A model of possession for collective sports

Working paper

Camille Grange, Rodolphe Charrier, Géraldine Del Mondo, Antoine Dutot, Eric Sanlaville

March 2025

Contents

1 Context

In sport science, different models exist for representing human beings, leading to different hypothesis of skill acquisition for them. According to the model used and the resulting type of skill acquisition, a coach will apply an appropriate pedagogical approach. For example, if we consider the human as a *linear system*, we hypothesize that learning occurs in a linear fashion. In this case, the coach's pedagogy will involve presenting a specific situation to the athlete, explaining how to tackle it, and then having him practice. Throughout this document, we refer to this approach as *prescriptive* pedagogy. On the other hand, if we view the human as a *complex system* (eco-dynamic system), we assume that skill acquisition is non-linear. In this case, the coach's pedagogy will focus on conveying general principles, which are not specific to any situation, and encouraging the athlete to practice while adapting these principles to different contexts. We refer to this second approach as *auto-organized* pedagogy.

Let us consider a collective sport opposing two teams of n players each, playing with a ball (or any similar portable object), where players of the same team can interact by passing the ball from one player to another. Specifically, the ball carrier can either move with the ball or pass it to another player. The aim of each team is to bring the ball to a target zone. We refer to a possession as a sequence of consecutive passes between the same team, until the ball is lost or the aim is reached (ball into the target zone). In what follows, for a given possession, we call the team carrying the ball the *attack* team, and the other one the *defense* team, which goal is to prevent the attack team from bringing the ball to the target zone.

In this work, we first propose a general graph-based representation of the spatio-temporal phenomenon that is a possession in a collective sport. More precisely, we propose a model of a possession based on graph theory, taking in consideration both spatial and temporal information. Indeed, in sport analysis, the spatial information is often omitted, resulting in studies of passing networks for instance. Thus, this work integrates in the model the spatial information as the temporal evolution. In addition to tackle this spatio-temporal modelization, scarce in the sport analysis literature, this work is also motivated by presenting a model able to differentiate the two pedagogies introduced above.

Specifically, the second goal of this work is to apply the proposed model to sport data and to assess its ability to classify (and even predict?) two types of attacking team's behavior, one under auto-organized pedagogy, and the other under prescriptive pedagogy. The sport data available have been collected from the following protocol we describe roughly, for both rugby and basketball. We are given two attack teams, and one defense team that can take three different initial position on the field. On the one hand, the coach applies prescriptive pedagogy to the first attack team, namely he provides detailed instructions for each initial position of the defense. On the other hand, the coach applies auto-organized pedagogy to the second attack team, providing general principles without tackling the three specific defense positions of the protocol. Then, we observe the possessions made by the two attack teams on the same set of defense initial positions (we repeat several times the three possible initial positions).

The rest of the document is structured as follows. In Section 2, we present the generic graph-based model of a possession, and in Section 3, we apply it to rugby and basketball. In Section 4, we propose different measures that can be evaluated from the model. In Section 5, we present preliminary results on rugby data.

2 Description of the model

In this section, we define a graph-based model that contains the following main information:

- Spatial state of the game (spatial information):
  - Absolute spatial position of the ball carrier.
  - Some relative spatial positions of players.
- Evolution of the game (temporal information):
  - Spatial changes.
  - Thematic changes.

Let us detail below the nature of these information. In what follows, we present the modelization of information related to the attack team only. Notice that the information of the defense team in the model could be easily integrated.

2.1 Spatial and temporal information

2.1.1 Spatial information

We define a *zone* as a connected bounded part of the Euclidean space of 2 dimension. Let A be a zone and $(B_i)_{i\in\mathcal{I}}$ be a finite set of zones. We say that $(B_i)_{i\in\mathcal{I}}$ is a partition of A if

$$\bigcup_{i\in\mathcal{I}} B_i = A \quad \text{and} \quad \bigcap_{i\in\mathcal{I}} B_i = \emptyset\,.$$

Henceforth, we consider the field as a zone called F.

Absolute spatial position. We partition F into m_{abs} zones $(A_i)_{i\in[m_{\text{abs}}]}$. This partition does not change over time and represents an absolute reference over the field. We call each A_i an *absolute* zone. We note

$$\mathcal{A} = \{A_1, \dots, A_{m_{\text{abs}}}\}$$

the set of absolute zones. For a given instant time, we define the absolute spatial position of the game as $A \in \mathcal{A}$. The absolute position indicates in which absolute zone is the ball carrier at this instante time.

Relative spatial position. For a given instant time $t \in \mathbb{R}^+$, we partition the zone $F \setminus \{pos(t)\}$, where $pos(t) \in F$ is the position of the ball carrier, into m_{rel} zones $(R_j(t))_{j\in[m_{\text{rel}}]}$. Thus, this partition evolves over time according to the position of the ball carrier. We call each $R_j(t)$ a *relative* zone. Notice that the function that splits the field into m_{rel} zones is the same for any position $pos(t)$, but the resulting splitting will differ according to the ball carrier position. Thus, for a given instant time, we define the relative spatial position of the game as a tuple $(N_1, N_2, \dots, N_{m_{\text{rel}}})$ such that $\sum_{j=1}^{m_{\text{rel}}} N_j = n-1$ and $N_j \in \mathbb{N}, \forall j \in [m_{\text{rel}}]$. The relative position indicates the number of players (other than the ball carrier) in each relative zone. Precisely, N_j is the number of players in zone $R_j(t)$. We note

$$\mathcal{R} = \{(N_1, \dots, N_{m_{\text{rel}}}) : \sum_{j=1}^{m_{\text{rel}}} N_j = n-1, N_j \in \mathbb{N}\}$$

the set of all possible relative positions.

2.1.2 Temporal information

In our model, we consider temporal information as a change of spatial state or a thematic change. The spatial state of the game, absolute plus relative, is defined above. Moreover, we define a thematic change as a element in $\mathcal{TC} = \{tc_1, \dots, tc_k\}$, for $k \in \mathbb{N}$. A thematic change represents a change that is different from a spatial change. Thus, it depends on the nature of the collective sport we tackle. For instance, for rugby, we can consider a back pass, a foot pass or even the action of tackling a player as thematic changes. For basket, we can consider the following thematic changes: an hand-to-hand pass, a regular pass or a screen to the ball carrier.

Next, we explain how spatial and temporal information are expressed in our model.

2.2 Paths on a skeleton graph

A possession is represented as a labeled path on a skeleton graph. Note that the skeleton graph is defined *a priori*, i.e. without observing any possession.

2.2.1 Skeleton graph

Let us consider the fully-connected non-oriented graph with the vertices corresponding to all the elements in $\mathcal{A} \times \mathcal{R}$. In other words, we consider the clique (with self-loops on each vertex) where a vertex is a couple of an absolute and a relative spatial position. Notice that each edge represent a possible transition from a spatial state of the game to another one. In specific collective sports, some edges could not be considered if not representing a possible transition. We add to the clique two vertices, "Target" and "¬Target" that will indicate if the ball ends in the target zone or not. Each of the two vertices are connected to all vertices in $\mathcal{A} \times \mathcal{R}$. We call $\mathcal{K}$ this skeleton graph. Notice that the number of vertices of $\mathcal{K}$ is upper-bounded by $(m_{\text{abs}} n^{m_{\text{rel}}-1} + 2)$. In Figure 1, we provide an example of the skeleton graph for the case of $n = 2$ players, $m_{\text{rel}} = 2$ relative zones and $m_{\text{abs}} = 2$ absolute zones.

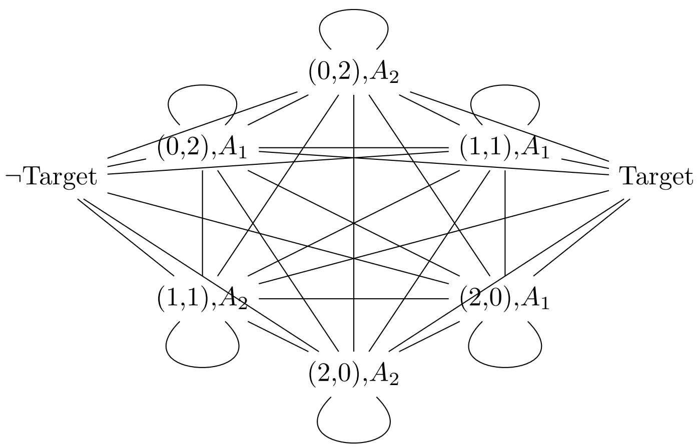


Figure 1: Example of the skeleton graph for $n = 2$, $m_{\text{rel}} = 2$ and $m_{\text{abs}} = 2$.

2.2.2 Labeled path

We represent a possession as a path on the skeleton graph, augmented with labels (both on vertices and edges it comes across). Precisely, we label each vertex by a time interval $T = [t_{\text{start}}, t_{\text{end}}[$ representing the period during which the spatial position of the vertex is valid. We also label each label by the nature of the temporal change: if the temporal change is absolute and/or relative spatial change, we add no label because this information is already contained in the edge's origin-destination vertices. However, we indicate on labels any thematic change. We define the path by recursion as follows.

Initialization. The path starts at the vertex that corresponds to the initial spatial position, absolute and relative, of the possession (which is always the same in our protocol). We set $t_{\text{start}} = 0$ the starting time of the time interval label of this first vertex. In other words, we initialize the starting time of the possession to 0.

Recurrence. Let $v \in \mathcal{K}$ be the last vertex of the path, which time interval label T has a known starting time $t_{\text{start}} = t$. We detect a change at time $t' > t$.

- If it is a spatial change (absolute and/or relative): Let us note $v' \in \mathcal{K}$ the vertex corresponding to the new spatial position of the game.
  - We set $t_{\text{end}} = t'$ the ending time of vertex v (i.e. T = [t,t'[).

- We add the arc (v, v') to the path.
- We add the vertex v' to the path, with the starting time of its time interval $t_{\text{start}} = t'$.

• If it is a thematic change: Let us note $[t', t'']$ the interval of time during the thematic change happens. We will see that we do not take into account what happens during the thematic change but only in what it results in. Let us note $v'' \in \mathcal{K}$ the vertex corresponding to the spatial position of the game at time t''.

  - We set $t_{\text{end}} = t'$ the ending time of vertex v (i.e. $T =$ [t,t'[).
  - We add the arc (v, v'') to the path and label it "Thematic".
  - We add the vertex v'' to the path, with the starting time of the time interval $t_{\text{start}} = t''$.

The path ends when the possession ends. If the possession is a success, we add to the last vertex of the path the edge leading to vertex "Target" and we add this (last) vertex to the path. Otherwise, if the possession is a failure, we add to the last vertex of the path the edge leading to vertex "¬ Target" and we add this (last) vertex to the path.

In Figure 2, we present an example of a path on the skeleton graph of the previous example. The initial vertex is $((0, 2), A_1)$. The possession lasts 5 units of time, where the only thematic change lasts 0.9 units of time (involving a relative spatial change). The path represents a successful possession, and is composed of 4 vertices and 3 edges. The vertices/edges of the paths are in red, and their labels are in blue.

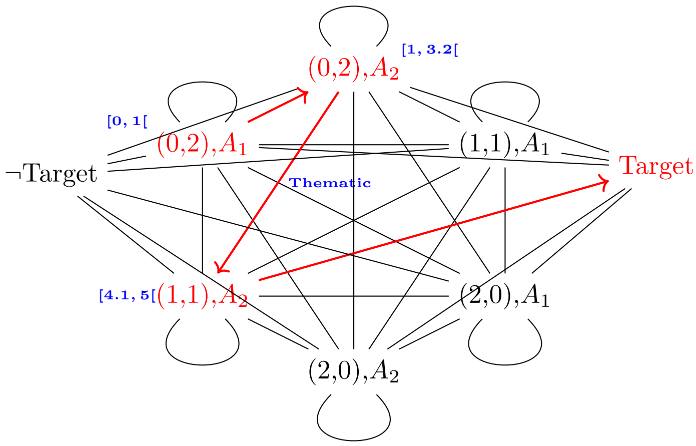


Figure 2: Example of a path on a skeleton graph.

3 Application to rugby and basketball

In this section, we apply our model of possession to two collective sports: rugby and basketball. For each of them, we choose the parameters of the model (partitions in m_{abs} absolute zones and m_{rel} relative zones, and the nature of thematic changes). This choice is lead by the specificity of the sport. On the one hand, the aim in rugby is to make the ball progress through the field toward the try line, which makes us naturally consider absolute zones parallel to the try line. Moreover, any player ahead of the ball carrier is offside and cannot receive the ball, thus we do not specifically express the number of players ahead in the relative zones because it is a temporary situation. On the other hand, the goal in basket is to shoot, and for that, to manage making free space between the ball carrier and the basket. The position from where the ball carrier shoots determine the number of points, which is expressed with the same partitioning in absolute zones. For the relative zones, the partitioning is naturally oriented toward the basket.

3.1 Rugby

The field is a rectangle, where one of the smallest edge is the try line. The aim of the attacking team, composed of $n = 6$ players, is to bring the ball to the try line.

3.1.1 Spatial information

Absolute zones. We decompose the field into 3 absolute zones: B (Back), M (Middle) and F (Front), i.e.

$$\mathcal{A} = \{B, M, F\}\,.$$

The absolute zones are depicted in Figure 3.

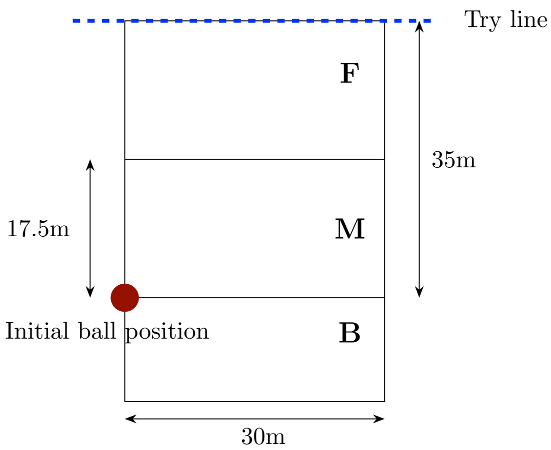


Figure 3: Absolute zones in rugby.

Relative zones. For a given position $pos(t)$, we divide the field into 2 relative zones L (Left) and R (Right), that are the zones to the left and to the right of the line passing through $pos(t)$ and perpendicular to the try line. Thus, we consider

$$\mathcal{R} = \{(l, r) : l + r = 5,\, l, r \in \mathbb{N}\}\,.$$

See Figure 4 for illustration.

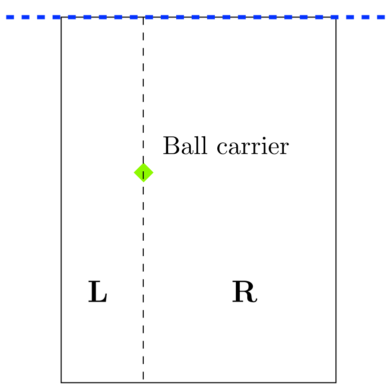


Figure 4: Relative zones in rugby.

3.1.2 Temporal information

Thematic changes. We consider thematic changes as passes of different nature. Specifically, we consider 3 type of passes: back pass (bp), diagonal foot pass (d-fp) and straight foot pass (s-fp).

3.2 Basketball

The field is a rectangle. The aim of the attacking team, composed of $n = 5$ is to shoot the ball into the basket.

3.2.1 Spatial information

Absolute zones. We divide the field into 3 zones: K (Key), 2P (2-point zone) and 3P (3-point zone), i.e.

$$\mathcal{A} = \{K, 2P, 3P\}.$$

The absolute zones are depicted in Figure 5.

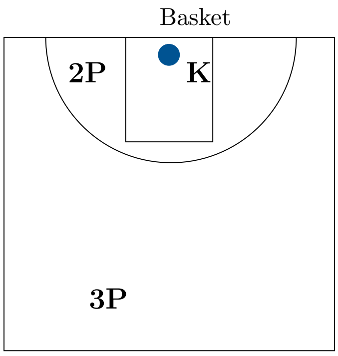


Figure 5: Absolute zones in basketball.

Relative zones. For a given position $pos(t)$, we divide the field into 4 relative zones: North-West (NW), North-East (NE), South-West (SW) and South-East (SE). The split is done with the line passing through the basket and $pos(t)$, and the perpendicular line passing through $pos(t)$. Thus, we consider the set of relative positions

$$\mathcal{R} = \{(nw, ne, sw, se) : nw + ne + sw + se = 4,\ nw, ne, sw, se \in \mathbb{N}\}.$$

See Figure 6 for illustration.

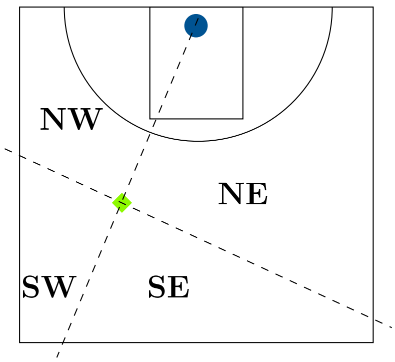


Figure 6: Relative zones in basketball.

3.2.2 Temporal information

Thematic changes. We consider passes and screen as thematic changes. Specifically, we consider the followings passes: hand-off (hp), regular (rp) and deviate (dp) pass. Moreover, we also consider the following screens: carrier screen (cs) or non-carrier screen (ncs).

4 Measures on the model

Our model is designed to be applied to the data protocol, generating graphs, with the goal of comparing the two pedagogical approaches. To facilitate this comparison, we introduce several measures based on the resulting graphs. Specifically, the protocol provides a set of possessions under the prescriptive pedagogy and another set under the auto-organized pedagogy. Consequently, the resulting graphs consist of two sets of paths on the same underlying skeleton graph, all starting from the same vertex (as the initial position on the field is fixed).

We consider measures at different level: *local* measures, at the level of a path, or *global* measures, at the level of a graph. The latter graph is the sum (i.e. weighted union) of paths of a set (corresponding to one pedagogy, or one pedagogy and one defense, etc.), thus it is a subgraph of $\mathcal{K}$. Moreover, for each kind of measures, we distinguish two types: *similarity* and *features*. On the one hand, similarity relates to a comparison between two mathematical objects (pair of paths or par of subgraphs in this case), such as a distance, meaning that the comparison is pairwise. On the other hand, features are absolute indicators of a given mathematical object, absolute in the sense that their values does not depend on other objects. We summarize in Table 1 several possible measures, and we describe some of them in details next. Notice that the aim of this section is to find relevant measures, which can be different according to the collective sport we study.

	Global	Local
Similarity	Edit distance (with labels) Matrix distances	Size of symmetric difference Maximal common subpaths Longest common subsequence
Features	Density Degree distribution Centrality Number of triangles	Path length Number of thematic labels Length between thematic labels Matching (fuzzy) patterns

Table 1: Different possible measures.

4.1 Local similarity

We propose next several ways of measuring the distance between two paths. Notice that these two paths start with the same vertex and can have different lengths. Let us note $P(v, \mathcal{K})$ the set of paths starting from the vertex v in $\mathcal{K}$. In our model, v is the vertex corresponding to the initial spatial position of the attack team. In what follows, we propose several definitions of a function

$$\Delta : P(v, \mathcal{K}) \times P(v, \mathcal{K}) \to \mathbb{R}^+ ,$$

which represents the distance between a pair of paths. The distance between two idendical paths must be equal to 0, and take larger values when paths are *less alike*. This is this notion of *likeness* that we define with the function Δ.

4.1.1 Maximal common subpaths distance

Possessions under the prescriptive pedagogy, for a same defense scenario, have the same instructions. In other words, the players should follow the same sequence of actions. Thus, a natural hypothesis is that two paths corresponding to two possessions under this pedagogy should have more and longer subpaths in common than two paths corresponding to two possessions under auto-organized pedagogy. We say that a subpath is common to two paths if it is contained in the two paths, wherever its position. We say that a common subpath if *maximal* if it is a common subpath, and that it is not contained in at least one of the two paths when adding any vertex to it. Notice that for a given pair of paths, there can be several maximal common subpaths (with no vertices in common), possibly with different lengths (equal to the number of edges of the path).

Formally, for two paths $p_1, p_2 \in P(v, \mathcal{K})$, we get the vector of the lengths of the maximal common subpaths $l(p_1, p_2) = \begin{pmatrix} l_1 \\ l_2 \\ \vdots \\ l_{N(p_1,p_2)} \end{pmatrix} \subseteq \mathbb{N}_*^{N(p_1,p_2)}$. Note that the size of the vector $N(p_1, p_2)$ depends on the number of maximal common subpaths between the two paths.

We display in Figure 7 an example.

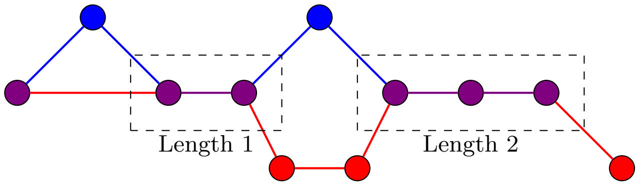


Figure 7: Example of two paths (p_1 in blue and p_2 in red) with common vertices/edges (in violet). There are two maximal common subpaths (in dashed rectangle), of lengths 1 and 2.

In this example, $l(p_1, p_2) = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$.

We can define the distance function $\Delta : (p_1, p_2) \mapsto f(l(p_1, p_2))$, where f is a function that decreases with the largest coordinate of the lengths vector. For instance, we can take

$$f(l(p_1, p_2)) = \max(|p_1|, |p_2|) - \max_i l_i, \quad \text{or} \quad f(l(p_1, p_2)) = \max(|p_1|, |p_2|) - \sum_i l_i\,,$$

where $|p|$ denotes the length of path p (number of edges). Notice that at this stage, we consider that two vertices, respectively two edges, are common if they are the same vertex, resp. edge, in the skeleton graph. Yet, we do not precise if they have the same label or not. Thereafter, it seems that the label on vertices (time interval) should not be considered, whereas the label on edges should.

There is the possibility to refine the notion of maximal common subpath by weighting its length by the distance between the two respective starting vertices for instance. This could express that finding the same sequence of actions at the beginning or at the end of a path might not mean the same thing. This idea appears in the next proposition of distance.

Note that we can also relax the notion of maximal common subpath by allowing, for instance, one vertex of difference. It would be a *fuzzy* maximal common subpath.

4.1.2 Sliding vertex-to-vertex distance

Distance between two vertices. We note $d_v : V_{\mathcal{K}} \times V_{\mathcal{K}} \to \mathbb{R}^+$ a distance function between two vertices of $\mathcal{K}$. For instance, we could define the following distance. Let $v_1 = ((N_1, \ldots, N_{m_{\text{rel}}}), A)$ and $v_2 = ((N'_1, \ldots, N'_{m_{\text{rel}}}), A')$ be two vertices of $\mathcal{K}$ (except the targets vertices). We can define the *relative* distance d_v^r as the sum of the difference of number of players in each relative zone, namely $\sum_i |N_i - N'_i|$. We can define the *absolute* distance d_v^a as the minimum number of absolute zones a player should change to go from A to A'. For instance, in the case of rugby, the absolute distance between Back and Middle is 1, and between Back and Front is 2. Eventually, we combine the two distances to define a common distance between two nodes, e.g.

$$d_v(v_1, v_2) = d_v^r(v_1, v_2) + 2 \cdot d_v^a(v_1, v_2)\,.$$

Distance between two paths of same length. We note d_p the distance function between two paths p_1, p_2 of same length (len). Notice that we set $[\![0, \ldots, len-1]\!]$ the indices of the vertices of respective paths. We consider the function

$$d_p(p_1, p_2) = g(d_v(p_1^0, p_2^0), d_v(p_1^1, p_2^1), \ldots, d_v(p_1^{len-1}, p_2^{len-1}))\,,$$

where p^i denotes the vertex of index i in path p, and g is a function non-decreasing with the values of d_v. This represents a distance vertex-to-vertex between the two paths, namely that the distance is a function of the distances pairwise vertex-to-vertex along the paths. Example is given in Figure 8. For instance, we can choose g as

$$g(d_v(p_1^0, p_2^0), \ldots, d_v(p_1^{len-1}, p_2^{len-1})) = \frac{1}{len} \sum_k d_v(p_1^k, p_2^k)\,.$$

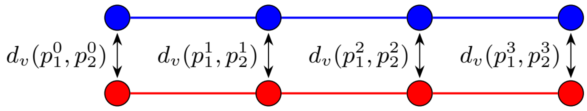


Figure 8: Example of computation of distances between two paths of same length.

Distance between two paths. We define a distance between two paths $p_1, p_2 \in P(v, \mathcal{K})$. We suppose without loss of generality that $|p_1| \leq |p_2|$, and we set $\delta := |p_2| - |p_1|$ the length difference. We consider the following function:

$$\Delta(p_1, p_2) = f(d_p(p_1, p_2^{[0]}), d_p(p_1, p_2^{[1]}), \ldots, d_p(p_1, p_2^{[\delta]}))\,,$$

where $p_2^{[i]}$ is the subpath (of length $|p_1|$) of p_2 corresponding the vertices of indices $[\![i, \ldots, i+len_1-1]\!]$, and where f is a function non-decreasing with the values of d_p. Example is given in Figure 9. For instance, we could choose

$$f(d_p(p_1, p_2^{[0]}), \ldots, d_p(p_1, p_2^{[\delta]})) = \frac{1}{\delta} \sum_{i=0}^{\delta} i \cdot d_p(p_1, p_2^{[i]})\,.$$

or

$$f(d_p(p_1, p_2^{[0]}), \ldots, d_p(p_1, p_2^{[\delta]})) = \min_i d_p(p_1, p_2^{[i]})\,.$$

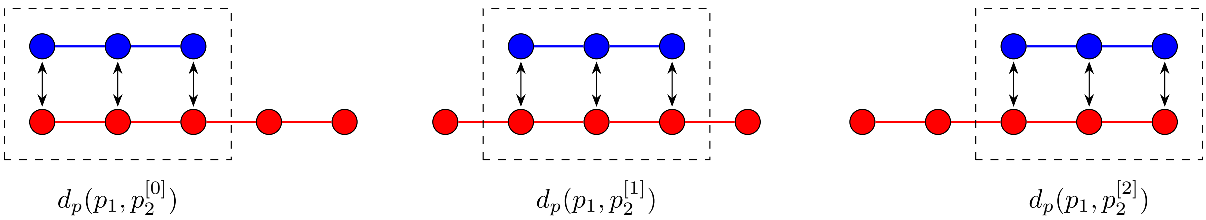


Figure 9: Example of the computation of distance between two paths (p_1 in red, p_2 in blue). The length difference is equal to 2, thus it requires 3 computations of distances between subpaths of same length.

Slight modifications. We can define a slightly modified distance than above by considering the position in p_2 of the initial vertex of the subpath $p_2^{[i]}$ when computing $d_p(p_1, p_2^{[i]})$. For that, we take into account that for a position that is advanced in p_2, the distance is expected to be larger because time has passed between the two starting vertices of the two respective paths. We propose

$$d_p(p_1, p_2^{[i]}) = \sum_k \frac{1}{i+k} d_v(p_1^k, p_2^k)\,.$$

4.1.3 Jaccard distance

Let $p = (e_0, e_1, \ldots, e_m)$ and $p' = (e'_0, e'_1, \ldots, e'_{m'})$ be two paths (described by a sequence of edges). Notice that for the moment, we do not consider the labels, neither on vertices nor edges. If we see the sequence of edges as a set, $E = \{e_0, \ldots, e_m\}$ respectively $E' = \{e'_0, \ldots, e'_m\}$, we can compute the Jaccard distance, also called Jaccard index (Jaccard, 1901), between these two paths as follows:

$$J_d^{\text{edge}}(p, p') = 1 - \frac{|E \cap E'|}{|E \cup E'|} = \frac{|E \Delta E'|}{|E \cup E'|}\,,$$

where Δ denotes in this case the symmetric difference. Note that we can also consider the Jaccard distance between the sets of vertices of each paths, namely,

$$J_d^{\text{vertex}}(p, p') = \frac{|V \Delta V'|}{|V \cup V'|}\,,$$

where V, respectively V', is the set of vertices of p, resp. p'. The distance J_d^{vertex} does not take into account the time relation between the vertices, whereas J_d^{edge} expresses at least the time relation between two consecutive vertices. Thus, J_d^{edge} seems more appropriate. Notice that we could also define the Jaccard distance between the sets of all subpaths of size 2 or 3 etc in order to take into account the succession of actions.

4.2 Divergence.

We define the *divergence* between two paths as follows.

4.2.1 Distance between two vertices

We define $d_v : V_{\mathcal{K}} \times V_{\mathcal{K}} \to \mathbb{R}^+$ a distance function between two vertices of $\mathcal{K}$ (except vertices $Target$ and $\neg Target$). This distance is defined as the sum of the the *absolute-wise* distance d_v^{abs} and the *relative-wise* distance d_v^{rel} defined below.

Absolute-wise distance. The *absolute-wise* distance $d_v^{\text{abs}} : \mathcal{A} \times \mathcal{A} \to \mathbb{N}$ represents the minimum number of absolute zones a player should change to go from one absolute zone to an other. Formally, if we consider the partition of the field into absolute zones as a floor plan, the distance between two zones is the length of the shortest path between the two corresponding vertices in the dual graph (illustrated in Figure 10).

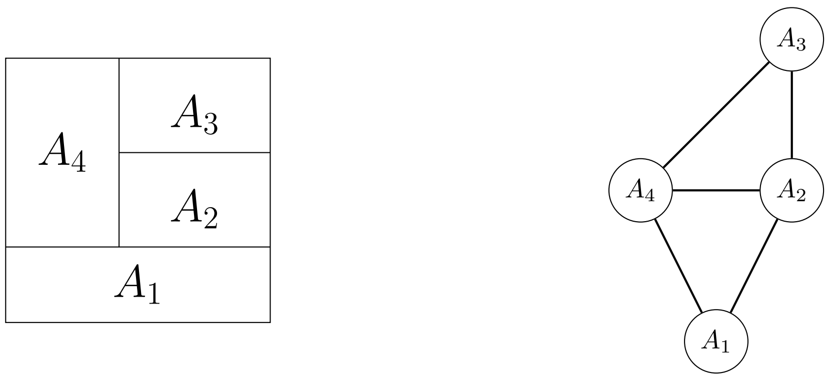


Figure 10: Example of a floor plan of absolute zones (left) and its dual graph (right). In the dual graph, each vertex represents a zone, and there exists an edges between two vertices if the two corresponding zones are adjacent in the floor plan. In this example, for instance, $d_v^{\text{abs}}(A_1, A_2) = 1$ and $d_v^{\text{abs}}(A_1, A_3) = 2$.

Relative-wise distance. The *relative-wise* distance $d_v^{\text{rel}} : \mathcal{R} \times \mathcal{R} \to \mathbb{N}$ represents the minimum number of changes of relative zones the players should do to transform one relative position into the other. Formally, if we consider the partition of the field into relative zones (which is independent of the position of the ball carrier because only the size of each relative zone changes, not the adjancencies) as a floor plan, the distance between two relative positions is the minimum flow between these two affectations in the dual graph (where edges capacities are maximum, equal to $n - 1$). We illustrate this notion in Figure 11.

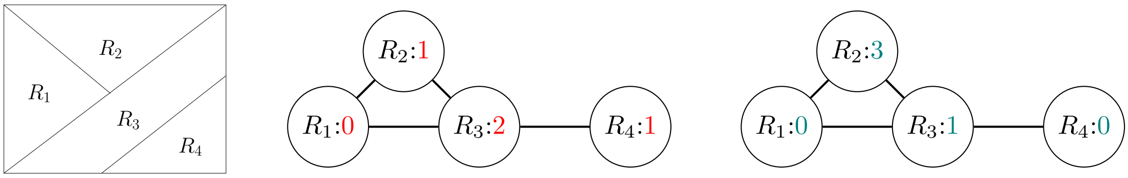


Figure 11: Example of a floor plan of relative zones (left) and two relative positions (right - red and green) represented by assignements of number of players on vertices of the dual graph. In this exemple, the distance between these two relative positions is $d_v^{\text{rel}}((0,1,2,1),(0,3,1,0)) = 3$.

In conclusion, for $v_1 = ((N_1,\ldots,N_{m_{\text{rel}}}),A)$ and $v_2 = ((N'_1,\ldots,N'_{m_{\text{rel}}}),A')$ two vertices, the distance between two nodes is:

$$d_v(v_1,v_2) = d_v^{\text{abs}}(A,A') + d_v^{\text{rel}}((N_1,\ldots,N_{m_{\text{rel}}}),(N'_1,\ldots,N'_{m_{\text{rel}}}))\,.$$

Application to rugby. For the case of rugby, the distance between two vertices is expressed as follows, for which the absolute and relative-wise distances are easy to express. Let $v = ((N_1,N_2),A)$ and $v' = ((N'_1,N'_2),A')$ two vertices. The distance between them is

$$d_v(v,v') = |N_1 - N'_1| + \begin{cases} 0 & \text{if} \quad A = A' \\ 1 & \text{if} \quad (A,A') \in \{(B,M),(M,B),(M,F),(F,M)\} \\ 2 & \text{if} \quad (A,A') \in \{(B,F),(F,B)\} \end{cases}$$

4.2.2 Distance between two arcs' semantic

We define the distance $d_s : (\mathcal{TC} \cup \{\emptyset\}) \times (\mathcal{TC} \cup \{\emptyset\}) \to \mathbb{N}$ between two arcs semantic as follows. We recall that in our model, an arc can have eather a thematic label in $\mathcal{TC}$, or has no label (representing a spatial change only). We define the distance between the labels of arcs a and a' as follows, where we note $lab(a)$ the label of a, equal to $\emptyset$ if a has no label:

$$d_s(lab(a),lab(a')) = \begin{cases} 0 & \text{if} \quad lab(a) = lab(a') \\ 1 & \text{if} \quad lab(a) \neq lab(a'), \text{ and } lab(a), lab(a') \in \mathcal{TC} \\ 2 & \text{if} \quad lab(a) \neq lab(a'), \text{ and } lab(a) = \emptyset \text{ or } lab(a') = \emptyset \end{cases}$$

4.2.3 Distance between two arcs

We define the distance $d_{\text{arc}} : E_{\mathcal{K}}^2 \times E_{\mathcal{K}}^2 \to \mathbb{N}$ between two arcs $a = v_1 \to v_2$ and $a' = v'_1 \to v'_2$ as follows:

$$d_{\text{arc}}(a,a') = \frac{d_v(v_1,v'_1)}{2} + \frac{d_v(v_2,v'_2)}{2} + d_s(lab(a),lab(a'))\,.$$

We schematized in Figure 12 the different distances (between vertices and semantic labels) involved in the distance between two arcs. We provide a numerical example in Figure 13.

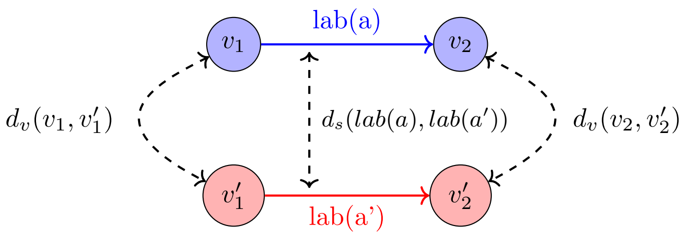


Figure 12: Distances involved when comparing two arcs (blue and red).

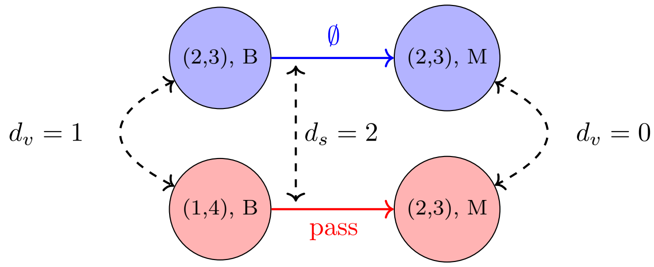


Figure 13: Example of the computation of distance between two arcs (blue and red) for rugby. The distance is $d_{\mathrm{arc}}(a, a') = 1 + 0 + 2 = 3$.

4.2.4 Definition of the divergence

Let p_1 and p_2 two paths (beginning by the same vertex) of length l_1, respectively l_2. We rank the arcs of each path by increasing order, beginning by 1. We note $r(a) \in \mathbb{N}^*$ the rank of arc a.

Definition 4.1. *Let us asssume at this point that $l_1 \le l_2$. We define the divergence of p_1 from p_2, with time-window of size $k \in 2\mathbb{N} + 1$, as follows:*

$$div(p_1, p_2) = \sum_{a_1 \in p_1} \min_{\substack{a_2 \in p_2 \\ r(a_1) - \frac{k-1}{2} \le r(a_2) \le r(a_1) + \frac{k-1}{2}}} d_{arc}(a_1, a_2) + \frac{\epsilon}{l_1} \cdot |r(a_1) - r(a_2)|\,,$$

where $\epsilon > 0$.

The time-window parameter represents the size of the set of arcs of p_2 we compare with an arc of p_1, ranks centered in the rank of the arc of p_1. For instance, choosing $k = 3$ means that each arc a_1 of p_1 is compared with the 3 arcs in p_2 of ranks $\{r(a_1) - 1, r(a_1), r(a_1) + 1\}$. Notice that for the first, resp. the last, arc of p_1, this set is, resp. can be, of size $k - 1$. We illustrate in Figure 14 the computations involded in div for $k = 3$. Each dashed black arrow represents the distance d_{arc} between two arcs (Figure 12).

Figure 14: Example of the arcs' distances involved in the divergence of path p_1 (in blue) from path p_2 (in red). Numbers in blue, resp. in red, are the ranks of arcs for each respective path.

Notice that considering the time-window parameter $k = 3$ is arbitrary, such as all the numerical values choosen for the definition of d_v and d_s, and can be changed according to the meta-parameters of the model (number of absolute zones etc.) and the specificities of the collective sport considered (nature and diversity of semantic etc.).

Next, we define the symmetric divergence of two paths p_1 and p_2 by *symmetrizing* the function div, so that we do not need any assumption on the lengths of the paths.

Definition 4.2. *Let p_1 and p_2 two paths. We define the generalized divergence between them as follows:*

$$div_{sym}(p_1, p_2) = \mathbb{1}_{\{l_1 < l_2\}} div(p_1, p_2) + \mathbb{1}_{\{l_1 > l_2\}} div(p_1, p_2) + \mathbb{1}_{\{l_1 = l_2\}} \frac{div(p_1, p_2) + div(p_2, p_1)}{2}\,.$$

We state below several properties of the (symmetric) divergence and the other distances introduced above.

Property 4.3. *The functions d_v^{abs}, d_v^{rel} and d_s, are distances on $\mathcal{A}$, $\mathcal{R}$ and $\mathcal{TC} \cup \{\emptyset\}$ respectively. It results that d_{arc} is a distance on $E_{\mathcal{K}}^2$.*

Proof. The non-negativity and symmetry properties for each function is clear. We prove the triangle inequality with a proof of contradiction for d_v^{abs} and d_v^{rel}, and with a proof of cases for d_s. Moreover, because d_{arc} is a linear combination of the three above-mentionnend distances, it is also a distance. □

Property 4.4. *The symetric divergence is non-negative and symmetric. However, it does not (seem to) respect the triangle inequality.*

Proof. The non-negativity and symmetry is clear. The (supposed) violation of the triangle inequality comes from the minimum in the definition of the divergence *div*. □

Remark 4.5. *An other natural definition of the divergence could have been to replace the minimum by a sum as following:*

$$div(p_1, p_2) = \sum_{a_1 \in p_1} \sum_{\substack{a_2 \in p_2 \\ r(a_1) - \frac{k-1}{2} \leq r(a_2) \leq r(a_1) + \frac{k-1}{2}}} d_{arc}(a_1, a_2) + \frac{\epsilon}{l_1} \cdot |r(a_1) - r(a_2)|\,.$$

In this case, the divergence is symmetric, and respects the triangle inequality. However, its is not anymore non-negative. Specifically, the divergence of a path with itself can be non-zero.

4.3 Local features

We can consider the following local features for a labeled path representing a possession:

- Path length
- Number of thematic labels
- Average number of vertices between two consecutive thematic labels
- Average time spend on a vertex

4.4 Global similarity

4.4.1 Edit distance

We consider the following edit distance between two graphs G_1 and G_2 (Sanfeliu and Fu, 1983). The edit distance between G_1 and G_2 is the minimum number of unitary operations necessary to transform G_1 into G_2 (and vice versa), where the unitary operations are:

- Add a vertex
- Remove a vertex
- Add an edge
- Remove an egde

The definition of the edit distance above is the standard one found in the literature. Note that it can be adapted, by modifying (the cost of) the unitary operations, to suit the best what represents the distance between two graphs for our specific problem. In future work, we could for instance integrate the notion of edges with/without thematic labels etc.

4.5 Global features

For these type of measures, we consider the subgraph representing the sum (weighted union) of all paths of a set of possessions. For our protocol, we are interested in the set of possessions under one specific pedagogy. Notice that, because all paths begin by the same vertex, the subgraph is rooted. Let us note $G = (V, E) \subseteq \mathcal{K}$ the subgraph, where V is the set of vertices and E is the set of edges.

4.5.1 Density

The density of a graph is defined by the ratio between the number of edges it contains and the number of all possible edges between all pair of vertices (Diestel, 2005). Notice that in our case, a self-loop on a vertex is possible. Thus, the density d of G is defined by

$$d = \frac{|E|}{\frac{1}{2}|V|(|V|-1)+|V|} = 2 \cdot \frac{|E|}{|V|(|V|+1)} \,.$$

A slightly modified version of the density could also be interesting in our modelization. We can define the *augmented* density d_{augm} of G by the ratio between the number of edges it contains and the number of edges of the skeleton graph $\mathcal{K} = (V_{\mathcal{K}}, E_{\mathcal{K}})$. In other words,

$$d_{\mathrm{augm}} = \frac{|E|}{|E_{\mathcal{K}}|} = \frac{|E|}{\frac{1}{2}(|V_{\mathcal{K}}|-2)(|V_{\mathcal{K}}|-1)+2(|V_{\mathcal{K}}|-2)} = 2 \cdot \frac{|E|}{|V_{\mathcal{K}}|^2+|V_{\mathcal{K}}|-6} \,.$$

4.5.2 Centralities

There exist many ways to measure centrality of edges and vertices of a graph, i.e. providing them a rank corresponding to their *importance* (Newman, 2018; Van Steen, 2010). We list some of them below.

Degree centrality. The degree centrality of a vertex v in a graph is its degree.

Closeness centrality The closeness centrality of a vertex v in a graph is

$$C_{\mathrm{close}}(v) = \frac{n-1}{\sum_{v'} d(v', v)} \,,$$

where $n-1$ is the number of reachable vertices from v, and $d(v', v)$ is the distance from v to v'.

Betweenness centrality The betweenness centrality of a vertex v in a graph is

$$C_{\mathrm{between}}(v) = \sum_{v_1, v_2} \frac{\sigma_{v_1,v_2}(v)}{\sigma_{v_1,v_2}} \,,$$

where σ_{v_1,v_2} is the number of shortest paths between the pair of vertices (v_1, v_2), and $\sigma_{v_1,v_2}(v)$ is the number of these shortest paths passing through v.

Pagerank centrality The pagerank of a vertex v in a graph is the ranking provided by the PageRank algorithm on the graph. Essentially, the higher the score, the more likely a random path leads to the vertex.

Katz centrality For a given $\alpha \in [0, 1]$, the Katz centrality of vertex v in a graph is

$$C_{\mathrm{Katz}}(v) = \sum_{k=1}^{\infty} \sum_{v'} \alpha^k (A^k)_{v',v} \,,$$

where A is the adjacency matrix of the graph.

5 Results for rugby

5.1 Data set

To describe the data of our protocol, we refer to the following elements:

$$\text{city} \in \{\text{agen}, \text{beziers}, \text{toulouse}, \text{racing}, \text{vannes}\}$$

$$\text{serie} \in \{\text{pre-test}, \text{intervention}, \text{post-test}\}$$
$$\text{pedagogy} \in \{\text{auto-organized}, \text{prescriptive}\}$$
$$\text{defense} \in \text{training} = \{\text{tight}, \text{open}, \text{foot}\} \cup \text{new} = \{\text{general}, \text{specific}\}$$

In our protocol, a unique possession is identified by the tuple (city, pedagogy, n_p), where $n_p \in \{1, \dots, 44\}$ is a possession number. For the series *pre-test* and *intervention*, we consider only the three training defenses. For the serie *post-test*, we consider two additional defenses which are the new defenses (thus five defenses in total for *post-test*). For a tuple (city, serie, pedagogy, defense) we have 4 possessions. Hence, for instance, a tuple (city, serie, pedagogy) represents 12 possessions for serie *pre-test* or *intervention*, and 20 possessions for serie *post-test*. Thus, in total, the data represent 440 possessions.

Notice that for this protocol, the coordinates in the field of each player are collected every $\tau = 20$ milliseconds, namely that the time in data is discretized. This discretization can be handle easily by evaluating the situation of the game every time step τ (instead of continuously as assumed in our theoretical model). This does not affect the resulting implemented path of the model because τ is much shorter than any possible thematic change, and also shorter than any consecutive spatial changes. Let us see below how we implement a path.

5.2 Implementation

The labeled path corresponding to a possession is described by two files .csv, one containing the vertices and the other containing the edges. The file of the vertices of the path contains the following information:

- Vertex rank: the ID of the vertex in the path.
- Relative and absolute position: indicates the corresponding vertex of the skeleton graph.
- Start and end time label: the time interval label of the vertex in the path. The time is expressed in seconds.
- Leaf: it is one of the vertex "Target" or "¬ Target" (named here "'Try" or "Failure" in the context of rugby).

The file of the edges of the path contains the following information:

- Edge rank: the ID of the edge in the path.
- Origin-Destination: indicates the vertices' ranks of the edges's origin and destination, noted as the couple (rank_origin, rank_destination).
- Change label: indicates the temporal change(s) label(s) of the edge in the path. Notice that for the specific case of the last edge (leading to the Leaf), the label specifies how ends the possession ("Ball lost by the carrier", "Failed pass" or '"Successful try"). Also notice that in this implementation, we also indicate in the label when there is an absolute/relative spatial change to ease the reading of the edge file.

We provide in Figures 15 and 16 two examples of possessions.

Vertex rank	Relative position	Absolute position	Start time label	End time label	Leaf
0	(0, 5)	Back	0.0	0.0	
1	(1, 4)	Back	0.56	1.56	
2	(1, 4)	Middle	1.56	1.64	
3	(3, 2)	Back	2.02	2.3	
4	(3, 2)	Middle	2.3	3.16	
5	(4, 1)	Middle	3.16	4.72	
6					Failure

Edge rank	Origin-Destination	Change label
0	(0, 1)	Successful pass hand no contact with relative change
1	(1, 2)	Absolute change
2	(2, 3)	Successful pass hand no contact with absolute and relative changes
3	(3, 4)	Absolute change
4	(4, 5)	Relative change
5	(5, 6)	Ball lost by the carrier

Figure 15: Vertex (left) and edge (right) files of the possession (agen, auto-organized, possession 1).

Notice that the interval of the first vertex of a path is always $[0, 0]$ because each possession starts with a pass.

Vertex rank	Relative position	Absolute position	Start time label	End time label	Leaf
0	(0, 5)	Back	0.0	0.0	
1	(1, 4)	Middle	0.74	1.32	
2	(2, 3)	Middle	1.32	1.44	
3	(1, 4)	Middle	1.74	3.58	
4	(0, 5)	Middle	3.58	3.86	
5	(0, 5)	Middle	4.3	5.08	
6	(0, 5)	Front	5.08	5.6	
7	(1, 4)	Front	5.88	6.6	
8	(2, 3)	Front	6.86	7.44	
9	(3, 2)	Front	7.76	8.4	
10	(4, 1)	Front	8.78	11.22	
11					Try

Edge rank	Origin-Destination	Change label
0	(0, 1)	Successful pass hand no contact with absolute and relative changes
1	(1, 2)	Relative change
2	(2, 3)	Successful pass hand no contact with relative change
3	(3, 4)	Relative change
4	(4, 5)	Successful pass hand contact
5	(5, 6)	Absolute change
6	(6, 7)	Successful pass hand contact with relative change
7	(7, 8)	Successful pass hand no contact with relative change
8	(8, 9)	Successful pass hand no contact with relative change
9	(9, 10)	Successful pass hand no contact with relative change
10	(10, 11)	Successful try

Figure 16: Vertex (left) and edge (right) files of possession (toulouse, prescriptive, possession 10).

5.3 Numerical results

In this subsection, we provide some preliminary results on the application of our model to differentiate the two pedagogies. Before presenting the result of several local/global measures introduced in Secion 4, we compute the success and progression of a possession (see definition below) in order to complement these results. Notice that success and progression do not result from our model but only from the raw data.

5.3.1 Preliminary computations: success rate and progression

Let us begin by computing the average success rate of a set of possessions, where 1 represents Success (i.e. Try) and 0 represents Failure. We recall that, on the contrary to the path length or other local features, these results do not depend on the model we study. To provide a more precise/relevant score to the possession, not binary but continuous, we also compute the average progression of the possessions. We define the progression of a possession as a pourcentage, thus between 0% and 100%, representing the portion of the field that the team manage to cross and bring the ball at the end of the possession. In other words, it represents the ratio of the progression line over the length of the field. For example, if the possession ends with a try, the progression is equal to 100%. But if the possession ends with the ball lost 30 meters from the initial starting line, thus the progression is equal to $\frac{30}{35} \cdot 100 \approx 85.7\%$ (see Figure 3 for the dimensions of the field). We display the results in Table 2.

Serie	Pedagogy	Av. success rate	Av. progression
Pre-test	All	0.41	58.7%
	Auto-org.	0.37	58.4%
	Prescriptive	0.46	59.0%
Intervention	Auto-org.	0.45	65.3%
	Prescriptive	0.6	70.1%
Post-test	Au.-Training	0.43	67.0%
	Pr.-Training	0.51	70.1%
	Au.-New	0.35	59.1%
	Pr.-New	0.36	56.6%

Table 2: Statistical results on success rate and progression.

In addition to the mean progression, we provide the Cumulative Distribution Function of the progression, which provide more details, in Figure 17 for *pre-test* and *intervention*, and in Figure 18 for *post-test*.

5.3.2 Occurence centrality

We display in Figure 19 the skeleton graph, in a way that it is link to the topology of the field. The bottom row is composed of all vertices with absolute position "Back", the middle row, "Middle" and the top row, "Front". Moreover, we place vertices with a low value of players to the right in

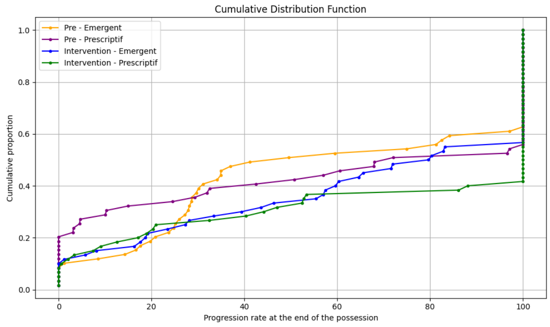


Figure 17: Cumulative distribution of the progression for *pre-test* and *intervention* series.

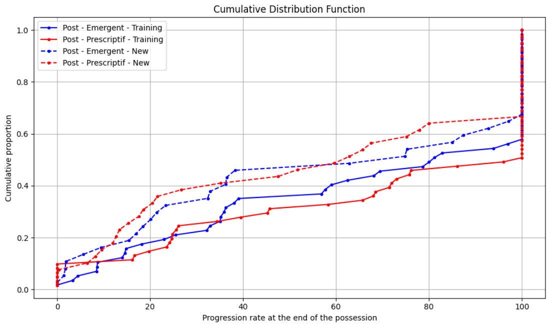


Figure 18: Cumulative distribution of the progression for *post-test* serie.

the relative position at the left because it is more likely that the ball carrier is at the left of the field in such a position.

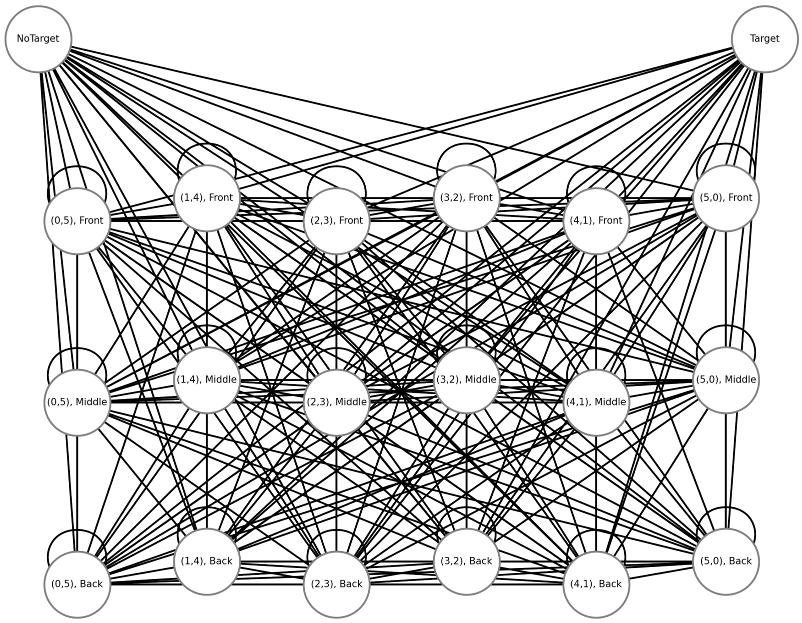
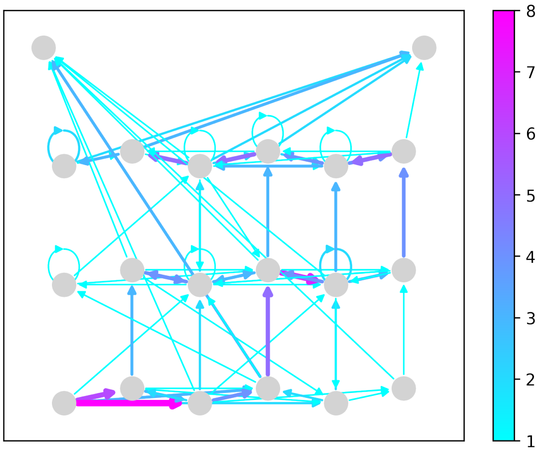


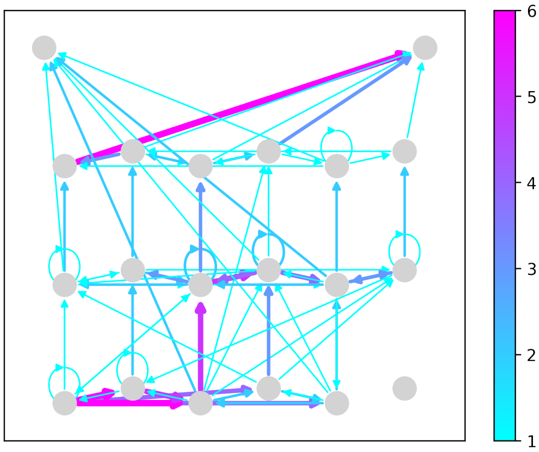
Figure 19: Skeleton graph.

In Figure 20, we display the subgraphs of all possessions of each couple (Pedagogy-Defense) during the Intervention phase (which represent 20 possessions). We define the subgraph of a set of possessions as union of all the paths where we keep in memory the occurence of each edge, respectively each vertex, (i.e. it is the *sum*, or the weighted union, of all paths). In other words, we represent the occurence edges centrality for each given set of possessions. We display on the same sets of possessions the occurence vertices centrality in Figure 20.

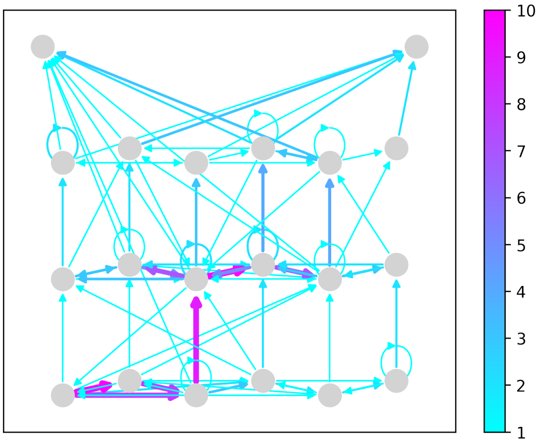
Discussion. Let us first discuss about occurence edges centralities (Figure 20). We observe that, starting from the bottom left corner (vertex $(0, 5)$, *Back*, starting position of each possession), frequent edges appear for both prescriptive and auto-organized pedagogies. However, notice that for (Auto-organized - Open), the most frequent edges are not totally at the beginning of the possession. Moreover, for prescriptive pedagogy, the most frequent edge leading to "Target" changes over the different defenses.



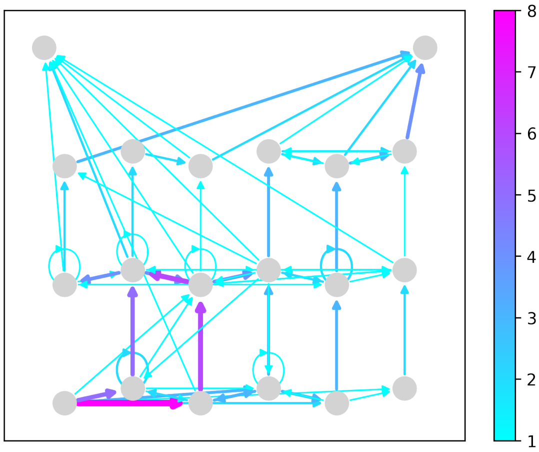
(a) Auto-organized - Tight



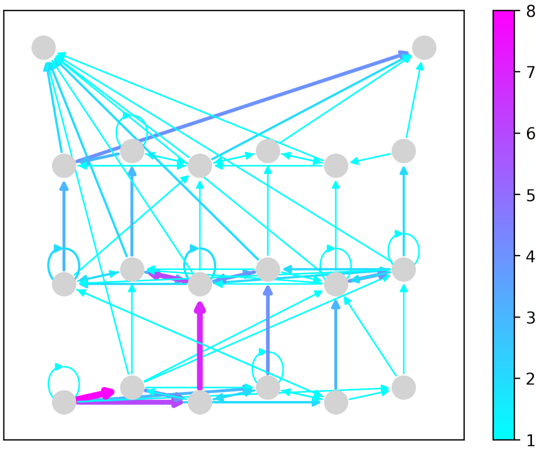
(b) Prescriptive - Tight



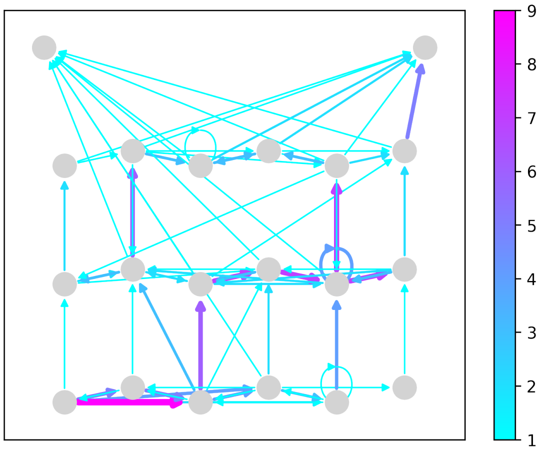
(c) Auto-organized - Open



(d) Prescriptive - Open

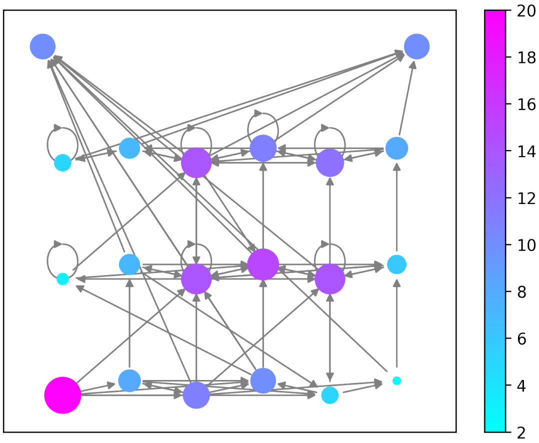


(e) Auto-organized - Foot

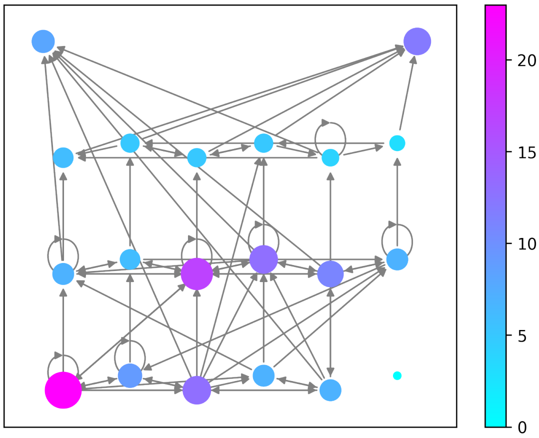


(f) Prescriptive - Foot

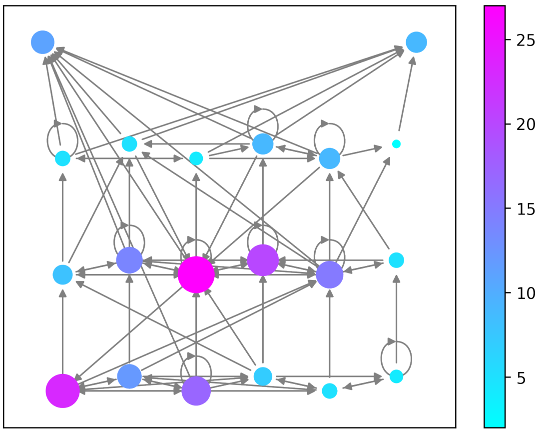
Figure 20: Occurence edges centrality of subgraphs for (Pedagogy - Defense) during Intervention. Colors and width represent occurrences of edges.



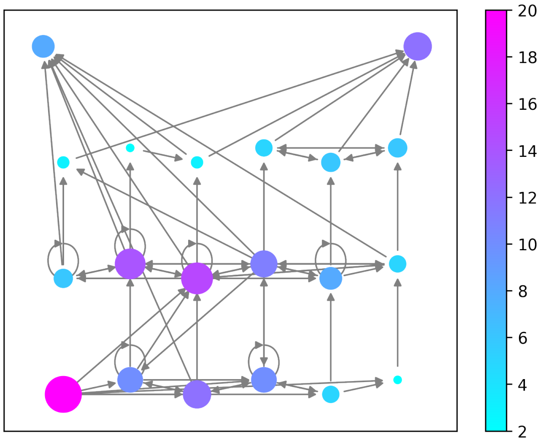
(a) Auto-organized - Tight



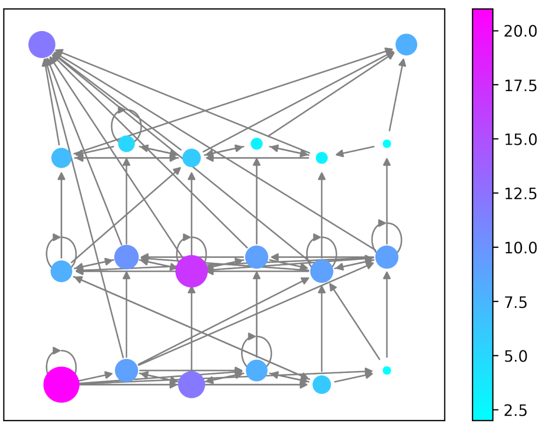
(b) Prescriptive - Tight



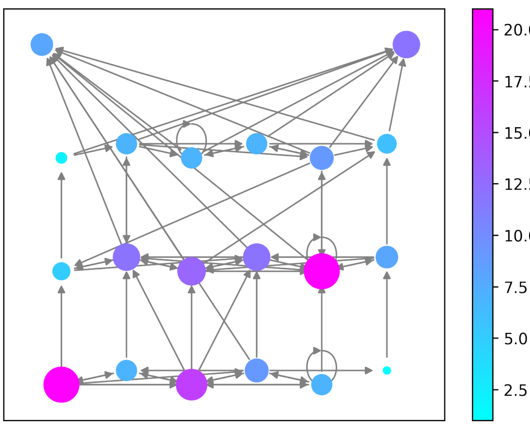
(c) Auto-organized - Open



(d) Prescriptive - Open

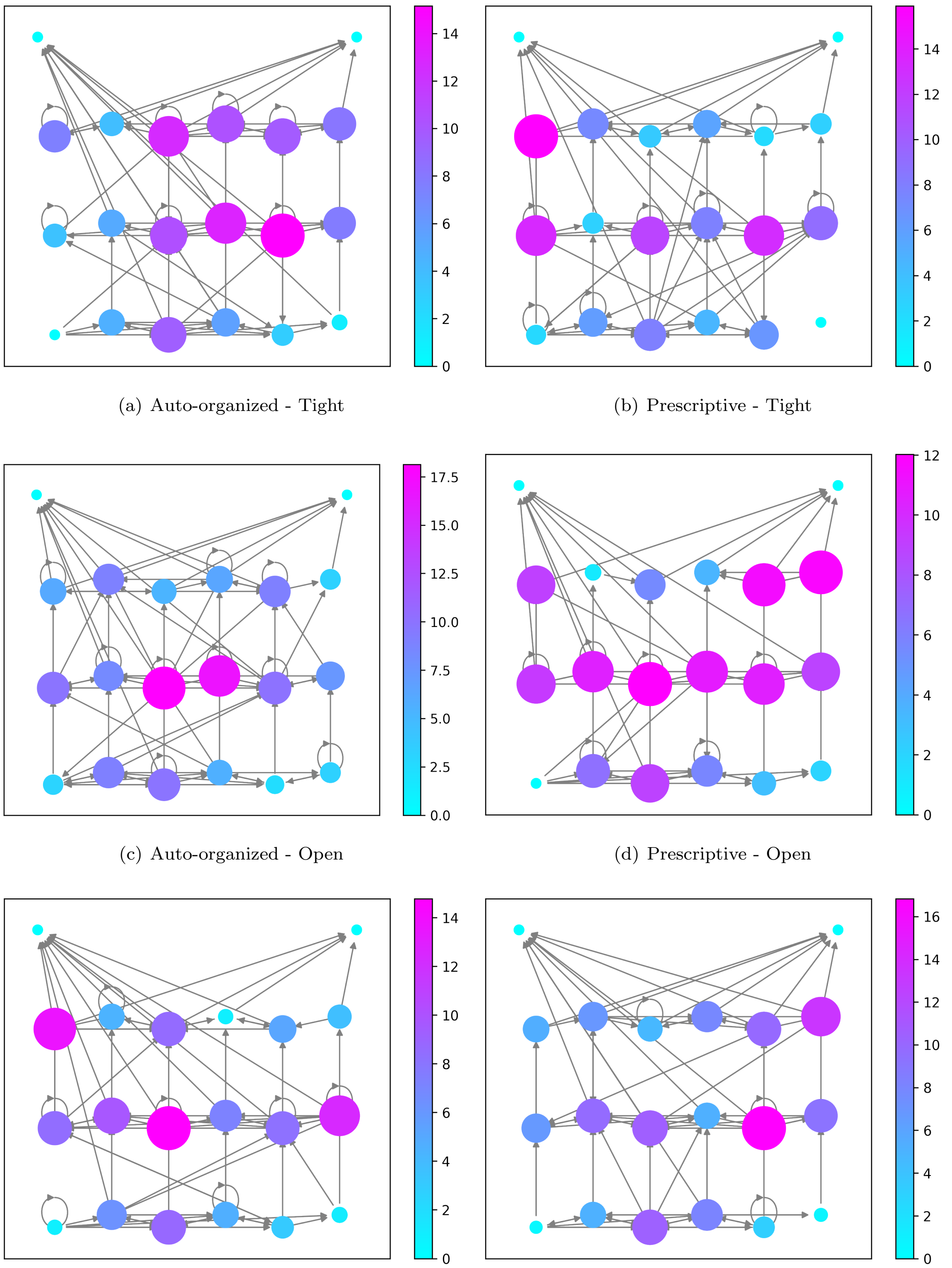


(e) Auto-organized - Foot



(f) Prescriptive - Foot

Figure 21: Occurence vertices centrality of subgraphs for (Pedagogy - Defense) during Intervention. Colors and size represent occurrences of vertices.



(a) Auto-organized - Tight

(b) Prescriptive - Tight

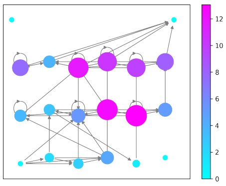
(c) Auto-organized - Open

(d) Prescriptive - Open

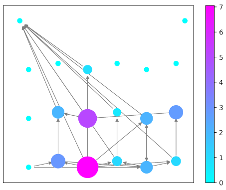
(e) Auto-organized - Foot

(f) Prescriptive - Foot

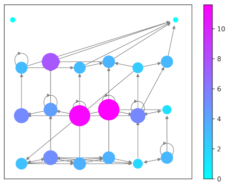
Figure 22: Time spent on vertices of subgraphs for (Pedagogy - Defense) during Intervention.



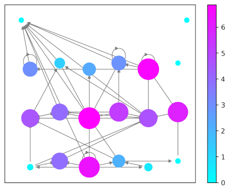
(a) Au.-org. - Tight - Success



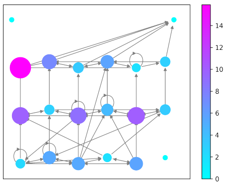
(b) Au.-org. - Tight - Fail



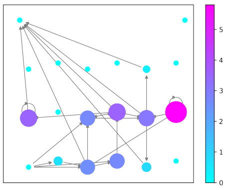
(e) Au.-org. - Open - Success



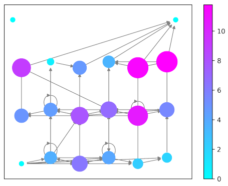
(f) Au.-org. - Open - Fail



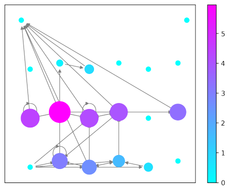
(c) Pres. - Tight - Success



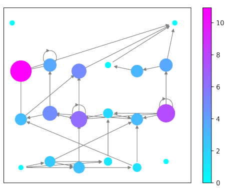
(d) Pres. - Tight - Fail



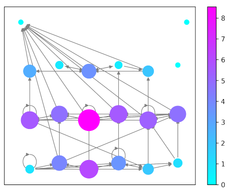
(g) Pres. - Open - Success



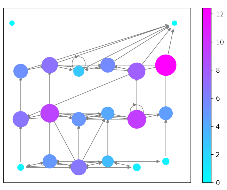
(h) Pres. - Open - Fail



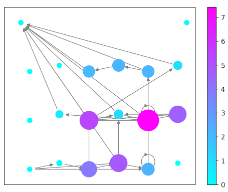
(i) Au.-org. - Foot - Success



(j) Au.-org. - Foot - Fail



(k) Pres. - Foot - Success



(l) Pres. - Foot - Fail

Figure 23: Time spent on vertices centrality of subgraphs of successful possessions (left columns) and failed possessions(right columns) for (Pedagogy - Defense) during Intervention.

5.3.3 Local features: path length

For each set of possessions corresponding to a couple (serie, pedagogy), we compute the mean length and the standard deviation For the Post-test serie, we filter also by the type of defense. The length of a path is defined by its number of vertices. We display the results in Table 3.

Serie	Pedagogy	Av. length	Std length
Pre-test	All	9.03	4.10
	Auto-org.	9.07	4.37
	Prescriptive	9.0	3.82
Intervention	Auto-org.	9.48	5.54
	Prescriptive	8.68	3.47
Post-test	Au.-Training	9.09	4.20
	Pr.-Training	8.11	2.97
	Au.-New	9.08	5.65
	Pr.-New	7.28	3.11

Table 3: Statistical results on path lengths.

In Figure 24, we display the results of Table 3 to ease their visualization. Precisely, we represent the evolution of the average length with a confidence interval that expresses the randomness that can come from the experiment. We display as a box this confidence interval for each result during and after the intervention of the coach, where the box's width corresponds to the difference of length between the two results in Pre-test. We also add as a complement the average success rate of each set of possessions studied (written next to the corresponding cross point).

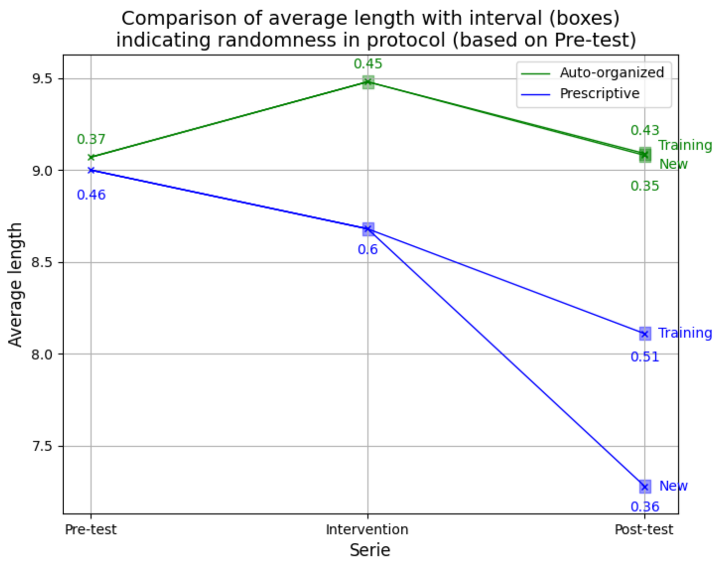


Figure 24: Evolution of average path length (with average success rate displayed).

Discussion. The evolution of the path lengths discloses a difference between auto-organized and prescriptive pedagogy. Indeed, during Intervention, the length of auto-organized paths is in average larger (around 10%) than prescriptive paths. It could be interpreted as the fact that auto-organized possessions need more re-organization (and thus spatial/thematic changes) leading to more vertices in the path, whereas prescriptive possessions follow a scenario defined in advance.

5.3.4 Local features: thematic labels

We compute in Table 4 the average number of vertices between two consecutive thematic labels for a set of possessions, and the average number of thematic labels. In the case of our implementation, we consider only one type of thematic label, which is a pass.

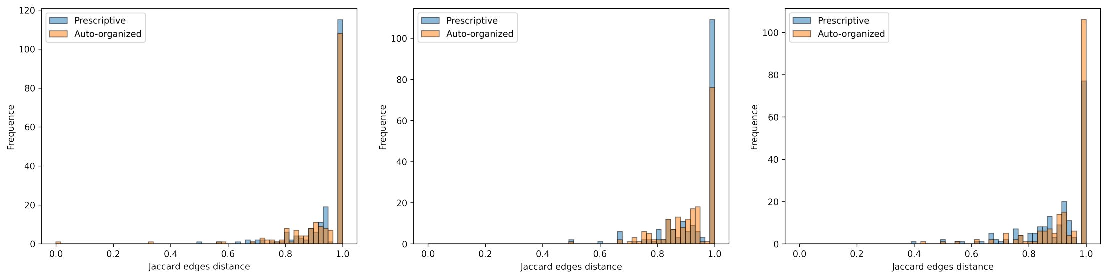
Serie - Defense	Pedagogy	Avg. nb vertices between passes	Avg. nb passes
Intervention - Tight	Auto-org.	0.81	4.1
	Prescriptive	0.75	3.5
Intervention - Open	Auto-org.	0.92	4.65
	Prescriptive	1.10	2.9
Intervention - Foot	Auto-org.	0.52	4.1
	Prescriptive	0.91	3.8

Table 4: Statistical results on passes.

Discussion. We observe that the average number of passes is always larger whatever the defense, in auto-organized possessions than in prescriptive ones. A significant difference is observed for the Open defense. Notice also that the average number of vertices between two consecutive passes is roughly the same for the Tight and Open defenses, whereas it is almost multiplied by 2 for the Foot defense.

5.3.5 Local similarity: Jaccard distance

Let us consider the set of possessions for (Pedagogy - Defense) during Intervention. We compute the Jaccard edges distances for all pair of possessions (190 pairs) and we display the distribution in Figure 25.



(a) Tight (b) Open (c) Foot

Figure 25: Distribution of the pairwise Jaccard edges distances for each defense during Intervention.

Discussion. The Jaccard edge distance provides a similarity measures between two paths. Namely, for a given defense, the more the distribution takes low values, the more the paths are similar pairwise. We hardly distinguish the two distributions, one for auto-organized and one for prescriptive pedagogy. However, we can guess that for Open and Foot defenses, the lowest values are mainly taken by prescriptive pedagogy, whereas this is the opposite for Tight defense.

5.3.6 Global similarity: edit distance

We compute the edit distance between pair of subgraphs representing the two pedagogies, for a given serie and defense. Notice that we display the smallest edit distance found after 120s of run.

5.3.7 Global features: density

For a given set of paths, we build the graph corresponding to the union of all the paths (which is a subgraph of $\mathcal{K}$). For each graph, we compute the augmented density (defined in Section 4). Notice

Serie	Defense	Edit distance
Pre-test	Tight	45
	Open	40
	Foot	39
Intervention	Tight	49
	Open	47
	Foot	41
Post-test	Tight	48
	Open	39
	Foot	47

Table 5: Edit distance between each pair of (directed) subgraphs of pedagogy Auto-organized VS Prescriptive.

that, on the contrary to the occurence edges/vertices centrality in Subsection 5.3.2, we take into account the time passing through each path, i.e. we consider that the edges of the subgraph are oriented. In this case, the number of arcs induced in the skeleton graph is $(|\mathcal{A} \times \mathcal{R}|^2 + 2|\mathcal{A} \times \mathcal{R}|)$, equal to 360 for rugby.

Serie	Pedagogy	Avg. augmented density	Details of augmented density
Pre-test	Auto-org.	0.2	Tight: 0.197 Open: 0.2 Foot: 0.203
	Prescriptive	0.211	Tight: 0.228 Open: 0.2 Foot: 0.206
Intervention	Auto-org.	0.228	Tight: 0.222 Open: 0.244 Foot: 0.217
	Prescriptive	0.205	Tight: 0.214 Open: 0.197 Foot: 0.203
Post-test	Auto-org.	Training: 0.199 New: 0.192	Tight: 0.236 Open: 0.183 Foot: 0.178 General: 0.181 Specific: 0.203
	Prescriptive	Training: 0.197 New: 0.181	Tight: 0.219 Open: 0.175 Foot: 0.197 General: 0.197 Specific: 0.164

Table 6: Augmented density considering time labels.

We display the results of Table 6 in Figure 26. Notice that the average density values are the round points, the boxes express the interval of confidence i.e. the randomness coming from real-world experiments, and the density values for each defense are cross points.

Discussion. The augmented density represents *diversity* of the edges of a set of possessions. The larger the augmented density, the more varied the actions taken in the set. During Intervention, we observe that the augmented density of the subgraph of auto-organized paths is larger than the one of the subgraph of prescriptive paths. It strengthens the idea that, for a given defense, the auto-organized possessions take more varied actions (i.e. more different edges) than the prescriptive possessions.

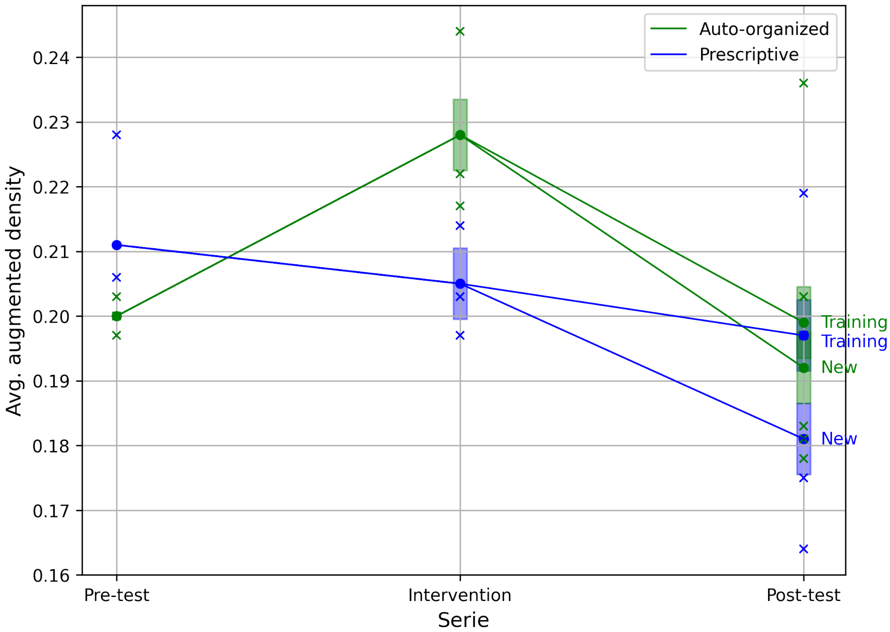
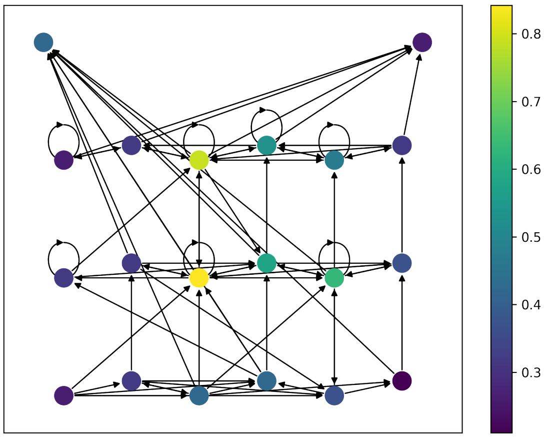


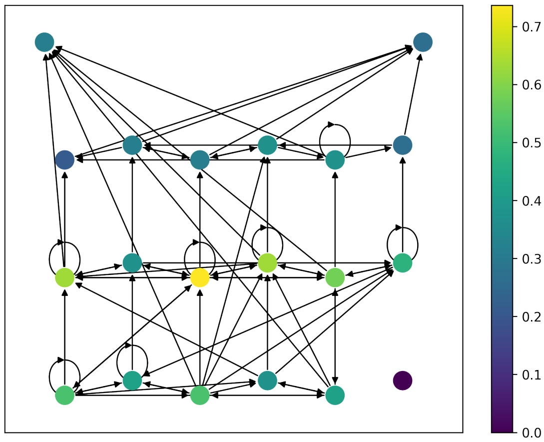
Figure 26: Evolution of augmented density.

5.3.8 Global features: different centralities

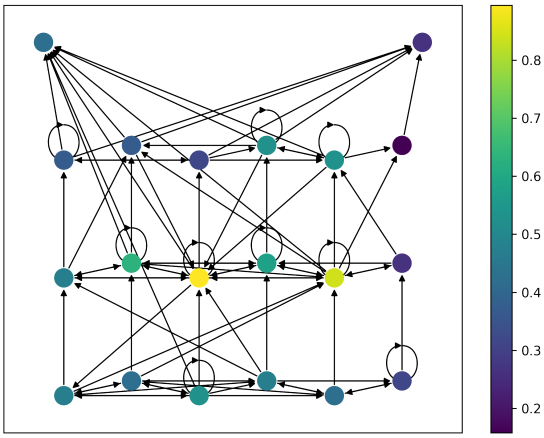
We present below different centralities on the subgraphs representing a given pedagogy and a given defense during Intervention. Notice that for some of them, we do not take into account the "Target" and "¬Target" because these vertices do not seem meaningful.



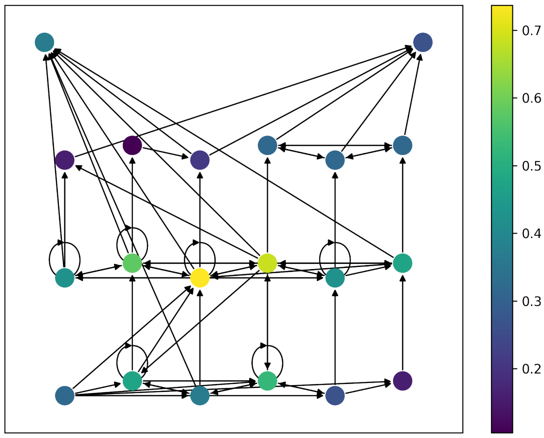
(a) Auto-organized - Tight



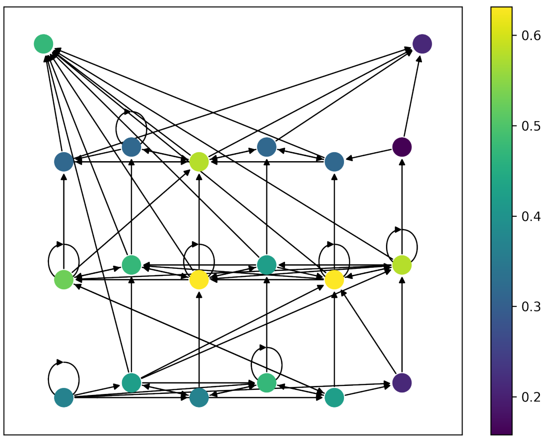
(b) Prescriptive - Tight



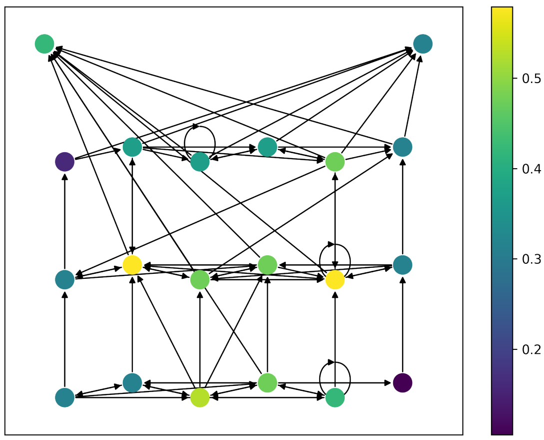
(c) Auto-organized - Open



(d) Prescriptive - Open

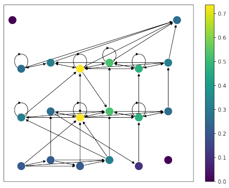


(e) Auto-organized - Foot

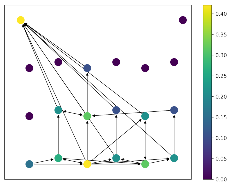


(f) Prescriptive - Foot

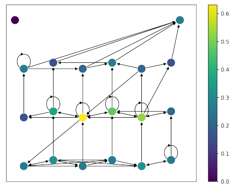
Figure 27: Degree centrality of subgraphs for (Pedagogy - Defense) during Intervention.



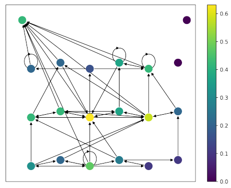
(a) Au.-org. - Tight - Success



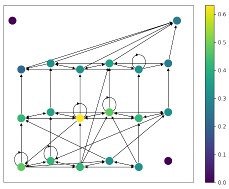
(b) Au.-org. - Tight - Fail



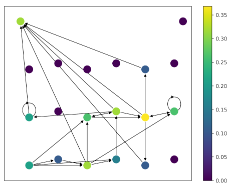
(e) Au.-org. - Open - Success



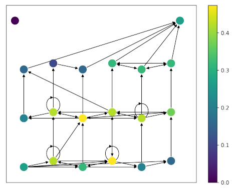
(f) Au.-org. - Open - Fail



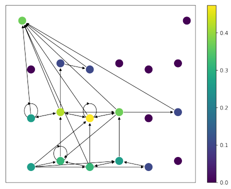
(c) Pres. - Tight - Success



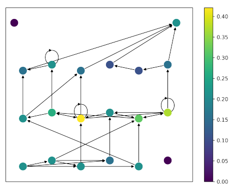
(d) Pres. - Tight - Fail



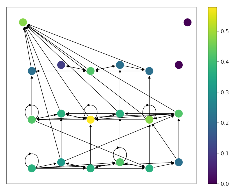
(g) Pres. - Open - Success



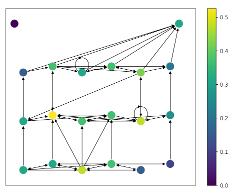
(h) Pres. - Open - Fail



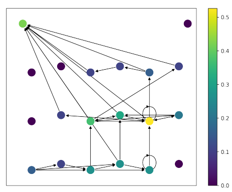
(i) Au.-org. - Foot - Success



(j) Au.-org. - Foot - Fail

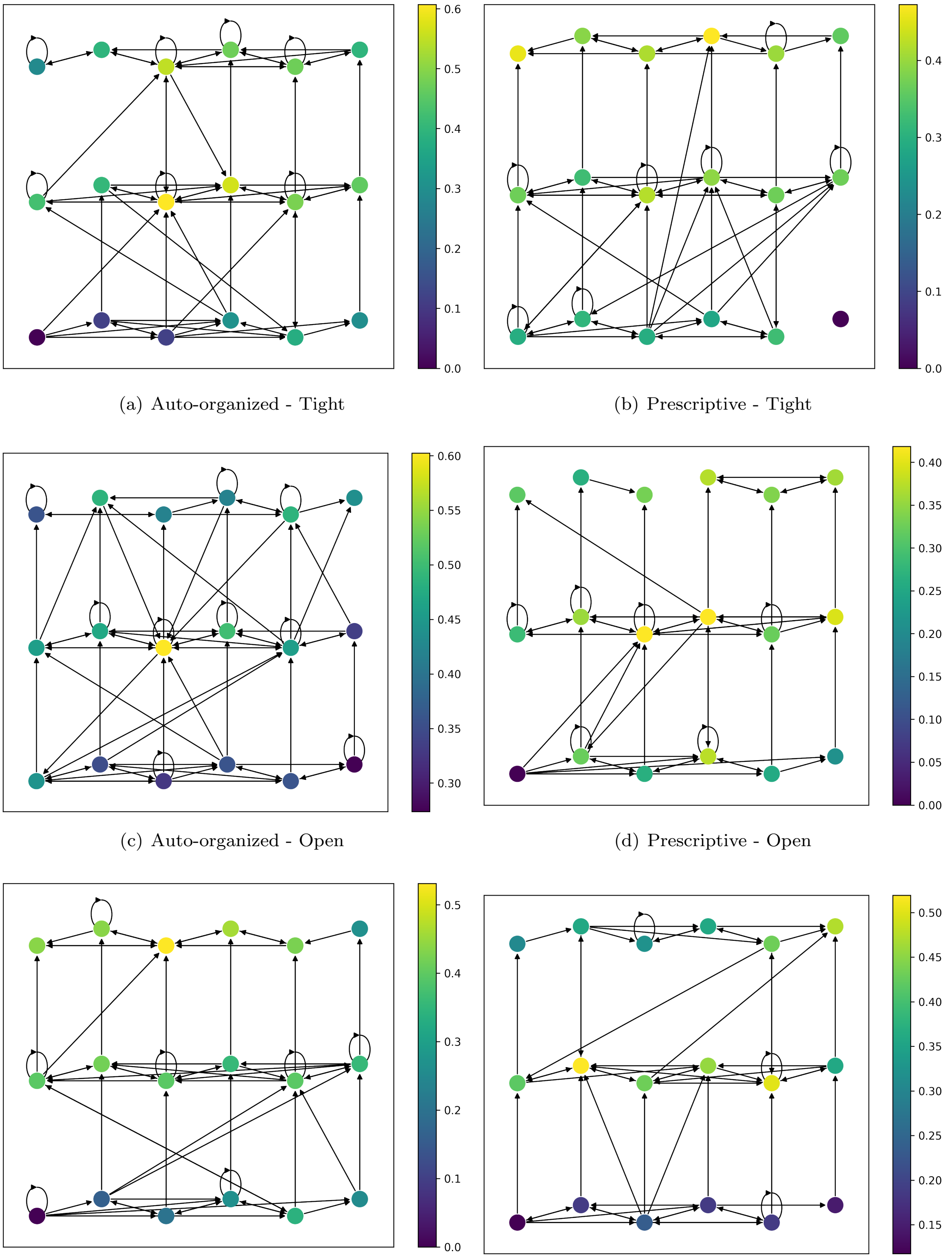


(k) Pres. - Foot - Success



(l) Pres. - Foot - Fail

Figure 28: Degree centrality of subgraphs of successful possessions (left columns) and failed possessions(right columns) for (Pedagogy - Defense) during Intervention.



(a) Auto-organized - Tight

(b) Prescriptive - Tight

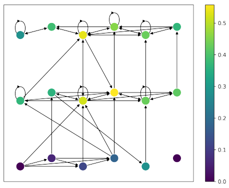
(c) Auto-organized - Open

(d) Prescriptive - Open

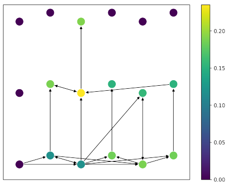
(e) Auto-organized - Foot

(f) Prescriptive - Foot

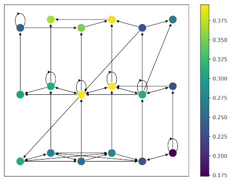
Figure 29: Closeness centrality of subgraphs for (Pedagogy - Defense) during Intervention.



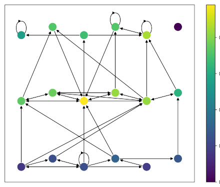
(a) Au.-org. - Tight - Success



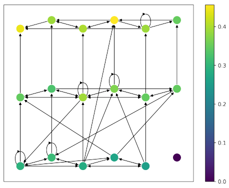
(b) Au.-org. - Tight - Fail



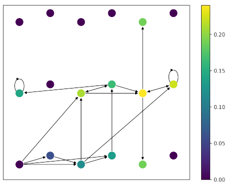
(e) Au.-org. - Open - Success



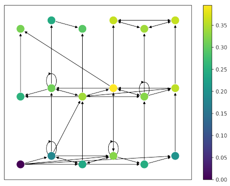
(f) Au.-org. - Open - Fail



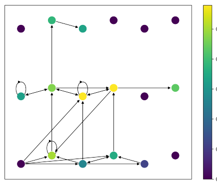
(c) Pres. - Tight - Success



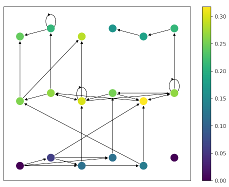
(d) Pres. - Tight - Fail



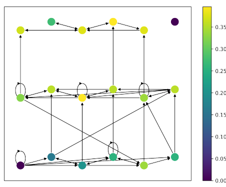
(g) Pres. - Open - Success



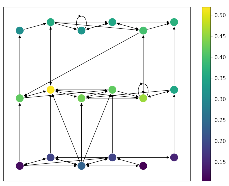
(h) Pres. - Open - Fail



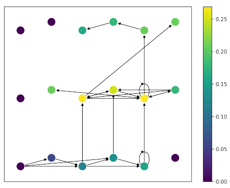
(i) Au.-org. - Foot - Success



(j) Au.-org. - Foot - Fail

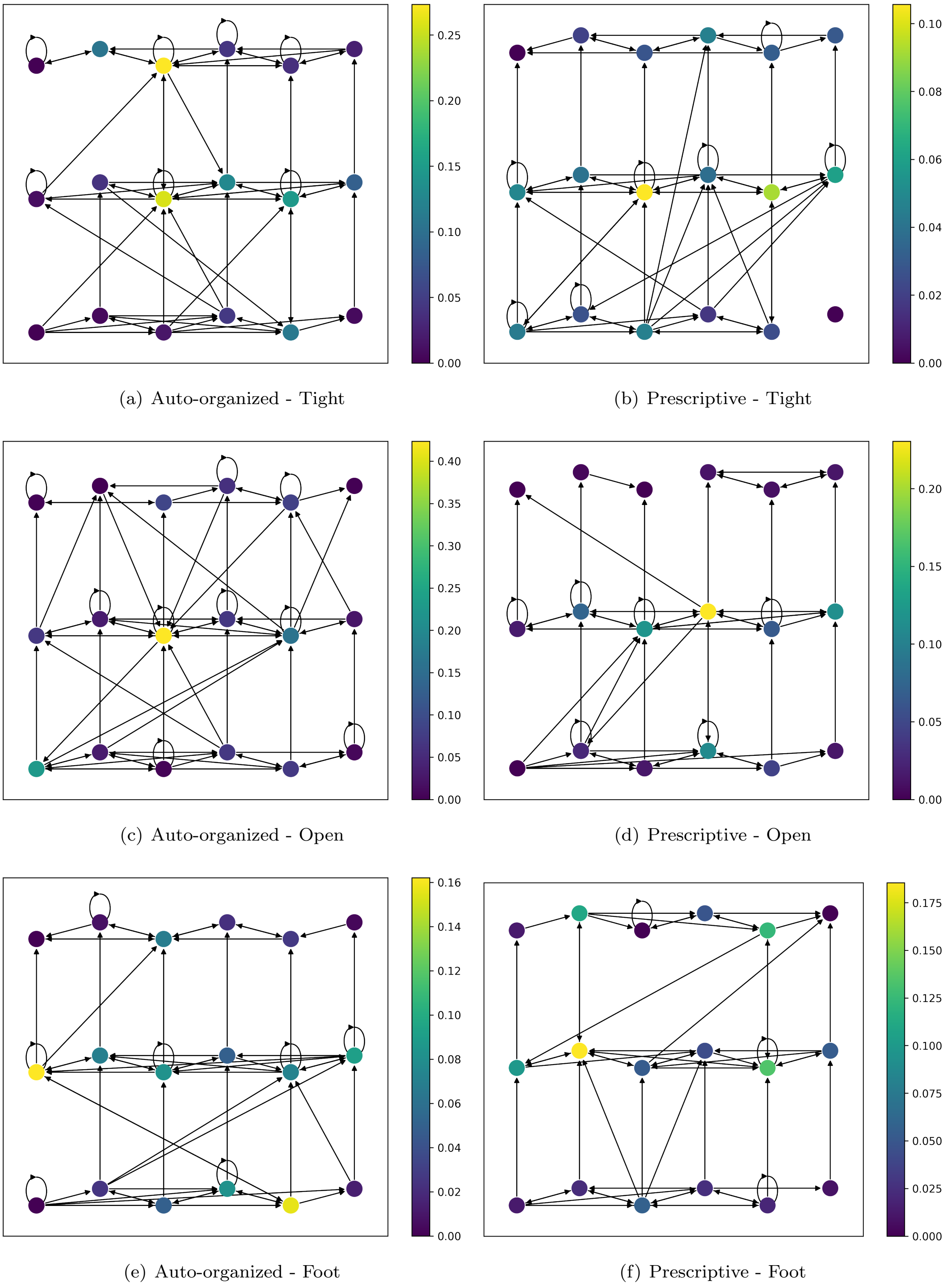


(k) Pres. - Foot - Success



(l) Pres. - Foot - Fail

Figure 30: Closeness centrality of subgraphs of successful possessions (left columns) and failed possessions(right columns) for (Pedagogy - Defense) during Intervention.



(a) Auto-organized - Tight

(b) Prescriptive - Tight

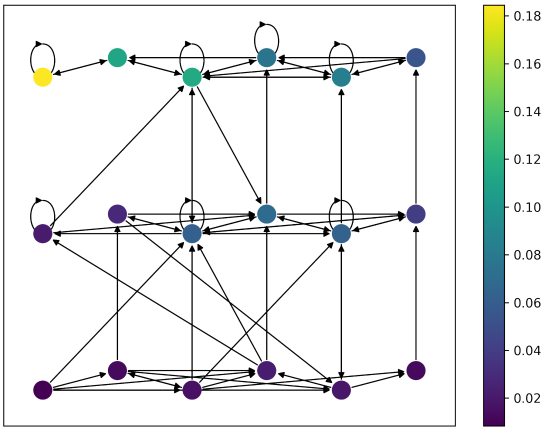
(c) Auto-organized - Open

(d) Prescriptive - Open

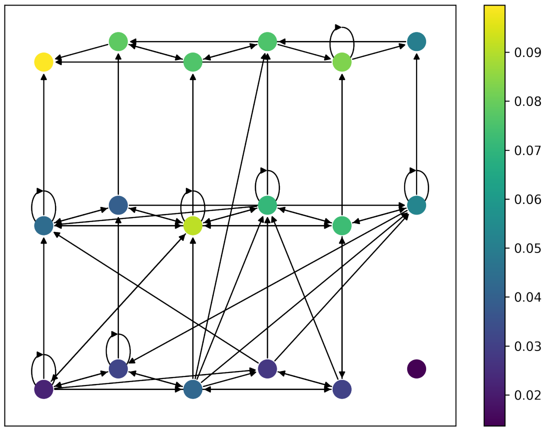
(e) Auto-organized - Foot

(f) Prescriptive - Foot

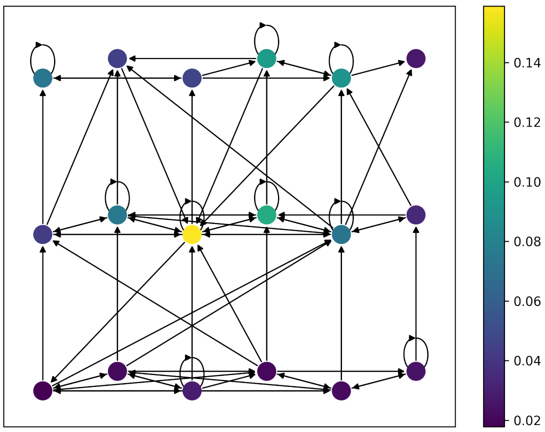
Figure 31: Betweenness centrality of subgraphs for (Pedagogy - Defense) during Intervention, without considering Target and NoTarget vertices.



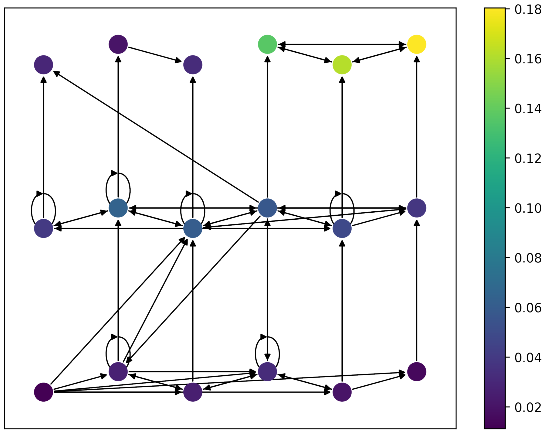
(a) Auto-organized - Tight



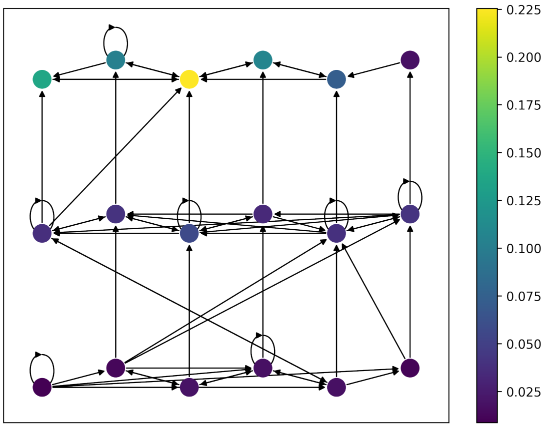
(b) Prescriptive - Tight



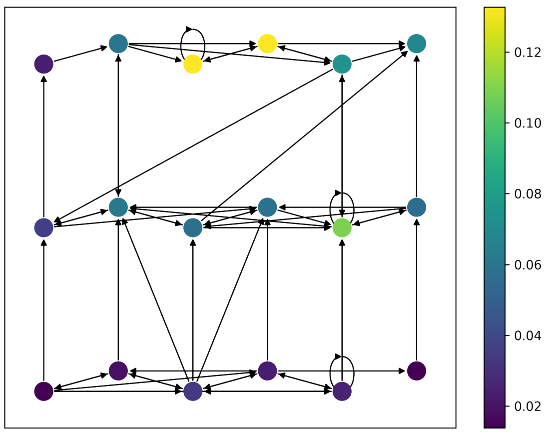
(c) Auto-organized - Open



(d) Prescriptive - Open



(e) Auto-organized - Foot



(f) Prescriptive - Foot

Figure 32: Pagerank centrality of subgraphs for (Pedagogy - Defense) during Intervention, without considering Target and NoTarget vertices.

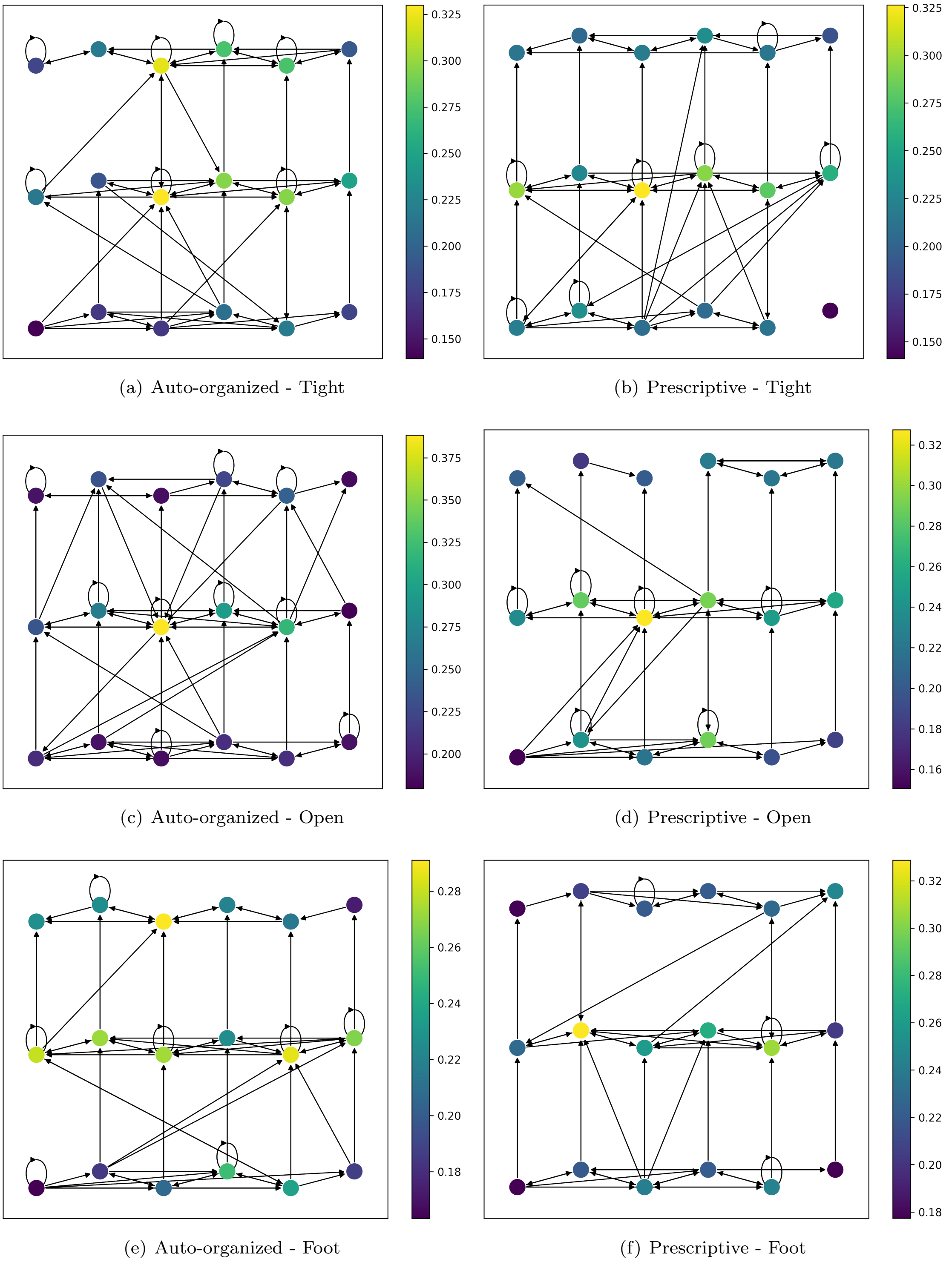


Figure 33: Katz centrality of subgraphs for (Pedagogy - Defense) during Intervention, without considering Target and NoTarget vertices.

References

Diestel, R. (2005). Graph theory 3rd ed. *Graduate texts in mathematics*, 173(33):12.

Jaccard, P. (1901). Étude comparative de la distribution florale dans une portion des alpes et des jura. *Bull Soc Vaudoise Sci Nat*, 37:547–579.

Newman, M. (2018). *Networks*. Oxford university press.

Sanfeliu, A. and Fu, K.-S. (1983). A distance measure between attributed relational graphs for pattern recognition. *IEEE transactions on systems, man, and cybernetics*, (3):353–362.

Van Steen, M. (2010). Graph theory and complex networks. *An introduction*, 144(1).