https://doi.org/10.1038/s44260-025-00047-x

Shape graphs and the instantaneous inference of tactical positions in soccer

Ulrik Brandes ✉, Hadi Sotudeh, Doğan Parlak, Paolo Laffranchi & Mert Erkul

We propose shape graphs as instantaneous representations of spatial arrangements in association football (soccer). Shape graphs are a novel type of subgraph of Delaunay triangulations inspired by related applications in fingerprinting and facial expression detection. They provide a foundational data structure that supports various downstream tasks in a manner that is flexible, explainable, and efficient. While previous approaches aggregate spatial positions over periods of time to stabilize the underlying signal, we instead interpret each frame individually for increased explainability at the highest possible temporal resolution. As an example use case we introduce position plots, a novel visualization capturing the characteristic fluidity of relative positioning during a game while retaining the possibility to add match context.

Association football (soccer) is a territorial invasion game in which simple rules lead to complex collective behavior[1]. Consequently, spatial positioning is of paramount importance in football tactics[2]. To be able to coordinate quickly and with limited communication, players assume situational responsibilities for opponents and areas of the pitch. Conventional means to describe such collective behavior include tactical positions with labels such as 'left back' or 'center forward,' and team formations with labels such as '4-4-2' or '3-4-3,' summarizing a split of outfield players into backs, midfielders, and forwards[3].

However, these descriptors are largely static and decontextualized. Players' spatial responsibilities generally differ between match phases and even within possession sequences. They are moderated by opponent behavior, and may change temporarily or permanently following the demands of an unfolding match. Although these dynamics are widely acknowledged, media and match reports commonly content themselves with the average locations of players[4].

We here propose a novel framework to analyze spatial arrangements instantaneously, i.e., at the temporal resolution at which tracking data is available and without using information from other moments in time. Tracking data is utilized at an increasing rate in a variety of team sports[5], including football[6,7]. Our framework integrates and extends previous ideas to describe the dynamic spatial arrangements of teams. Instead of aggregating player locations over periods of time, we analyze each data frame separately to generate time series at the original temporal resolution. From a proximity network imposed onto the set of player locations, we are able to derive tactical positions, team formations, and other aspects of interest for each snapshot and therefore without losing temporal contexts such as match phases.

Our two main contributions are as follows.

- *Shape graphs:* The shape of a team's spatial organization is represented as a combination of structure (networks) and geometry (space) by augmenting each frame of spatial tracking data with a novel kind of proximity network.
- *Instantaneous tactical positions:* By inferring tactical positions from a single shape graph, each snapshot is interpreted separately. This yields instantaneous positions that facilitate the exploration of dynamic positioning in a novel visualization of such high-frequency time series of spatial information.After introducing these concepts in the next two sections, we demonstrate their use on publicly available tracking data.

Team shape

To characterize a team's dynamic spatial arrangement in terms of tactically relevant features such as compactness or defensive line height, it is common to derive time series of geometric descriptors such as centroids, lengths, and areas from tracking data[8,9]. While these are determined on a frame-by-frame basis, more complex features such as player activity or team formations are typically determined from data aggregated over sequences of frames[10–14]. We combine these two approaches and generate high-frequency time series of complex features.

Since players are trained to use other players as spatial references, proximity relations are a natural starting point for the definition of team shape. Delaunay triangulations capture many of these relations in a single structure[15] and are therefore used extensively for shape representation in football[13,14,16], computer graphics, biometry, and many other domains.

A Delaunay triangulation is a planar graph defined on a point set in the plane as exemplified in Fig. 1a. It can be determined efficiently and a characteristic feature is that no circumcircle of a triangle encloses another point. Its graph contains, for instance, the convex hull of the point set, pairwise nearest neighbor relations, and a geometric minimum spanning tree. It also maximizes the minimum angle between incident edges across all triangulations of the point set. Delaunay triangulations are dual to Voronoi

Social Networks Lab, ETH Zurich, Zurich, Switzerland. ✉e-mail: ubrandes@ethz.ch

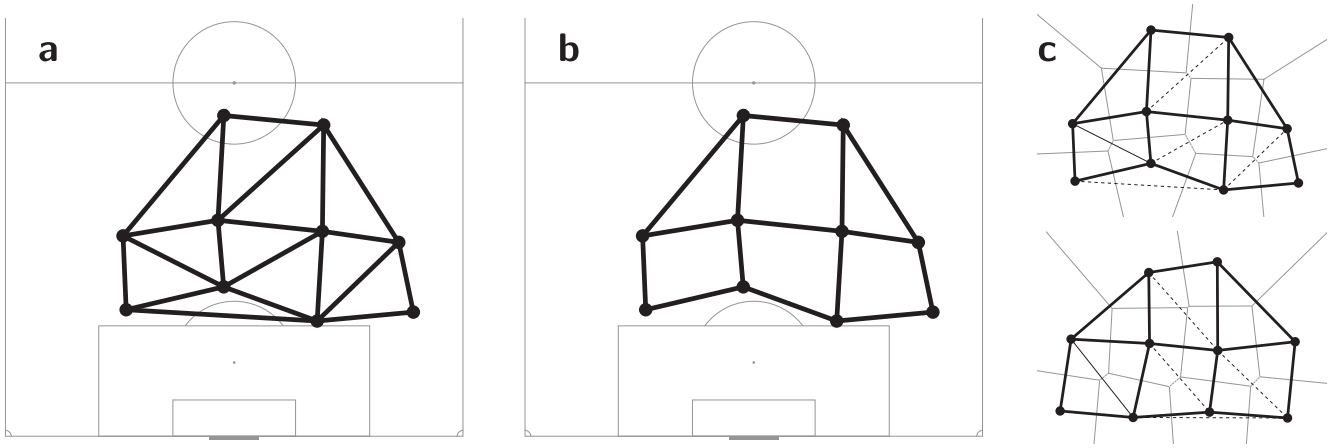


Fig. 1 | Representing the spatial arrangement of a team's ten outfield players in a 4-4-2 formation. We define a shape graph (**b**) as a subgraph of the Delaunay triangulation (**a**) that is more stable against shape-preserving perturbations (**c**). Note that small movements suffice to cause structural changes (dashed) in the Delaunay triangulation and its dual Voronoi diagram (gray).

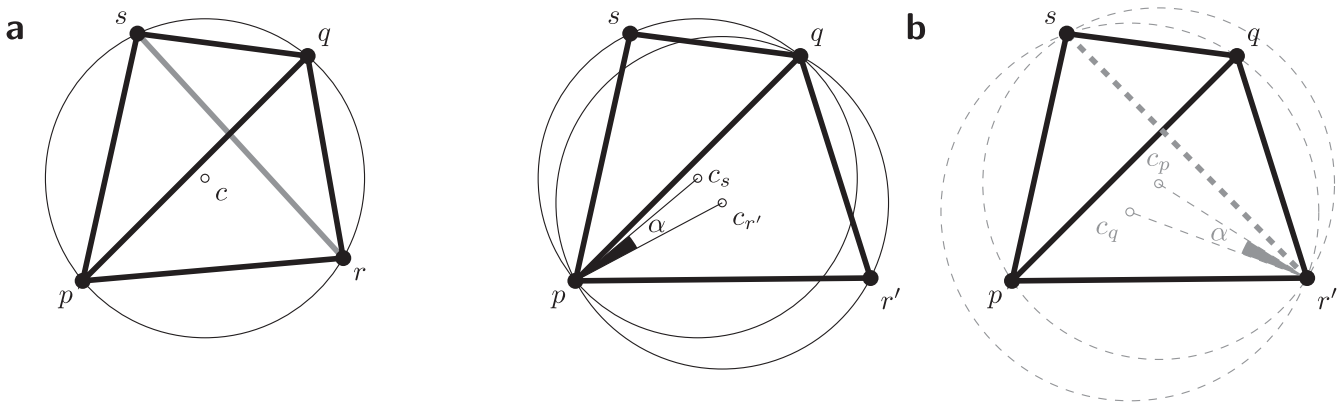


Fig. 2 | Co-circular and near-degenerate point configurations. a For four points on an otherwise empty circle with center c, either diagonal can be in the Delaunay triangulation. **b** After translating $r \to r'$, only pq is a Delaunay edge, because the two circles defined by p, q and one of r', s do not enclose the remaining point, whereas those defined by r', s and one of p, q do. The two pairs of circumcenters are necessarily separated by the same angle α.

diagrams, which were among the first techniques to model what would later be called pitch control[17–20].

Despite their many desirable properties, Delaunay triangulations also have disadvantages, especially in the context of moving objects. As shown in Fig. 2a, alignments of points can lead to ambiguities and thus sensitivity to perturbation.

Among the variety of approaches to measure sensitivity and cope with perturbations[21,22], we found a scale-invariant measure of angular stability[23] particularly suitable for our context. As depicted in Fig. 2b, it is defined as the angle formed by an endpoint of a Delaunay edge and the circumcenters of its two incident triangles. Since circumcenters are the vertices of the dual Voronoi diagram, the angle corresponds to the lengths of dual Voronoi edges in Fig. 1c. Removing all edges with angular stability below some angle α yields a so-called α-stable Delaunay graph. This subgraph, however, is often too sparse, because the joint removal of multiple edges with low stability may reduce connectivity, yield non-convex faces, and thus lead to shape descriptions that do not match the assumptions about spatial orientation of players.

We therefore introduce a new kind of Delaunay subgraph which we refer to as shape graph. A straightforward generalization of angular stability to non-triangular faces allows us to remove a least stable edge, update the stability of all edges in the newly created face, and iterate until no unstable edges remain. We use $\alpha = \frac{\pi}{4} = 45^\circ$ as the threshold and provide computational details in the section on Methods.

Figure 1c illustrates that our shape graphs are less sensitive to shape-preserving movement. Although we here consider only the outfield players of one team in our examples, the approach extends naturally to representations including the goalkeeper, the ball, or the opposition team as additional points.

In the next section, we show how to exploit shape graphs to infer tactical positions of players.

Instantaneous tactical positions

We reserve the term position for the tactical concept, and refer to a point with coordinates in space as a location. For the purpose of this contribution, we further distinguish between player *positions* as spatially, and player *roles* as behaviorally defined concepts. However, we do not aim for a unified terminology, because stakeholders such as coaching staff, data providers, and the media all have their own nomenclature for tactical positions, often mixing positions with roles. We do attempt to stay consistent with the most common terms, however.

A position such as left center back or central attacking midfielder describes where a player is generally expected to play in relation to his or her teammates. A role such as support striker or playmaker describes the actions expected of a player, which may differ even when the player appears in similar locations. Moreover, particular roles such as wing back or false nine can be viewed as combinations of positions assumed in different situations. Note that this differs from the terminology of Bialkowski et al.[10] and subsequent work, where the term role is used for a predominantly spatial concept.

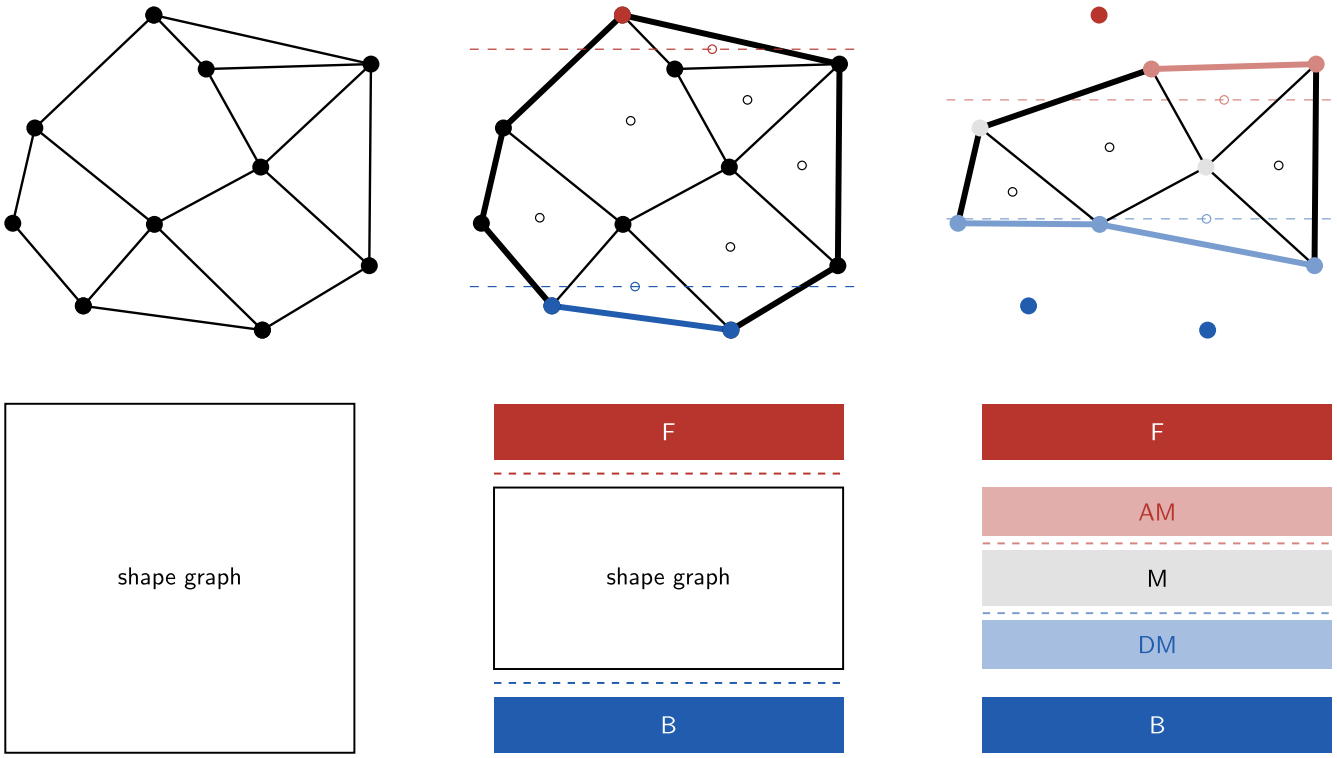
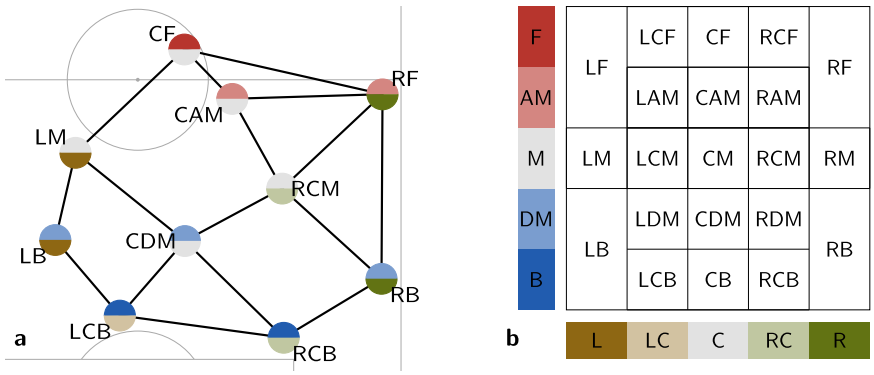


Fig. 3 | Decomposition of a shape graph into vertical positions. Centers of faces (small circles) serve as reference for splitting off the upper and lower parts of the convex hull, and the procedure is repeated once on the remaining midfielders.

Fig. 4 | Player positions are determined from the shape graph of each individual tracking data frame. a Based on structure and geometry, players are assigned color-coded vertical and horizontal levels. **b** Their combination is translated into a tactical position according to a matrix template. Because of their typical positioning patterns, left and right full backs (LB, RB) and wide forwards (LF, RF) are expected to be found in either of two matrix cells.



We propose a simple method to infer positions from shape graphs, which by their very definition ease a number of tasks:

- *Normalization:* The structure of shape graphs is, by definition, invariant under translation and scaling, and therefore robust against collective movements such as shifting, contraction, and expansion that teams use to cover space efficiently. This invariance eliminates the need for transformations common in other approaches[10,11,24].
- *Identity:* When determined separately for each frame of tracking data, the structure and geometry of shape graphs is independent of player identities. Since we infer positions from the geometric graph only, with no regard for who a node represents, no mutual constitution of players and positions through matching[10,12] is needed.
- *Orientation:* Finally, the space we are concerned with is oriented, from goal line to goal line and from touchline to touchline. Following the logic behind common terminology, and thus to avoid cognitive dissonance, we use the convention of pitches oriented vertically, with the focal team playing upward. This terminology also suggests that the assignment of position labels can be separated into a horizontal (left, central, right) and a vertical (backs, midfielders, forwards) component.

For all these advantages, our method is rather simple (see Fig. 3):

- Split off two levels at the extreme ends of each dimension, and
- repeat the procedure once to further discriminate central players.

Splits are determined independent of scale by using the centers of mass for each internal face as references. The two extreme centers define the dashed horizontal levels at which players are separated in Fig. 3. The procedure is applied analogously to the horizontal component, and our handling of edge cases is described in the Methods section.

As a result, we obtain up to five ordinal levels horizontally and vertically, or a 5×5-matrix of relative positions, which may be labeled, for instance, as shown in Fig. 4.

Results

The purpose of this section is two-fold: first, we aim to build trust in our method to detect meaningful positions, and second, we exemplify how it can

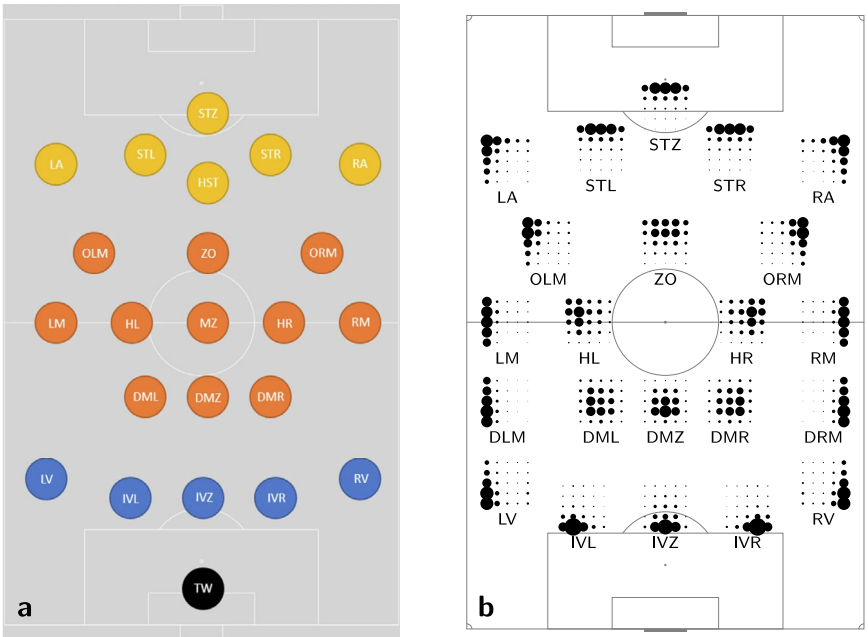


Fig. 5 | Comparison with position labels provider data. Position labels reproduced from Bassek et al.[25] (**a**) and the relative frequency by which our method assigns players with these labels to a cell in the 5 × 5 position matrix (**b**). No player in the data set is labeled HST (support striker) or MZ (central midfielder), but additional labels DLM, DRM (defensive left/right midfielder) are present.

be used in match and opposition analysis. For concreteness and replicability, we present results obtained for some publicly available tracking datasets.

Validity of position labels

There is no single concept of tactical position, as many coaches and analysts have their own approaches, players interpret positions differently, and the fluidity of the game leads to varying observable outcomes even if the underlying principles are the same. Hence, differences between any two means of assigning positions are inevitable.

The quantitative validation of our shape-graph derived labeling should therefore not be understood as an attempt to measure the accuracy of predictions, but as a demonstration that consistently assigned labels from a major commercial data provider yield consistent and distinguishable patterns of labels in our method.

Bassek et al.[25] recently published data from seven matches of the 2022/2023 season in the first and second division of German professional men's football. For each of these matches, the data include optically tracked locations at 25 frames per second (roughly 143,000 frames per match) and provider-assigned position labels as shown in Fig. 5a. To reduce noise and small-sample effects, we double the size of the data as follows. Assuming that there should be no systematic bias in lateral positioning, we add a copy of each frame mirrored across the longitudinal axis (from goal to goal), with players relabeled accordingly (LM to RM, DMR to DML, etc.). This yields between three and five match occurrences for positions LA/LR, LM/RM, DML/DMR, IVZ, and DMZ, and at least 16 for all others (except HST and MZ, which are not present in the data). Moreover, we only consider the 176 players who started a match or were substituted into the game no later than the 75th minute.

For about one million frames during which the ball is in play we determine our instantaneous horizontal and vertical levels. Level combinations of players with the same label in the data set are aggregated. After normalization, we obtain relative frequencies by which players with the same given label are assigned to each of the 25 matrix cells. Results are depicted in Fig. 5b, where the area of each disc is proportional to the share of frames in which the corresponding levels were assigned.

The evident patterns in Fig. 5b suggest that the positioning of players relative to their teammates as determined by our method is strongly associated with the position labels provided in the data. To test the association further, these distributions are used as prototypes for a simple classifier. For each of the seven matches we classify every player who featured by the 75th minute. If $Q = (q_{ij})_{1 \le i,j \le 5}$ is the distribution of level combinations for any one appearance of a player, it is assigned to the prototype $P = (p_{ij})_{1 \le i,j \le 5}$ from Fig. 5b that is most similar according to the Bhattacharyya coefficient $BC(P, Q) = \sum_{i=1}^{5} \sum_{j=1}^{5} \sqrt{p_{ij} q_{ij}}$. Figure 6a confirms that this classification matches the labels provided in the data in more than two-thirds of the cases (122 out of 176). Moreover, a metric multi-dimensional scaling of Hellinger distances $\sqrt{1 - BC(P, Q)}$ between positional distributions suggests that they cluster well and in alignment with given position labels.

Inspection of mismatches and clustering outliers revealed a few common scenarios suggesting that it is often the labels provided in the data, which are assigned before the start of the match[25], that are implausible.

- Players expected to be fielded on one side actually play on the other side. An example shown in Fig. 7a is Lukas Schleimer of 1. FC Nürnberg in a match against Fortuna Düsseldorf on October 15, 2022. Although commonly a striker and labeled as right center forward (STR), he actually played on the left in a 4-2-2-2 after coming on in the 65th minute.
- Positions may change during a match, which necessarily yields distributions that do not fall squarely into a single category. The example in Fig. 7b is Edmond Tapsoba of Bayer Leverkusen in a match against VfL Bochum on May 27, 2023. Labeled and starting as a left center back (IVL), he moves to the right side when the team changes from four to three at the back at half-time.
- Positioning may depend on ball possession. The example in Fig. 7c is Dawid Kownacki of Fortuna Düsseldorf in a match against Hansa Rostock on September 10, 2022. Labeled as the center forward (STZ) in a 4-2-3-1 formation, he tended more to the left when his team was out of possession.

Match analysis example

For an example analysis of position dynamics, we use better known data from a match that has been available for some time. It is the most recent in a small compilation published by Metrica[26], and contains anonymized event and tracking data for an unspecified match referred to as Sample Game 3. The data was collected at a rate of 25 frames per second for a total of 143,761 frames (close to 96 min), each containing location information for all players and the ball.

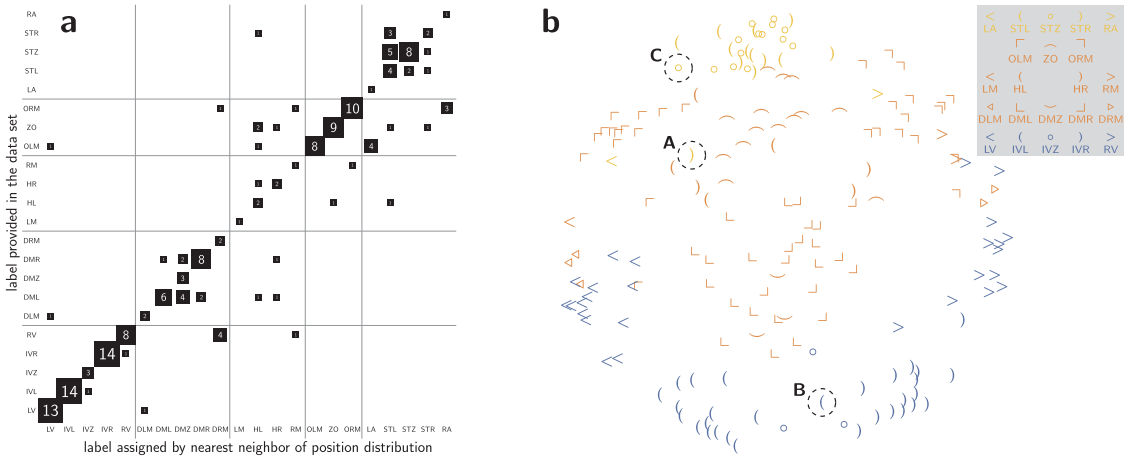


Fig. 6 | Comparison of given position labels with our position distributions for 176 player-match instances. a A classifier assigning the nearest prototype from Fig. 5b largely recovers the given position labels. **b** The scatter plot is a two-dimensional representation of similarity among position distributions associated with given position labels. Circled outliers are discussed in Fig. 7.

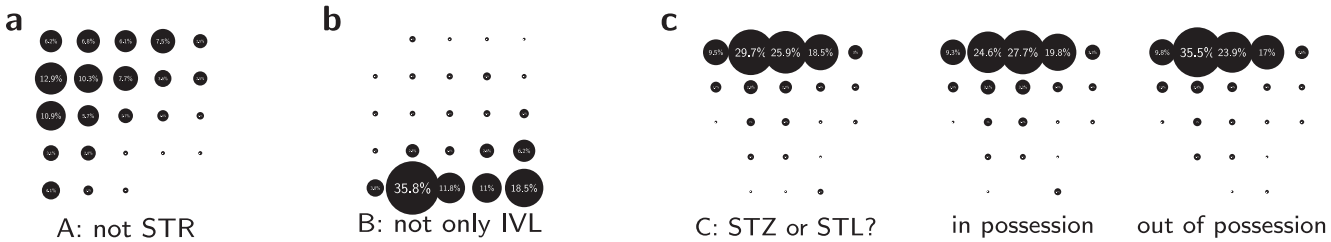
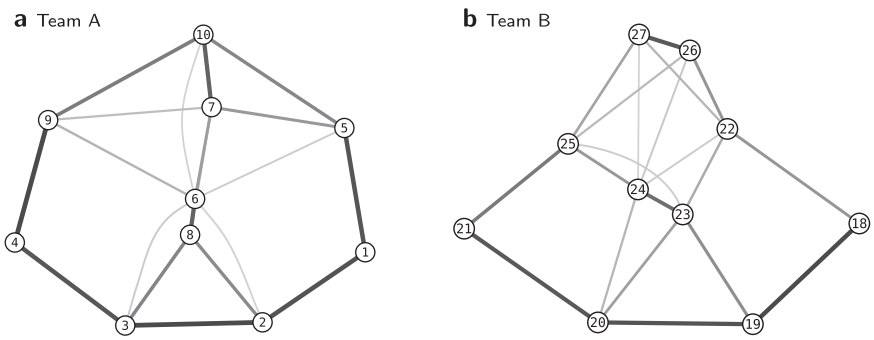


Fig. 7 | Examples of mismatches. Position distributions of the examples highlighted in Fig. 6b suggest that their labels provided with the tracking data are not suitable. While **a** is simply a misclassification, **b** results from a position change during the match, and **c** from the overlay of in- and out-of-possession phases.

Fig. 8 | Frequent shape graph edges. Edges appearing at least 20,000 times (i.e., in more than one third of all frames) during the first half of the Sample Game 3. Thicker and darker edges appear more often, with the maxima for both teams above 50,000 times. Players are placed in their average locations. **a** A single center forward and few edges between wingers indicate stricter positioning. **b** Four attackers with close average positions and many edges represent a more varied approach.



We used Python package kloppy v3.15[27] to infer times during which the ball is not in play from the event data. After dropping the corresponding frames from the tracking data, $T = 117,448$ frames remained.

Shape graphs were computed separately for the outfield players of each team in each frame. Frequent shape graph edges shown in Fig. 8 are restricted to the first half, because substitutions and tactical changes discussed below lead to a mix of principles in the second half.

Both teams play with a back four (four defenders at the back) and a double pivot (two midfielders in front of the back four). Their higher degree in the graph of frequent proximity edges suggests that Players #6 (Team A) and #24 (Team B) cover a substantially larger vertical range than their partnering defensive midfielders #8 and #23, who are in corresponding average locations but seem to play different roles.

Average locations of Team B give the false impression of a compact attack through the center, which seems different from the wide locations of #5 and #9 in Team A. Since the individual shape graphs are planar, however, crossing proximity edges already hint at frequent lateral movement. Since it is not possible to infer the actual dynamics from such aggregate representations, we next turn to time series of positions.

Using the color scheme from Fig. 4, we visualize the ordinal level pairs for all players as bivariate times series in Fig. 9. For readability and resolution, we aggregate over intervals of six seconds (150 frames) and display the modal level (most frequent color) in each time interval.

These position plots are inspired by visualization designs for collective movement[28], and since the colors have an interpretation that persists across teams and matches, they represent more

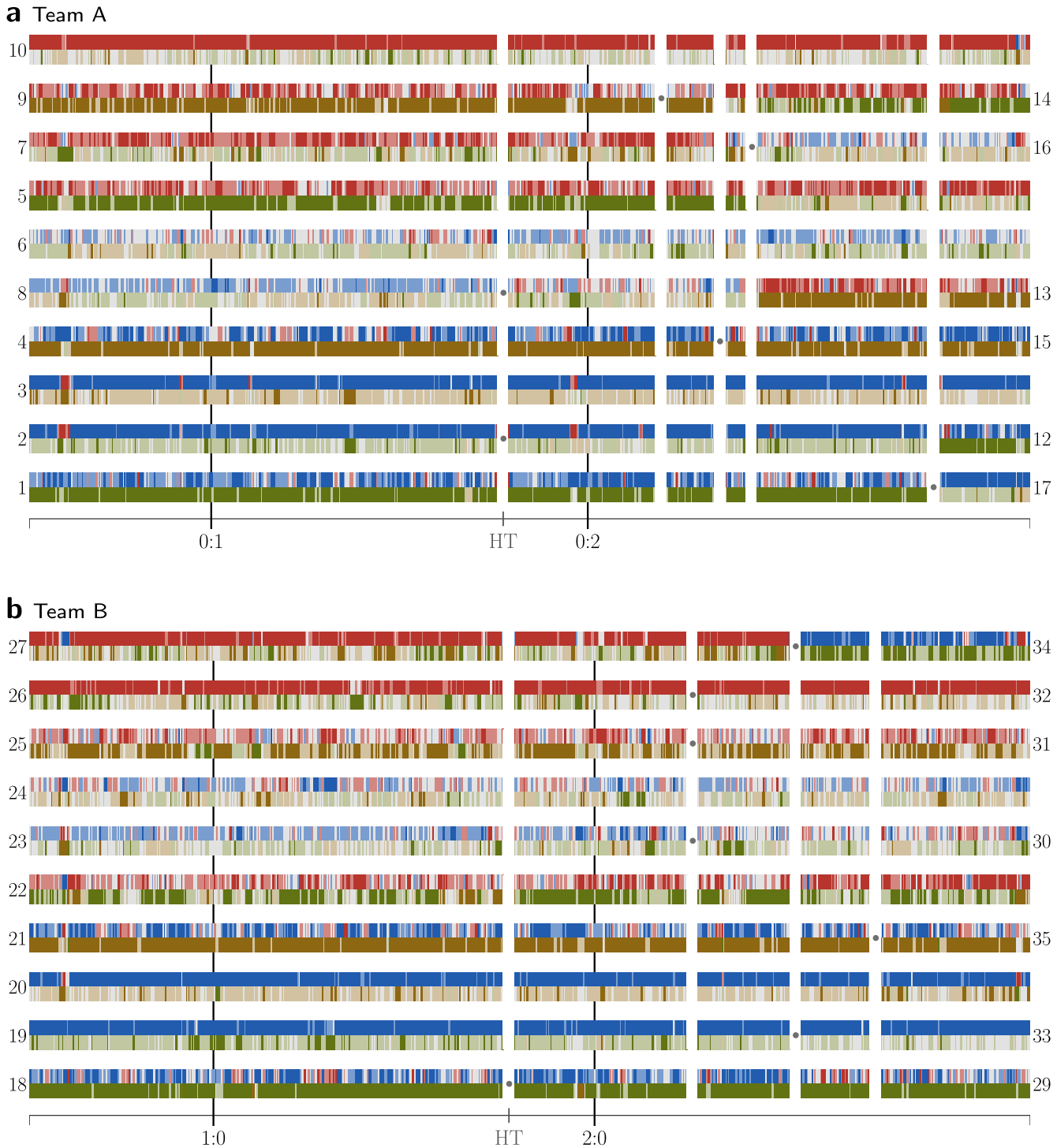


Fig. 9 | Position time series over the entire match. Color pairs represent vertical (upper half of each row) and horizontal (lower half) levels. Gaps indicate substitutions in the team, with dots linking the players involved.

information than seemingly similar depictions of cluster membership in related work[10,11,14].

In Team A, for instance, Player #10 is recognized as a center forward being almost always highest up the pitch (mostly red in the top row) and rarely moving wide of other players (only light colors in the bottom row). Player #12, on the other hand, replaces right center back (RCB) #2 at half-time and plays in the same position until right back (RB) #1 is replaced by another RCB, #17, and #12 moves to RB instead. Categorizing all other players in the same way, we conclude that Team A play out of a 4-2-3-1 formation. With respect to the aggregate plots of Fig. 8, we observe that the wide players rarely move to the other side.

Team B, on the other hand, plays out of a 4-4-2 formation in which the two center forwards and wingers switch sides often. This explains the seemingly compact attacking formation in Fig. 8b, because player rotations visible from the position plots result in more central average locations.

An interesting change of tactics can be observed by combining the two charts. When Team A introduce #16 as LDM for AM #7, #14 rotates right, #5 into the center, and #13, having already played more advanced

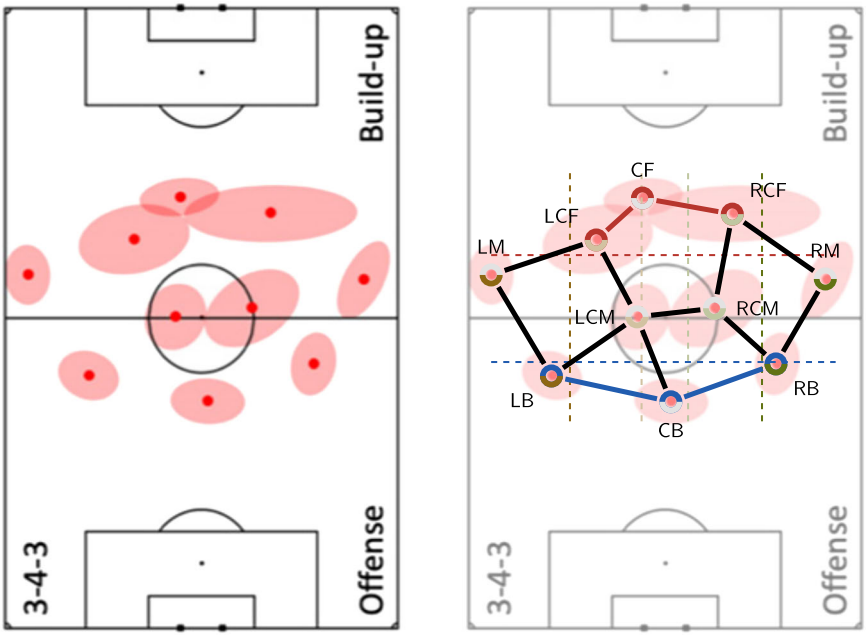


Fig. 10 | Automated labeling of aggregate configurations. Bauer et al.[38] (their Fig. 3) approximate distributions of normalized locations during match phases. With shape graphs defined on their centroids, position labels can be assigned automatically.

than #8 after coming on at half-time, moves to LF. It appears that Team B respond by replacing a CF with RB #34 and moving #29 into positions between the new RB and RM #30. During this final stretch of the match, Team B appear to abandon the 4-4-2 and free up #29 for man-marking #13.

In addition to general trends, coordinated exceptions become apparent. We have already noted that the center forwards of Team B are swapping sides frequently, whereas the wide forwards of Team A are not. A temporary change involving more players can be observed when the center backs of Team A exchange vertical positions with midfielders early in each period, likely to expect a corner kick or a free kick in the opposition half. Similarly, after conceding the second goal, both the full backs LB #4, RB #1) and wide forwards (LF #9, RF #5) of Team A drop deeper until the next substitution is made.

Position plots allow for much more detailed analysis when augmented with match phases, main events, ball location, and other information. Such extensions, however, are beyond the scope of this paper.

Discussion

Shape graphs represent simultaneously the structure and geometry of relative positions and dispense with the need to normalize player locations by translation or scaling. They capture meaningful proximity relationships because players are trained to make use of four different types of spatial references: their teammates, their opponents, the ball, and the pitch[29]. Although the shape graphs presented here make use of the first type only, inclusion of additional references is straightforward. Shape graphs defined on the locations of all outfield players in both teams, for example, would allow for the derivation of time series that represent marking and pressing dynamics[30–32].

Delaunay triangulations of moving points change when passing through ambiguous configurations of zero angular stability, and our criterion for preemptive removal of edges with low angular stability shifts changes to moments in which a third point is close enough to challenge the existence of an edge representing mutual spatial referencing of two players.

As our approach is computationally efficient and instantaneous, it is suitable for use with live tracking data. Future work could include the storage of shape graphs in kinetic data structures[33] for efficiency, and in graph databases for querying moments of interest[34,35].

By deriving them instantaneously from shape graphs, we have separated the detection of positions from their matching with players. Momentary tactical positions of players are inferred from their relative spatial locations. It is therefore not necessary to assume team-level persistence of the set of positions over any period of time. In fact, previous work can be viewed as focusing on the identification of spatial arrangements that are stable over periods of time, and therefore interested just as much in temporal segmentation as in positions and formations. With the template-matching approach of Müller-Budack et al.[12] a notable exception, position labels are generally assigned by hand after positions have been determined.

Our assignment is systematic and interpretable; and it may be applied to any configuration, including the output of aggregation approaches. After the first vertical split in the example of Fig. 10, the shape graph of the four midfielders forms a horizontal path, so that no further distinction takes place. In the horizontal partition, inside players are split equidistantly (see Methods section), but the center forward is close to being interpreted as another left center forward.

On a temporal scale, the tactical tweaks pointed out in the Results section are fairly coarse grained. The time a team in transition needs to return to their default shape, and frequent patterns of positional changes are but two examples of interesting analyses at higher temporal resolution. Instantaneous positions may further be used to contextualize individual events. Passes, for instance, are often labeled with time, players involved, their locations, a pass type, and an outcome, but could be augmented with the tactical positions at this moment.

The position plots in Fig. 9 are but one example how to visualize position time series. They can be varied, for instance, by filtering, changing the level of resolution and annotation with further match events. The latter may include bookings, match phases, momentum, or any other information common for match time-lines and potentially driving positional changes or providing situational context. Our example should therefore not be considered a specific chart design, but a framework to construct static depictions of positional dynamics.

Methods

To complement the introduction of shape graphs and instantaneous positions, we outline in this section some additional details relevant for their implementation.

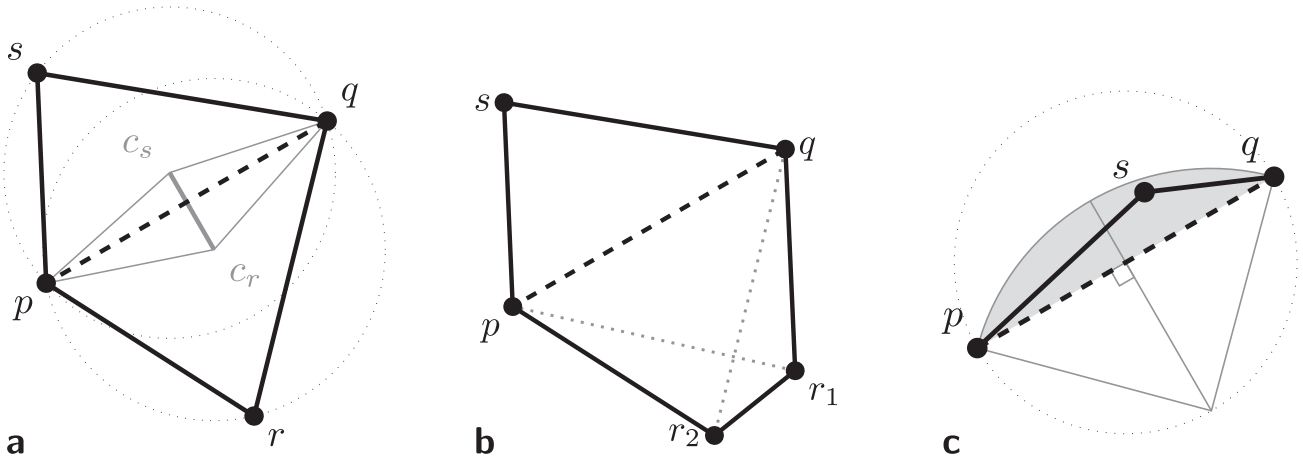


Fig. 11 | Delaunay edges *pq* are removed from the shape graph, if their angular stability is below $\alpha = 45°$. a For an internal edge of the triangulation we test $\angle c_s p c_r = \angle c_r q c_s < 45°$ or, equivalently, $\angle prq + \angle qsp > 135°$. **b** For an internal edge adjacent to a face that is no longer a triangle, the removal condition has to be fulfilled with all of its corners, i.e., for $\min_{i=1,\dots,k} \angle pr_i q$ replacing $\angle prq$. **c** For an edge on the outer face, the outside angle is set to 0°; the removal condition becomes $\angle psq > 135°$ or, equivalently, *s* being inside the gray area.

Shape graphs

Assume we are given a sequence of tracking data frames $P^{(t)} = \{p_1^{(t)}, \dots, p_n^{(t)}\}$ containing two-dimensional spatial locations $p_i^{(t)} \in \mathbb{R}^2$ for n objects at discrete time steps $t = 1, \dots, T$. In the present scenario with 25 Hz tracking and a focus on the outfield players, we have $T \approx 140,000$, and $n = 10$ for each team.

Given a finite point set $P \subset \mathbb{R}^2$ in general position, a Delaunay triangulation is a planar graph $D(P) = (P, E)$ such that every internal face is a triangle whose circumcircle does not enclose another point of P. The circumcenters of those triangles are the vertices of the dual Voronoi diagram.

For $p, q \in P$, consider the oriented line $\vec{pq}$ through p and q, and let $P_{pq} = \{r \in P \backslash \{p, q\} : (q - p) \times (r - p) < 0\}$ be the subset of points in the half-plane to its right. Consequently, P_{qp} contains those in the half-plane to the left, as this is to the right of the reverse line $\vec{qp} = \overleftarrow{pq}$.

Note that the circles through p, q and any third point $r \in P_{pq}$ are nested in the half-plane to the right (and reversely nested on the left), and that an innermost arc belongs to a point with maximum angle $\angle prq$. We therefore define $\alpha(p, q) = \max_{r \in P_{pq}} \angle prq$ and call it the angle on the right of $\vec{pq}$ (and on the left of $\vec{qp}$). If $P_{pq} = \emptyset$, we let $\alpha(p, q) = 0$; this occurs for p, q consecutive in counterclockwise order on the convex hull of P and corresponds to the angle with a point at infinity.

Any four points in general position form a quadrilateral in which the sum of the internal angles is $2\pi = 360°$. If these points are co-circular, the left and right angles of each diagonal each sum to $\pi = 180°$. If points $p, q \in P$ are connected by a Delaunay edge, $\alpha(p, q) + \alpha(q, p) \leq \pi$, and the edge is called $\pi - (\alpha(p, q) + \alpha(q, p))$-stable[23]. Note that this generalizes to edges in the convex hull. As was illustrated in Fig. 2a, edges with zero stability are exactly those for which circumcenters of the two adjacent faces coincide and it is ambiguous whether they are in the Delaunay triangulation.

Geometric properties, stability experiments, and discussions with domain experts led us to set $\frac{1}{4}\pi = 45°$ as a minimum angular stability requirement. Removing all below-threshold edges simultaneously, however, can lead to large faces far from being circular. Therefore, we propose a generalization of angular stability to non-triangular faces and remove edges of minimum stability iteratively to obtain a subgraph that is, in general, more connected than the $\frac{\pi}{4}$-stable subgraph[23] for which edges are filtered in a single sweep. Our generalization sets $\alpha(p, q)$ to the minimum angle with any vertex in the face to the right of $\vec{pq}$. See, for instance, r_1, r_2 in Fig. 11b. An edge is removed only if all vertices in the two faces incident to it are nearly co-circular, i.e., the resulting larger face is circle-like.

Algorithm 1. Shape graph

input : point set $P = \{p_1, \dots, p_n\}$
output: geometric graph $S = (P, E)$

$S \leftarrow$ Delaunay triangulation $D(P) = (P, E)$
$Q \leftarrow \emptyset$ (maximum priority queue)
for $e = \{p, q\} \in E$ **do**
  determine opposing angles $\alpha(p, q), \alpha(q, p)$
  $Q.insert \leftarrow e$ with priority $\alpha(p, q) + \alpha(q, p)$
while $Q \neq \emptyset$ *and* $Q.max > \frac{3}{4}\pi = 135°$ **do**
  $Q.extractMax \rightarrow e = \{p, q\}$
  remove e from E
  $F(p, q) \leftarrow$ newly merged face left and right of $\vec{pq}$
  $p_0, p_1, \dots, p_k \leftarrow$ vertices in $F(p, q)$ clockwise
  for $i = 1, \dots, k$ **do**
    if *right of* $\overrightarrow{p_{i-1}p_i}$ *is external face* **then**
      $\alpha(p_{i-1}, p_i) \leftarrow 0$
    else
      $\alpha(p_{i-1}, p_i) \leftarrow \min_{j=1,\dots,k-2} \{\angle p_{i-1} p_{(i+j \mod k)} p_i\}$
    $Q.update(\{p_{i-1}, p_i\}, \alpha(p_{i-1}, p_i) + \alpha(p_i, p_{i-1}))$

The shape graph of a set of n points is determined by Algorithm 1, where angles associated with edges in a newly merged face are updated after each removal. Worst-case running time is therefore $\mathcal{O}(n^2)$, as evidenced by any convex point set P, but in applications to football we typically have $n \leq 23$ (all players and the ball) and only few removals.

While essentially absent from match data, ties in least angular stability do occur in textbook examples of spatial arrangements. If two edges are tied for minimum stability and the removal of one leads to a stability update that lets us keep the other edge, we in fact keep both to maintain symmetry.

It is important to note that shape graphs are geometric graphs[36], i.e., combinations of network structures and spatial information. In addition to the combinatorial structure of player proximity, they also represent geometric properties of that structure, and therefore quantitative spatial information that can be organized in variables associated with the structural elements. Examples include the area of faces or the orientation of edges relative to the direction of play. This representation is lossless, and since the

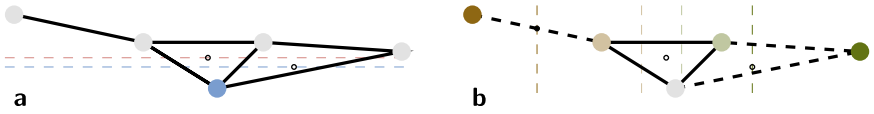


Fig. 12 | Special cases in the assignment of position levels. a Centered split. **b** Equidistant secondary split.

additional memory needed for their storage is linear in the number of points, the overhead is negligible.

Tactical positions

Splits obtained with face centers as reference are tolerant to shape changes in other parts of the team. The latter present difficulties for approaches based on pairwise distances and clustering[24,37].

When multiple players align on circles or straight lines, however, shape graphs may contain too few faces and potentially even bridging edges in the outer face. To ensure robust and meaningful assignment of positions, the following exceptions are introduced in our two-tier decomposition from above:

- If the shape graph contains steep bridging edges, the edges themselves are treated as faces, so that their center is in the middle of the edge. The slope of the bridging edge in Fig. 12 aligns with the vertical partition in 12a, but defines threshold for the first horizontal split in 12b.
- If a group of at least four vertices is maximally imbalanced, i.e., divided into a singleton, an empty center, and the rest, then the large group is assigned to the central partition. Otherwise, a configuration such as the one in Fig. 12a, which resembles the midfield of a 4-1-4-1 formation, would have four attacking midfielders.
- If all centers of internal faces and at least one vertex are in the middle third between the highest and lowest vertex, we split into equal thirds instead. See the secondary split in Fig. 12b for example.

Data availability

The data used in this article was made publicly available in Metrica (2021) and Bassek et al. (2025).

Received: 30 December 2024; Accepted: 5 June 2025;
Published online: 04 August 2025

References

1. Teoldo, I., Guilherme, J. & Garganta, J. *Football Intelligence: Training and Tactics for Soccer Success* (Taylor & Francis, 2022).
2. Wilson, J. *Inverting the Pyramid: The History of Soccer Tactics* (Nation Books, 2013).
3. Sotudeh, H. The principles of tactical formation identification in association football (soccer) – a survey. *Front. Sports Active Living* **6**, 1512386 (2025).
4. Whitmore, J. & Seidl, T. Shape analysis: automatically detecting formations. https://web.archive.org/web/20210301192414/https://www.statsperform.com/resource/shape-analysis-automatically-detecting-formations/ (2021).
5. Torres-Ronda, L., Beanland, E., Whitehead, S., Sweeting, A. & Clubb, J. Tracking systems in team sports: a narrative review of applications of the data and sport specific analysis. *Sports Med. Open* **8**, 15 (2022).
6. Memmert, D. & Raabe, D. *Data Analytics in Football: Positional Data Collection, Modelling and Analysis* (Taylor & Francis, 2024).
7. Andrienko, G. et al. Constructing spaces and times for tactical analysis in football. *IEEE Trans. Visual. Comput. Graph.* **27**, 2280–2297 (2021).
8. Bartlett, R., Button, C., Robins, M., Dutt-Mazumder, A. & Kennedy, G. Analysing team coordination patterns from player movement trajectories in soccer: methodological considerations. *Int. J. Perform. Anal. Sport* **12**, 398–424 (2012).
9. Low, B. et al. A systematic review of collective tactical behaviours in football using positional data. *Sports Med.* **50**, 343–385 (2020).
10. Bialkowski, A. et al. Large-scale analysis of soccer matches using spatiotemporal tracking data. In *IEEE International Conference on Data Mining*, 725–730 (Institute of Electrical and Electronics Engineers, 2014).
11. Shaw, L. & Glickman, M. Dynamic analysis of team strategy in professional football. In *Barça Sports Analytics Summit* (Barça Innovation Hub, 2019). https://barcainnovationhub.com/event/barca-sports-analytics-summit-2019/.
12. Müller-Budack, E., Theiner, J., Rein, R. & Ewerth, R. "Does 4-4-2 exist?": an analytics approach to understand and classify football team formations in single match situations. In *Proc. 2nd International Workshop on Multimedia Content Analysis in Sports*, MMSports '19, 25–33 (Association for Computing Machinery, 2019).
13. Narizuka, T. & Yamazaki, Y. Clustering algorithm for formations in football games. *Sci. Rep.* **9**, 13172 (2019).
14. Kim, H., Kim, B., Chung, D., Yoon, J. & Ko, S.-K. SoccerCPD: formation and role change-point detection in soccer matches using spatiotemporal tracking data. In *Proc. 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '22, 3146–3156 (Association for Computing Machinery, 2022).
15. Aurenhammer, F., Klein, R. & Lee, D.-T.*Voronoi Diagrams and Delaunay Triangulations* (World Scientific, 2013).
16. Raabe, D., Biermann, H., Bassek, M., Memmert, D. & Rein, R. The dual problem of space: Relative player positioning determines attacking success in elite men's football. *J. Sports Sci.* **42**, 1821–1830 (2024).
17. Kawashima, Yoshino & Aoki. Qualitative image analysis of group behaviour. In *1994 Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 690–693 (Institute of Electrical and Electronics Engineers, 1994).
18. Taki, T., Hasegawa, J. & Fukumura, T. Development of motion analysis system for quantitative evaluation of teamwork in soccer games. In *Proc. 3rd IEEE International Conference on Image Processing*, Vol. **3**, 815–818 (Institute of Electrical and Electronics Engineers, 1996).
19. Spearman, W. Quantifying pitch control. In *OptaPro Analytics Forum*. https://doi.org/10.13140/RG.2.2.22551.93603 (2016).
20. Fernández, J. & Bornn, L. Wide open spaces: a statistical technique for measuring space creation in professional soccer. In *MIT Sloan Sports Analytics Conference* (MIT Sloan School of Management, 2018). https://www.sloansportsconference.com/research-papers/wide-open-spaces-a-statistical-technique-for-measuring-space-creation-in-professional-soccer.
21. Abellanas, M., Hurtado, F. & Ramos, P. A. Structural tolerance and Delaunay triangulation. *Inform. Process. Lett.* **71**, 221–227 (1999).
22. Liang, X., Bishnu, A. & Asano, T. A robust fingerprint indexing scheme using minutia neighborhood structure and low-order delaunay triangles. *IEEE Trans. Inform. Forensics Secur.* **2**, 721–733 (2007).
23. Agarwal, P. K. et al. Stable delaunay graphs. *Discrete Comput. Geometry* **54**, 905–929 (2015).
24. FIFA High Performance, T. S. G. Enhanced Football Intelligence. https://www.fifatrainingcentre.com/media/native/world-cup-2022/Enhanced%20Football%20Intelligence%20EN.pdf (2022).
25. Bassek, M., Rein, R., Weber, H. & Memmert, D. An integrated dataset of spatiotemporal and event data in elite soccer. *Sci. Data* **12**, 195 (2025).
26. Metrica. Metrica Sports Sample Data (2021). https://github.com/metrica-sports/sample-data/tree/master/data/Sample_Game_3.

27. PySport. kloppy: standardizing soccer tracking and event data. https://kloppy.pysport.org/ (2024).
28. Buchmüller, J. F., Schlegel, U., Cakmak, E., Keim, D. A. & Dimara, E. SpatialRugs: a compact visualization of space and time for analyzing collective movement data. *Comput. Graph.* **101**, 23–34 (2021).
29. Escher, T. *Der Schlüssel zum Spiel: wie moderner Fußball funktioniert* (Rowohlt Taschenbuch Verlag, 2020).
30. Buldú, J. M. et al. Football tracking networks: Beyond event-based connectivity. In *Analytics in Sports Tomorrow* (Barça Innovation Hub, 2020). https://arxiv.org/abs/2011.06014.
31. Chacoma, A., Billoni, O. V. & Kuperman, M. N. Complexity emerges in measures of the marking dynamics in football games. *Phys. Rev. E* **106**, 044308 (2022).
32. Andrienko, G. et al. Visual analysis of pressure in football. *Data Mining and Knowledge Discovery* **31**, 1793–1839 (2017).
33. Guibas, L. Kinetic Data Structures. *Handbook of Data Structures and Applications*, 377–388 (Chapman and Hall/CRC, 2017).
34. Sha, L. et al. Chalkboarding: a new spatiotemporal query paradigm for sports play retrieval. In *Proc. 21st International Conference on Intelligent User Interfaces*, IUI '16, 336–347 (Association for Computing Machinery, 2016).
35. Stein, M. et al. From movement to events: improving soccer match annotations. In Kompatsiaris, I. et al. (eds.) *MultiMedia Modeling*, Lecture Notes in Computer Science, 130–142 (Springer International Publishing, 2019).
36. Pach, J. Geometric Graph Theory. In Preece, D. A. & Lamb, J. D. (eds.) *Surveys in Combinatorics, 1999*, vol. 267 of *London Mathematical Society Lecture Note Series*, 167–200 (Cambridge University Press, 1999).
37. Fernández de la Rosa, J. *A framework for the analytical and visual interpretation of complex spatiotemporal dynamics in soccer*. Ph.D. Thesis, Universitat Politécnica de Catalunya.http://www.tdx.cat/handle/10803/673529 (2022).
38. Bauer, P., Anzer, G. & Shaw, L. Putting team formations in association football into context. *J. Sports Anal.* **9**, 39–59 (2023).

Acknowledgements

Our method was developed on data provided by UEFA and FIFA for research purposes. We have benefited from feedback and discussions with professional match analysts, most notably Kevin Ehmes, Timo Gross, and Yannick Herkommer. Two reviewers provided constructive feedback that helped improve both content and exposition. This research was supported through ETH Grant ETH-012 21-2.

Author contributions

U.B. conceptualized the research. All authors contributed to the specifics of the methods and implemented various test. H.S. implemented a consolidated version and produced the results. U.B. wrote the manuscript text and prepared the figures; U.B. and H.S. revised the manuscript.

Funding

Open access funding provided by Swiss Federal Institute of Technology Zurich.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Ulrik Brandes.