ABCB motif in the Spanish league.

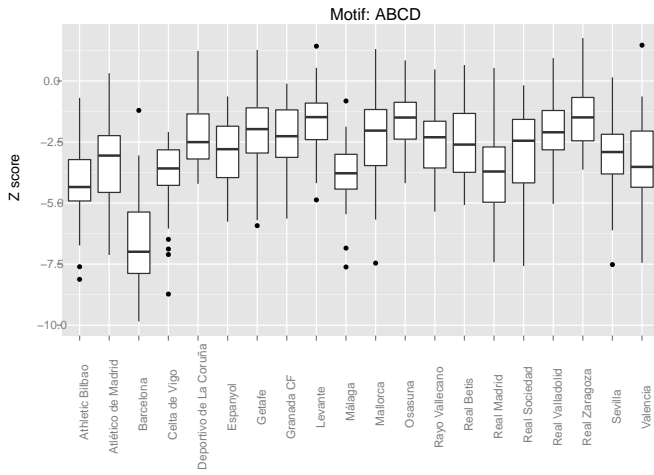


Figure 2: FC Barcelona uses the ABCD motif less often than the other teams.

Spanish league spreads 20 teams, 380 matches, and more than 250 thousands of passes. We quantify the motif characteristics of the teams using the aforementioned dataset. We first present results on the passing styles of teams in the Spanish first division and later on we compare our finding with the other European leagues and teams.

We compare the Spanish teams with respect to their ABAC motifs in Figure 1. Most of the teams have similar z-scores, i.e., apply the ABAC pass motif to comparable extent. However, FC Barcelona has a quite distinct strategy: applies ABAC motifs significantly more often than the other teams (the difference is at least 2.5 standard deviation). The trend is similar in case of the ABCD motif; the only difference is that the majority of the teams have notably larger z-scores than FC Barcelona (Figure 2). This means that FC Barcelona applies this motif significantly less frequently than the other teams. In general, FC Barcelona uses structured motifs (i.e., motifs with more back and forth passes such as ABAB, ABAC, and ABCB) more often than simpler ones compared to other teams. We present the results of the remaining motifs in Figure 3.

We next analyze the similarities and the differences of the teams' motif characteristics via cluster analysis. First, for each team, we construct a feature vector representing the



Figure 4: K-means clustering of the teams in the Spanish league. One of the four clusters contains only a single team, namely, FC Barcelona that has an unique style based on its passing motifs.

team's usage of motifs. We use the mean of the z-scores of the five distinct motifs as the features (by averaging the z-scores over 38 matches a team had in the season). Afterwards, we cluster the teams based on their five-motif long feature vectors. We use two methods for cluster analysis: k-means and hierarchical clustering. We illustrate the result of the k-means clustering in Figure 4 (the clusters are color-coded), where the ratio of the within the cluster and the total sum of squares is 90.3%. For example, the cluster that contains Atletico Madrid and Athletic Bilbao, among others, is characterized by extensive usage of ABAB and ABCA motifs. While most of the teams are clustered in three major groups, FC Barcelona is separated from the other teams. FC Barcelona is the only team in its cluster; hence, it has a distinctive motif characteristics.

The Ward hierarchical clustering algorithm reveals similar trend as shown in Figure 5. Again, FC Barcelona has a solitary style while the other teams are having resembling features. The implications of the two clustering schemes are consistent: FC Barcelona had a unique, significantly different passing style than any other team in the Spanish league.

Finally, we take a broader point of view and investigate whether the style of FC Barcelona remains unique if we con-

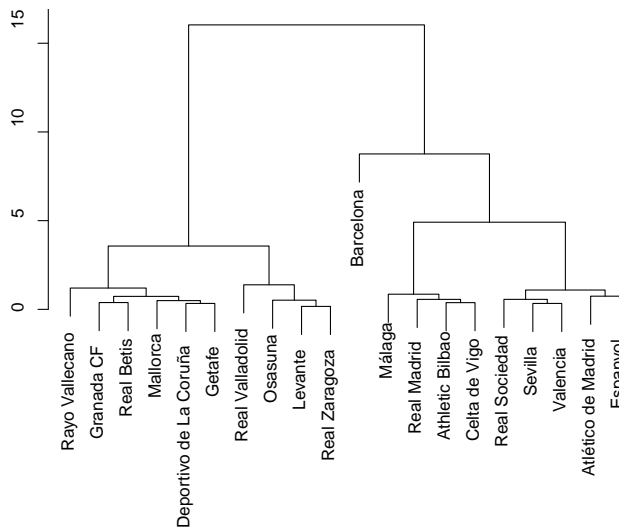


Figure 5: Ward hierarchical clustering of the soccer teams in the Spanish league. FC Barcelona does not belong to any major groups of the teams.

sider teams of four additional European soccer leagues. We show the teams in Figure 6 based on their motifs using principal component analysis. Although we analyze more teams that have more variation in their pass characteristics, FC Barcelona still able to maintain its rare, distinct style. It is surprising that Torino, an Italian team nearly relegated at the end of the season, has a style diverse from the vast majority of the considered teams and shares properties with teams like Lille, Milan, and Juventus—dominant teams in the French and Italian leagues. The distinctive feature of Torino’s strategy is that it involves less frequent usage of the ABCA motif.

4. FUTURE WORK

The presented results illustrate the potential of analyzing the flow motifs of soccer teams. There are several ways to extend the investigation of pass motifs to reveal finer-grained details of teams and players. As future work, we plan to address three areas: (i) condition the pass motifs based on the results of the matches, (ii) study the impact of home and away games on the prevalence of the motifs, and (iii) explore the players’ involvement in the different motifs.

5. CONCLUSIONS

We proposed a quantitative method to evaluate the styles of soccer teams through their passing structures. The analysis of the motifs in the pass networks allows us to compare and differentiate the styles of different teams. Although most teams tend to apply homogenous style, surprisingly, a unique strategy of soccer is also viable—and quite successful, as we have seen in the recent years. Our results shed light on the unique philosophy of FC Barcelona quantitatively: the famous tiki-taka does not consist of uncountable random passes but rather has a precise, finely constructed structure.

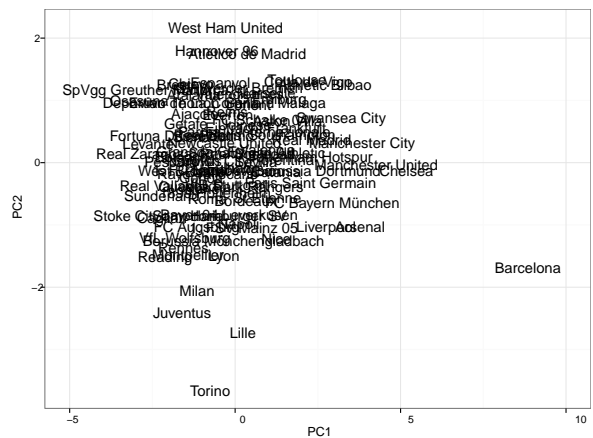


Figure 6: The style of soccer teams of the Spanish, Italian, English, French, and German soccer leagues. FC Barcelona has a unique style even on a European scale.

6. REFERENCES

- [1] C. Anderson and D. Sally. *The Numbers Game: Why Everything You Know about Football is Wrong*. Penguin UK, 2013.
- [2] J. Duch, J. S. Waitzman, and L. A. N. Amaral. Quantifying the performance of individual players in a team activity. *PloS one*, 5(6):e10937, 2010.
- [3] InStat Football. <http://instatfootball.com>. 2014.
- [4] P. Lucey, D. Oliver, P. Carr, J. Roth, and I. Matthews. Assessing team strategy using spatiotemporal data. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1366–1374. ACM, 2013.
- [5] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002.
- [6] T. Narizuka, K. Yamamoto, and Y. Yamazaki. Statistical properties of position-dependent ball-passing networks in football games. *arXiv preprint arXiv:1311.0641*, 2013.
- [7] OptaPro. <http://optasportspro.com>. 2014.
- [8] J. L. Peña and H. Touchette. A network theory analysis of football strategies. *arXiv preprint arXiv:1206.6904*, 2012.
- [9] Prozone. <http://prozonesports.com>. 2014.
- [10] Squawka. <http://www.squawka.com>. 2014.
- [11] The Independent. *The worst ever January transfer signings, including Fernando Torres, Andy Carroll and Eric Djemba-Djemba*. <http://www.independent.co.uk/sport/football/news-and-comment/the-worst-ever-january-transfer-signings-including-fernando-torres-andy-carroll-and-eric-djembadjemba-1858589.html>, 2014.
- [12] J. Wilson. *Inverting The Pyramid: The History of Soccer Tactics*. Nation Books, 2013.