A theory of data patterns and visual analytics of football

Gennady and Natalia Andrienko

Fraunhofer Institute IAIS, Germany

City, University of London, UK

www.geoanalytics.net

Football data

Player ID	X	Y	time	Frame N	Speed
0001	8.61	3.35	18:31:03.400	10000	0
0001	8.6	3.34	18:31:03.440	10001	1.01
0001	8.58	3.34	18:31:03.480	10002	0.97
0001	8.57	3.33	18:31:03.520	10003	1.15
0001	8.56	3.32	18:31:03.560	10004	1.15
0001	8.55	3.31	18:31:03.600	10005	1.15
0001	8.54	3.3	18:31:03.640	10006	1.15
0001	8.52	3.28	18:31:03.680	10007	0.98
0001	8.52	3.28	18:31:03.720	10008	0.98
0001	8.52	3.26	18:31:03.760	10009	0.62
0001	8.51	3.26	18:31:03.800	10010	0.62
0001	8.51	3.25	18:31:03.840	10011	0.63
0001	8.52	3.23	18:31:03.880	10012	0.23
0001	8.52	3.23	18:31:03.920	10013	0.31
0001	8.54	3.23	18:31:03.960	10014	0.66
0001	8.55	3.23	18:31:04.000	10015	0.66
0001	8.56	3.24	18:31:04.040	10016	0.66
0001	8.57	3.26	18:31:04.080	10017	0.98
0001	8.58	3.28	18:31:04.120	10018	1.27
0001	8.59	3.3	18:31:04.160	10019	1.27
0001	8.6	3.32	18:31:04.200	10020	0.84
0001	8.61	3.33	18:31:04.240	10021	0.71
0001	8.62	3.34	18:31:04.280	10022	0.69
0001	8.62	3.35	18:31:04.320	10023	0.69
0001	8.64	3.36	18:31:04.360	10024	0.69
0001	8.64	3.35	18:31:04.400	10025	0.69
0001	8.64	3.36	18:31:04.440	10026	0.69
0001	8.62	3.35	18:31:04.480	10027	0.69
0001	8.61	3.36	18:31:04.520	10028	0.69
0001	8.6	3.37	18:31:04.560	10029	0.9
0001	8.59	3.37	18:31:04.600	10030	1.09
0001	8.58	3.38	18:31:04.640	10031	1.27
0001	8.56	3.39	18:31:04.680	10032	1.81

Trajectories

PUID1	PUID2	Shirt N 1	Shirt N 2	Time	X	Y	Type	Frame N	XRec	YRec	Frame N 2	Evaluation	duration
0027G6		20	0	20:47:01.920	0.2	0	Kickoff	10008					0
0027G6	000191	20	18	20:47:01.920	0.2	0	Pass	10008	-7.6	-4.31	10030	successfullyComplete	880
000191	0002F5	18	6	20:47:03.280	-8.2	-4.7	Pass	10042	-11.4	0.26	10062	successfullyComplete	800
0002F5	0000OJ	6	21	20:47:04.680	-12	0.2	Pass	10077	-7.87	10.34	10100	successfullyComplete	920
0000OJ	0002F5	21	6	20:47:05.640	-7.9	10.3	Pass	10101	-10.41	3.7	10118	successfullyComplete	680
0002F5	000191	6	18	20:47:07.440	-8.7	2.2	Pass	10146	-5.86	-6.66	10172	successfullyComplete	1040
000191	0001HT	18	15	20:47:13.960	0.6	-12.9	Pass	10309	-7.44	8.98	10393	successfullyComplete	3360
0001HT	0000ZS	15	4	20:47:18.920	-5	8.9	Pass	10433	-2.04	-11.55	10481	successfullyComplete	1920
0000ZS	000191	4	18	20:47:22.000	-2.2	-11.2	Pass	10510	4.78	-6.56	10529	successfullyComplete	760
000191	0000ZS	18	4	20:47:22.880	5.7	-6.4	Pass	10532	-1.41	-9.35	10557	successfullyComplete	1000
0000ZS	0001HT	4	15	20:47:24.000	-2.2	-9.8	Pass	10560	-1.05	11.52	10602	successfullyComplete	1680
0001HT	0027G6	15	20	20:47:28.280	4.4	14.1	Pass	10667	14.69	8.25	10690	successfullyComplete	920
0027G6	00258K	20	2	20:47:32.200	10.8	-2.6	Pass	10765	23.99	-30.42	10807	successfullyComplete	1680
00258K	000191	2	18	20:47:35.600	24.6	-29.1	Pass	10850	20	-14.58	10879	successfullyComplete	1160
000191	0000ZS	18	4	20:47:38.320	20.1	-19.2	Pass	10918	14.29	-25.28	11039	successfullyComplete	4840
0000ZS	0002AU	4	19	20:47:39.680	9.5	-22.1	Pass	10952	17.64	-14.43	10976	successfullyComplete	960
0002AU	0000ZS	19	4	20:47:40.640	18.3	-12.8	Pass	10976	14.29	-25.28	11039	successfullyComplete	2520
0000ZS	00258K	4	2	20:47:43.200	14.1	-25.2	Pass	11040	29.95	-31.78	11102	successfullyComplete	2480
00258K	0000ZS	2	4	20:47:45.680	30.1	-32.4	Pass	11102	17.78	-27.97	11130	successfullyComplete	1120
0000ZS	0002F5	4	6	20:47:47.640	18.3	-25.7	Pass	11151	4.48	13.9	11258	successfullyComplete	4280
0002F5	0001HT	6	15	20:47:55.480	5.1	8.9	Pass	11347	4.38	-1.92	11380	successfullyComplete	1320
0001HT	0000SE	15	14	20:47:57.720	5.9	-1.7	Pass	11403	37.03	29.05	11493	successfullyComplete	3600
002G08		23	0	20:48:01.200	37	27.2	OtherBallA	11490					0
002G08	0000SE	23	14	20:48:01.280	37.4	27.3	Ground	11492					0
0000SE		14	0	20:48:01.360	37.5	28.9	OtherBallA	11494					0
0000SE		14	0	20:48:04.840	39.4	34.3	ThrowIn	11581					0
0000SE	0002F5	14	6	20:48:04.840	39.4	34.3	Pass	11581	17.93	30.98	11629	successfullyComplete	1920
0002F5	0000OJ	6	21	20:48:08.120	12.7	28	Pass	11663	15.53	19.9	11681	successfullyComplete	720
0000OJ	0001HT	21	15	20:48:08.880	15.5	20.3	Pass	11682	-1.84	17.95	11721	successfullyComplete	1560
0001HT	0002F5	15	6	20:48:10.480	-1.8	18	Pass	11722	4.23	23.81	11781	successfullyComplete	2360
0002F5	0000ZS	6	4	20:48:12.880	4.1	23.7	Pass	11782	-2.68	-16.54	11851	successfullyComplete	2760
0000ZS	0027G6	4	20	20:48:19.600	4.6	-14.7	Pass	11950	13.25	-25.66	11993	successfullyComplete	1720
0027G6	00258K	20	2	20:48:23.400	23.1	-27.1	Pass	12045	31.1	-31.72	12068	successfullyComplete	920
00258K	0000OJ	2	21	20:48:26.200	28.7	-28.4	Pass	12115	20.82	-11.98	12150	successfullyComplete	1400

Events

Football data $\xrightarrow{?}$ Understand football tactics

Player ID	X	Y	time	Frame N	Speed
0001	8.61	3.35	18:31:03.400	10000	0
0001	8.6	3.34	18:31:03.440	10001	1.01
0001	8.58	3.34	18:31:03.480	10002	0.97
0001	8.57	3.33	18:31:03.520	10003	1.15
0001	8.56	3.32	18:31:03.560	10004	1.15
0001	8.55	3.31	18:31:03.600	10005	1.15
0001	8.54	3.3	18:31:03.640	10006	1.15
0001	8.52	3.28	18:31:03.680	10007	0.98
0001	8.52	3.28	18:31:03.720	10008	0.98
0001	8.52	3.26	18:31:03.760	10009	0.62
0001	8.51	3.26	18:31:03.800	10010	0.62
0001	8.51	3.25	18:31:03.840	10011	0.63
0001	8.52	3.23	18:31:03.880	10012	0.23
0001	8.52	3.23	18:31:03.920	10013	0.31
0001	8.54	3.23	18:31:03.960	10014	0.66
0001	8.55	3.23	18:31:04.000	10015	0.66
0001	8.56	3.24	18:31:04.040	10016	0.66
0001	8.57	3.26	18:31:04.080	10017	0.98
0001	8.58	3.28	18:31:04.120	10018	1.27
0001	8.59	3.3	18:31:04.160	10019	1.27
0001	8.6	3.32	18:31:04.200	10020	0.84
0001	8.61	3.33	18:31:04.240	10021	0.71
0001	8.62	3.34	18:31:04.280	10022	0.69
0001	8.62	3.35	18:31:04.320	10023	0.69
0001	8.64	3.36	18:31:04.360	10024	0.69
0001	8.64	3.35	18:31:04.400	10025	0.69
0001	8.64	3.36	18:31:04.440	10026	0.69
0001	8.62	3.35	18:31:04.480	10027	0.69
0001	8.61	3.36	18:31:04.520	10028	0.69
0001	8.6	3.37	18:31:04.560	10029	0.9
0001	8.59	3.37	18:31:04.600	10030	1.09
0001	8.58	3.38	18:31:04.640	10031	1.27
0001	8.56	3.39	18:31:04.680	10032	1.81

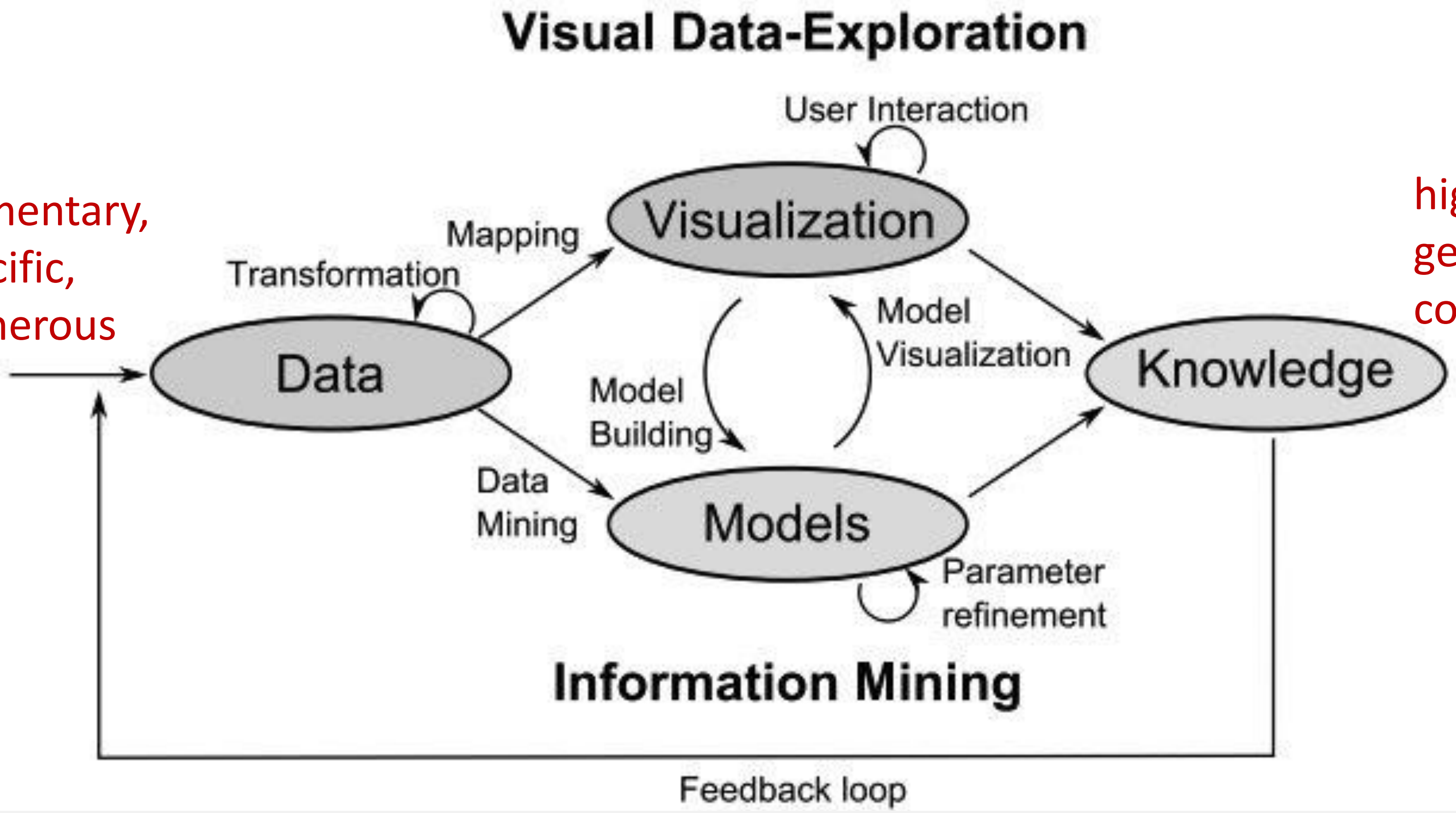
Trajectories

PUID1	PUID2	Shirt N 1	Shirt N 2	Time	X	Y	Type	Frame N	XRec	YRec	Frame N 2	Evaluation	duration
0027G6		20	0	20:47:01.920	0.2	0	Kickoff	10008					0
0027G6	000191	20	18	20:47:01.920	0.2	0	Pass	10008	-7.6	-4.31	10030	successfullyComplete	880
000191	0002F5	18	6	20:47:03.280	-8.2	-4.7	Pass	10042	-11.4	0.26	10062	successfullyComplete	800
0002F5	0000OJ	6	21	20:47:04.680	-12	0.2	Pass	10077	-7.87	10.34	10100	successfullyComplete	920
0000OJ	0002F5	21	6	20:47:05.640	-7.9	10.3	Pass	10101	-10.41	3.7	10118	successfullyComplete	680
0002F5	000191	6	18	20:47:07.440	-8.7	2.2	Pass	10146	-5.86	-6.66	10172	successfullyComplete	1040
000191	0001HT	18	15	20:47:13.960	0.6	-12.9	Pass	10309	-7.44	8.98	10393	successfullyComplete	3360
0001HT	0000ZS	15	4	20:47:18.920	-5	8.9	Pass	10433	-2.04	-11.55	10481	successfullyComplete	1920
0000ZS	000191	4	18	20:47:22.000	-2.2	-11.2	Pass	10510	4.78	-6.56	10529	successfullyComplete	760
000191	0000ZS	18	4	20:47:22.880	5.7	-6.4	Pass	10532	-1.41	-9.35	10557	successfullyComplete	1000
0000ZS	0001HT	4	15	20:47:24.000	-2.2	-9.8	Pass	10560	-1.05	11.52	10602	successfullyComplete	1680
0001HT	0027G6	15	20	20:47:28.280	4.4	14.1	Pass	10667	14.69	8.25	10690	successfullyComplete	920
0027G6	00258K	20	2	20:47:32.200	10.8	-2.6	Pass	10765	23.99	-30.42	10807	successfullyComplete	1680
00258K	000191	2	18	20:47:35.600	24.6	-29.1	Pass	10850	20	-14.58	10879	successfullyComplete	1160
000191	0000ZS	18	4	20:47:38.320	20.1	-19.2	Pass	10918	14.29	-25.28	11039	successfullyComplete	4840
0000ZS	0002AU	4	19	20:47:39.680	9.5	-22.1	Pass	10952	17.64	-14.43	10976	successfullyComplete	960
0002AU	0000ZS	19	4	20:47:40.640	18.3	-12.8	Pass	10976	14.29	-25.28	11039	successfullyComplete	2520
0000ZS	00258K	4	2	20:47:43.200	14.1	-25.2	Pass	11040	29.95	-31.78	11102	successfullyComplete	2480
00258K	0000ZS	2	4	20:47:45.680	30.1	-32.4	Pass	11102	17.78	-27.97	11130	successfullyComplete	1120
0000ZS	0002F5	4	6	20:47:47.640	18.3	-25.7	Pass	11151	4.48	13.9	11258	successfullyComplete	4280
0002F5	0001HT	6	15	20:47:55.480	5.1	8.9	Pass	11347	4.38	-1.92	11380	successfullyComplete	1320
0001HT	0000SE	15	14	20:47:57.720	5.9	-1.7	Pass	11403	37.03	29.05	11493	successfullyComplete	3600
002G08		23	0	20:48:01.200	37	27.2	OtherBallA	11490					0
002G08	0000SE	23	14	20:48:01.280	37.4	27.3	Ground	11492					0
0000SE		14	0	20:48:01.360	37.5	28.9	OtherBallA	11494					0
0000SE		14	0	20:48:04.840	39.4	34.3	ThrowIn	11581					0
0000SE	0002F5	14	6	20:48:04.840	39.4	34.3	Pass	11581	17.93	30.98	11629	successfullyComplete	1920
0002F5	0000OJ	6	21	20:48:08.120	12.7	28	Pass	11663	15.53	19.9	11681	successfullyComplete	720
0000OJ	0001HT	21	15	20:48:08.880	15.5	20.3	Pass	11682	-1.84	17.95	11721	successfullyComplete	1560
0001HT	0002F5	15	6	20:48:10.480	-1.8	18	Pass	11722	4.23	23.81	11781	successfullyComplete	2360
0002F5	0000ZS	6	4	20:48:12.880	4.1	23.7	Pass	11782	-2.68	-16.54	11851	successfullyComplete	2760
0000ZS	0027G6	4	20	20:48:19.600	4.6	-14.7	Pass	11950	13.25	-25.66	11993	successfullyComplete	1720
0027G6	00258K	20	2	20:48:23.400	23.1	-27.1	Pass	12045	31.1	-31.72	12068	successfullyComplete	920
00258K	0000OJ	2	21	20:48:26.200	28.7	-28.4	Pass	12115	20.82	-11.98	12150	successfullyComplete	1400

Events

General task of data analysis: data → knowledge

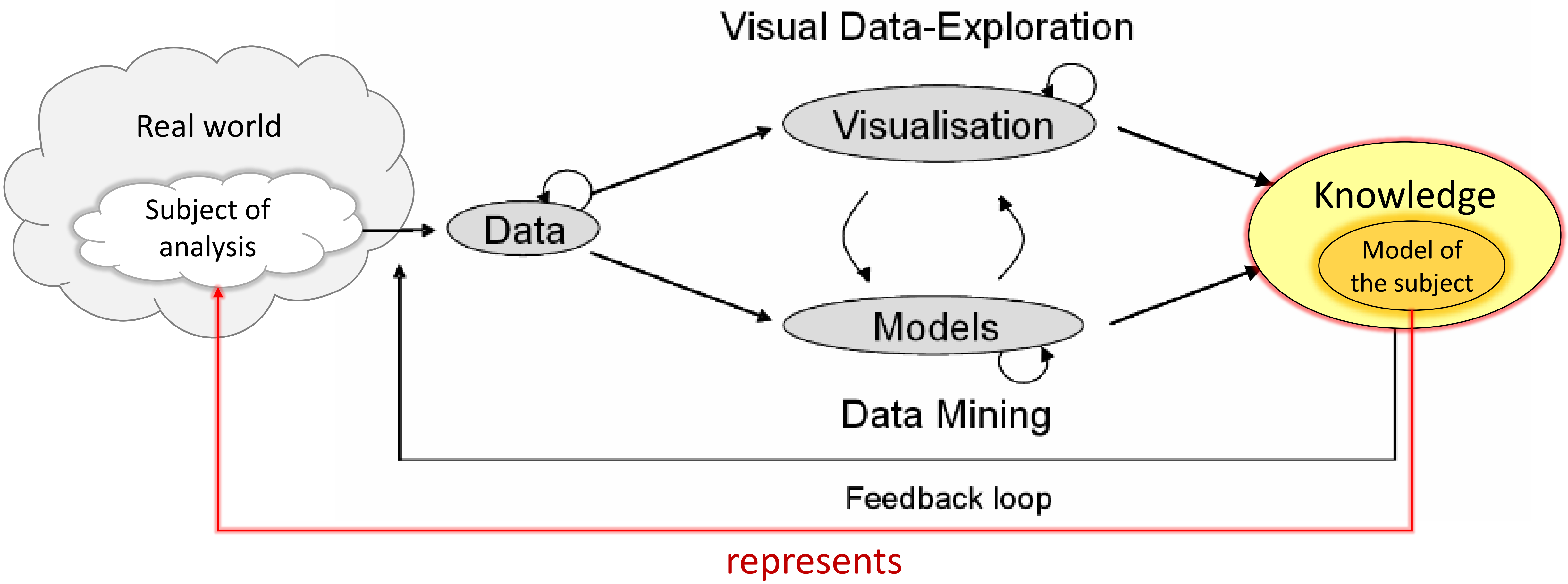
elementary,
specific,
numerous



high-level,
general,
concise

Keim D., Andrienko G., Fekete J.D., Görg C., Kohlhammer J., Melançon G. (2008)
Visual Analytics: Definition, Process, and Challenges.
In: Kerren A., Stasko J.T., Fekete JD., North C. (eds) Information Visualization. Lecture Notes in Computer Science, vol 4950.
Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-70956-5_7

Knowledge = model of reality



Andrienko, N., Lammarsch, T., Andrienko, G., Fuchs, G., Keim, D., Miksch, S. and Rind, A. (2018)
Viewing Visual Analytics as Model Building.
Computer Graphics Forum, 37: 275-299. https://doi.org/10.1111/cgf.13324

Model

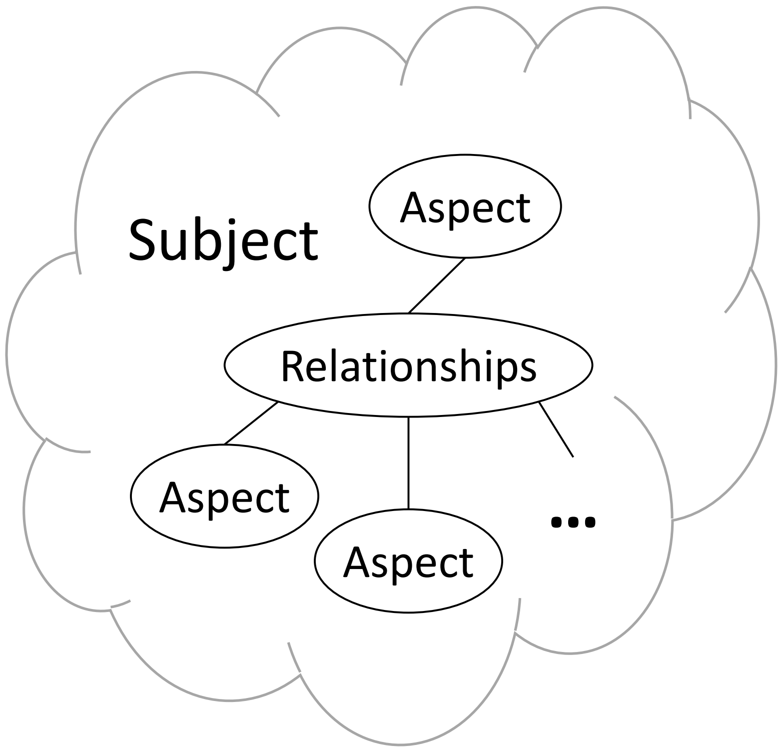
- “a schematic description or representation of something, especially a system or phenomenon*, that accounts for its properties and is used to study its characteristics”

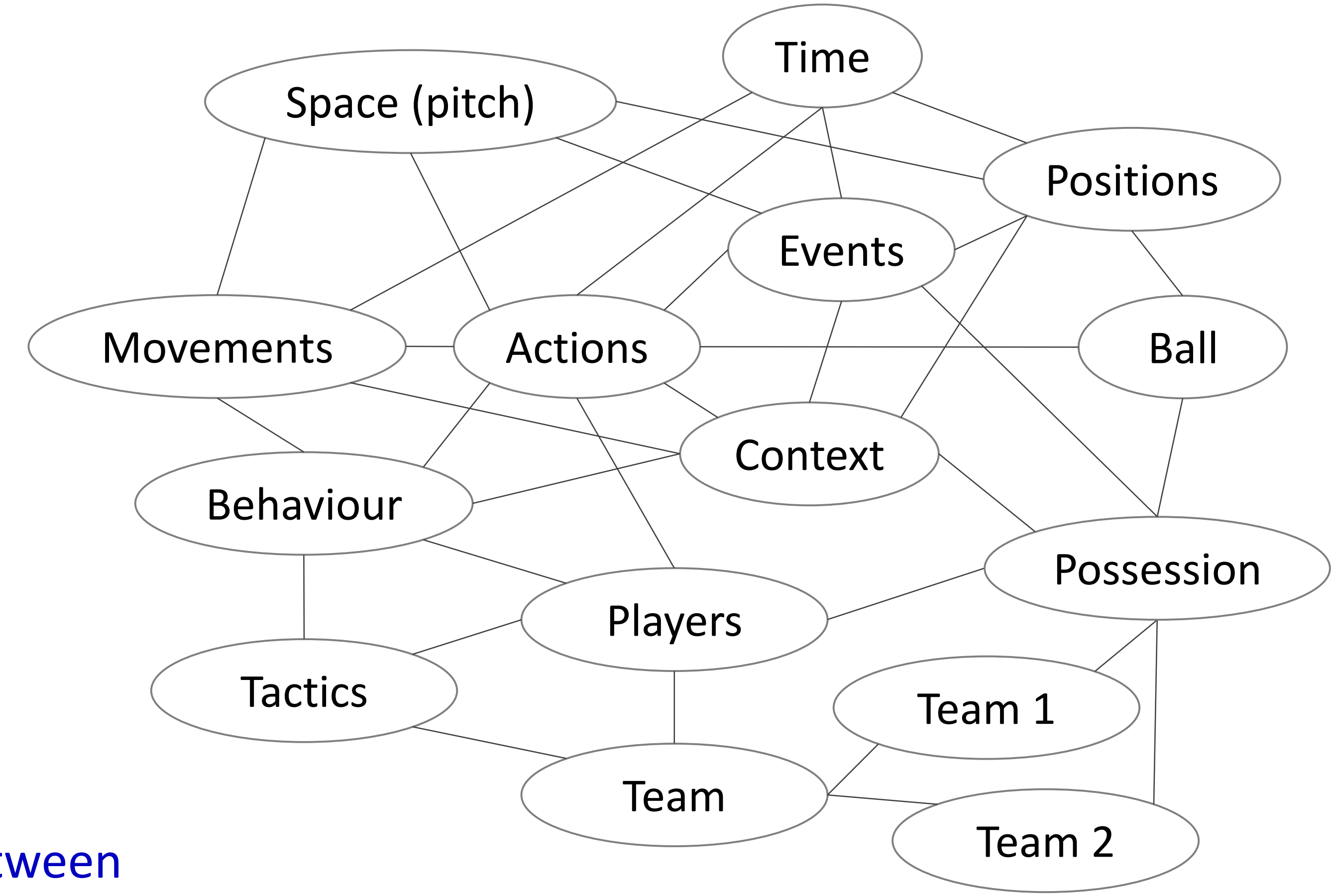
  (* = subject of analysis)

  The American Heritage Dictionary of the English Language, Fifth Edition. Houghton Mifflin Harcourt Trade, 2011.

- A model of a subject is built by analysing data reflecting the subject.
- A model needs to represent the subject and not the data.
- A model is a *generalisation* from data.
  - Its scope extends beyond the part of the subject directly reflected in the data.
- A model is a *simplified* representation of a subject.
  - Subjects are typically too complex to be represented fully.

Subject of analysis viewed as a system





A model needs to represent relationships between aspects or components of the analysis subject.

Need for abstraction

What we have: elementary connections

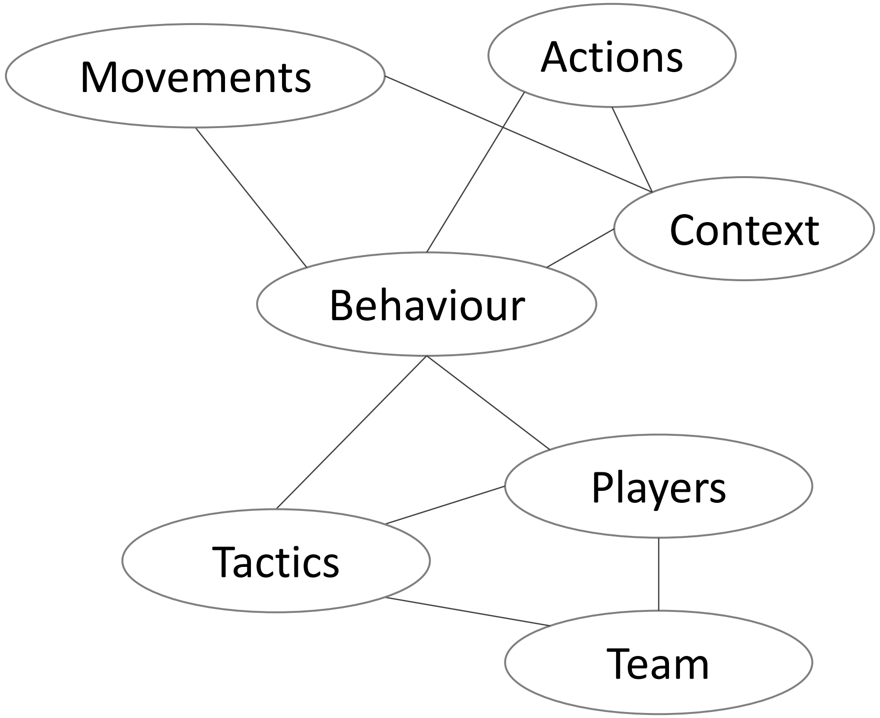
Player ID	X	Y	time	Frame N	Speed
0001	8.61	3.35	18:31:03.400	10000	0
0001	8.6	3.34	18:31:03.440	10001	1.01
0001	8.58	3.34	18:31:03.480	10002	0.97
0001	8.57	3.33	18:31:03.520	10003	1.15
0001	8.56	3.32	18:31:03.560	10004	1.15
0001	8.55	3.31	18:31:03.600	10005	1.15
0001	8.54	3.3	18:31:03.640	10006	1.15
0001	8.52	3.28	18:31:03.680	10007	0.98
0001	8.52	3.28	18:31:03.720	10008	0.98
0001	8.52	3.26	18:31:03.760	10009	0.62
0001	8.51	3.26	18:31:03.800	10010	0.62
0001	8.51	3.25	18:31:03.840	10011	0.63

. . .

Multiple data items need to be considered together as a <u>unified whole</u> (= a **data pattern**).

Transformation operation: **abstraction**

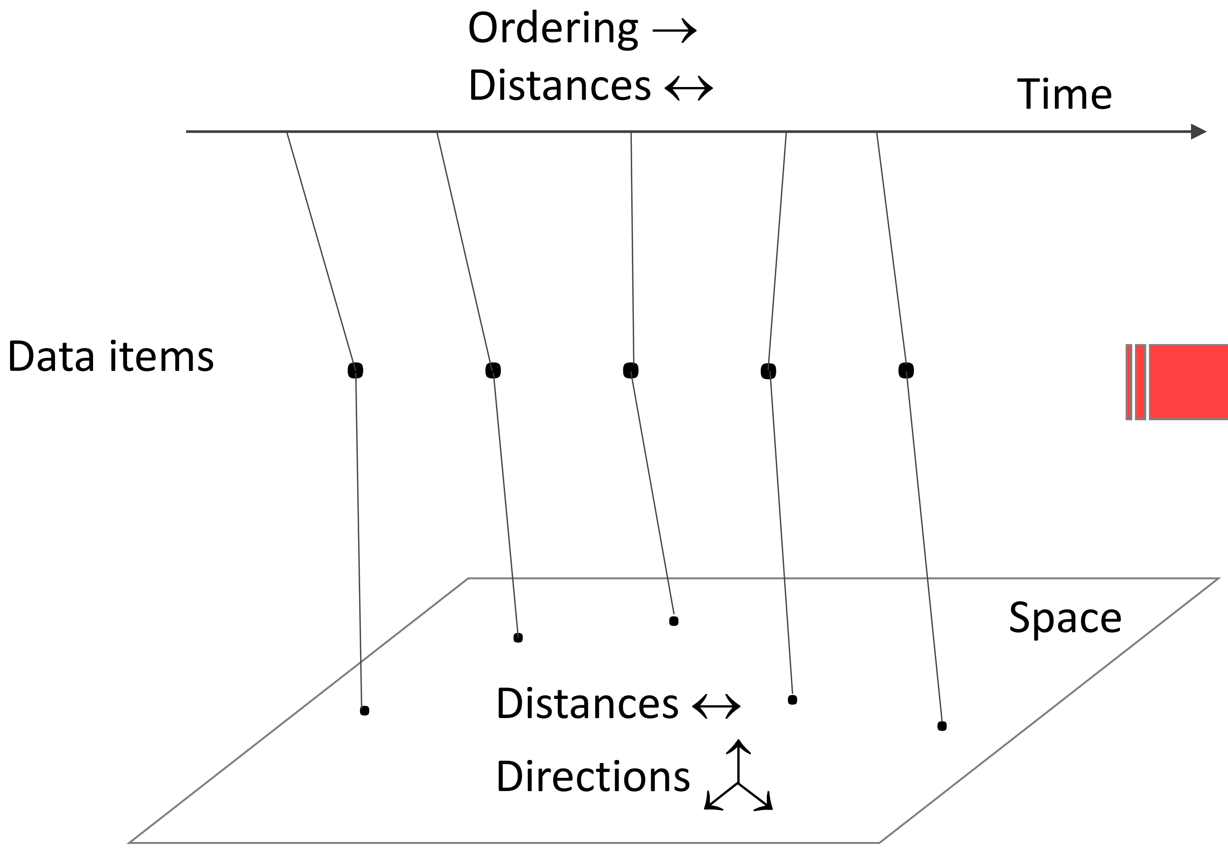
What we need: overall relationships and high-level constructs

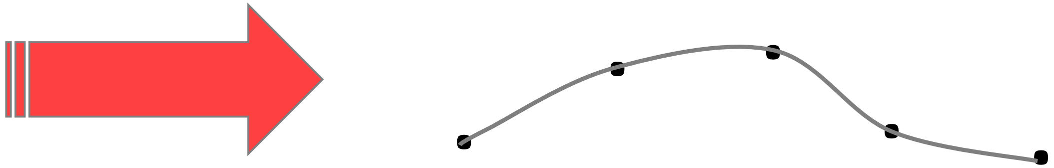


Abstraction and pattern

- Abstraction is the process of deriving ***general*** concepts, principles, or relationships ***from specific*** examples (instances).
  - Abstraction = “seeing the forest for the trees”
- In terms of data, an ***instance*** is a combination of elements of different components represented by one data record.
- A ***pattern*** is a combination of multiple instances that can be described together as a meaningful whole, without referring to the individual instances.
- ***Finding patterns in data*** is a way to achieve abstraction in data analysis.

Elementary relationships as a medium of unification

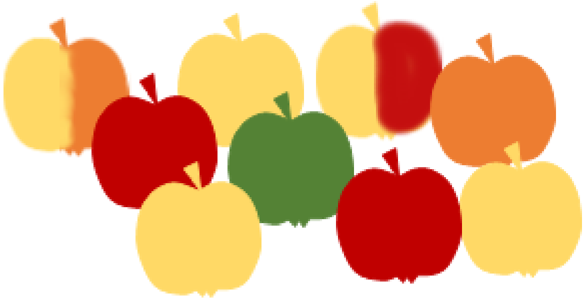




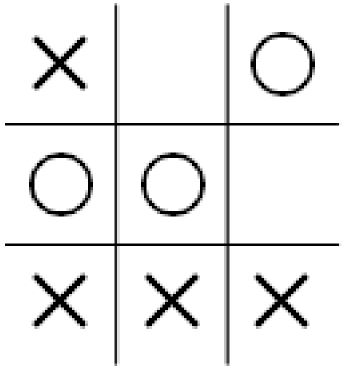
Relationships that exist between elements of subject components allow us to consider multiple data items as a single object.

Data distribution

Distribution of colours **over** a set of apples

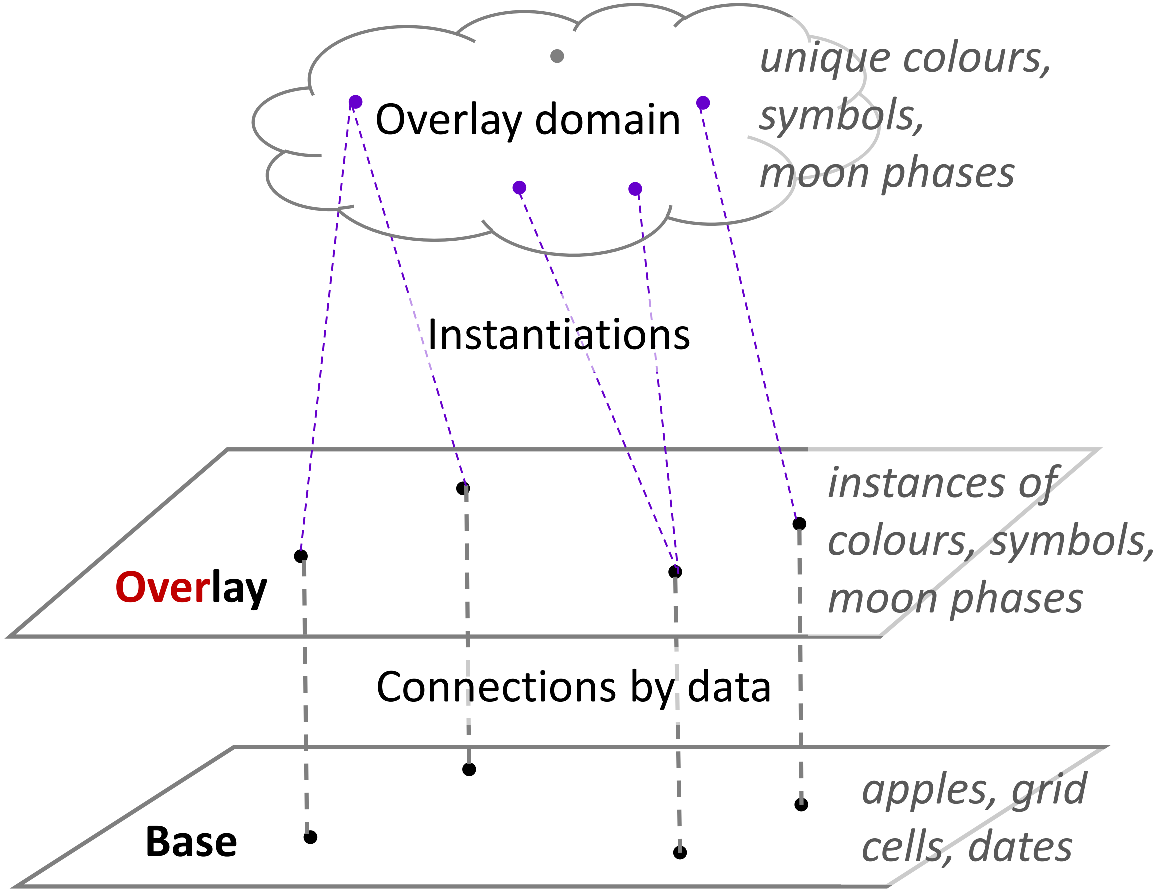


Distribution of X and O symbols **over** a grid



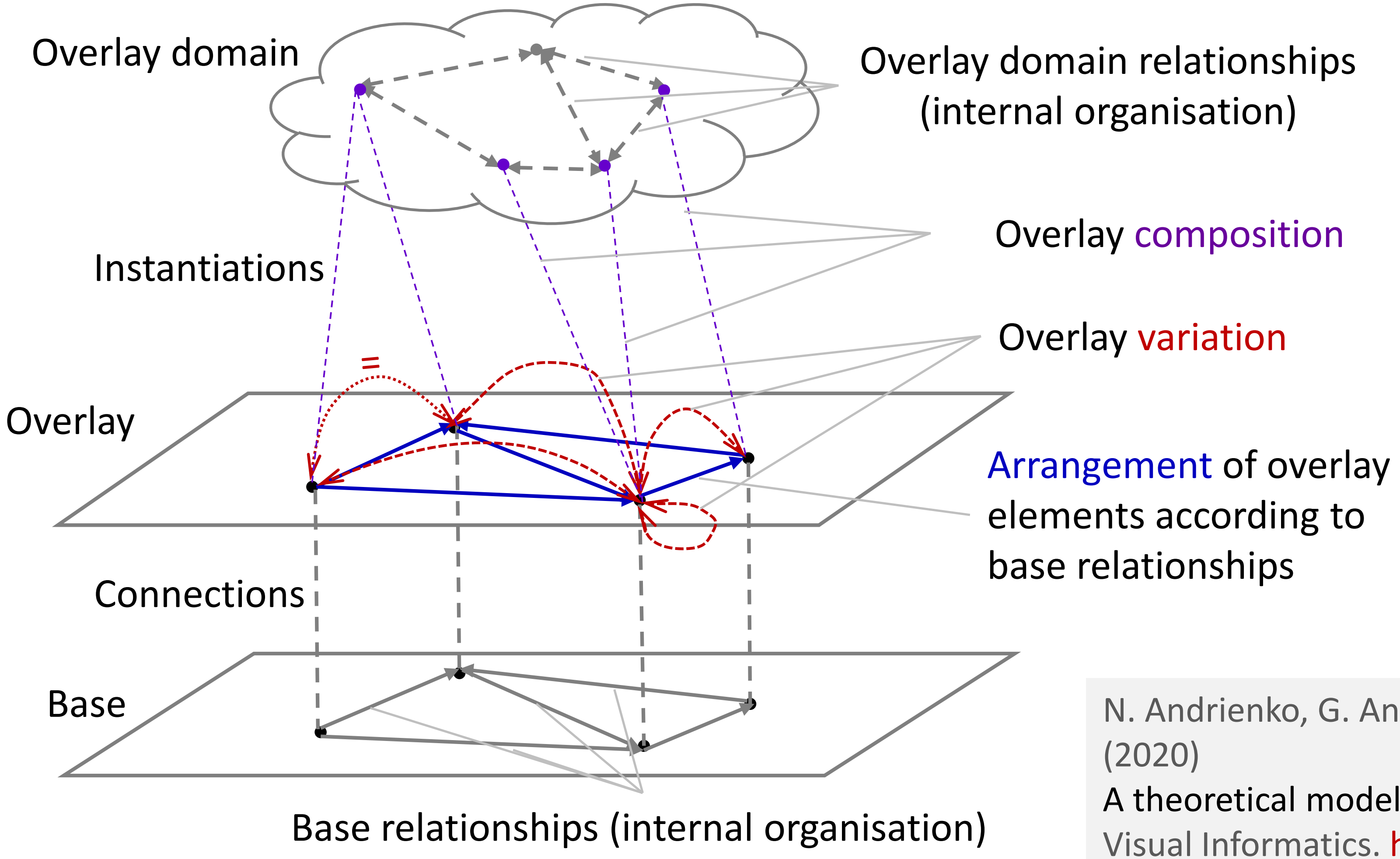
Distribution of moon phases **over** time

Examples



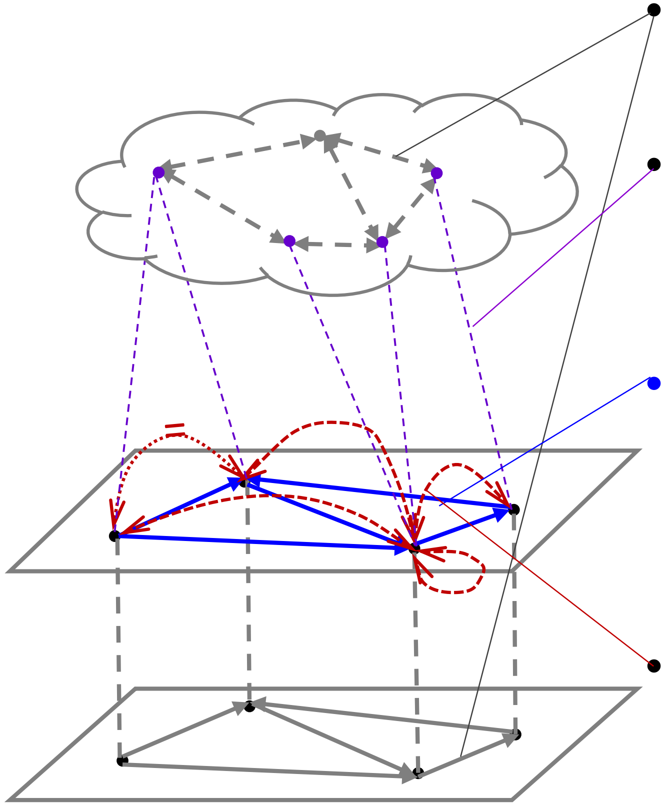
Schematic representation

Aspects of a data distribution



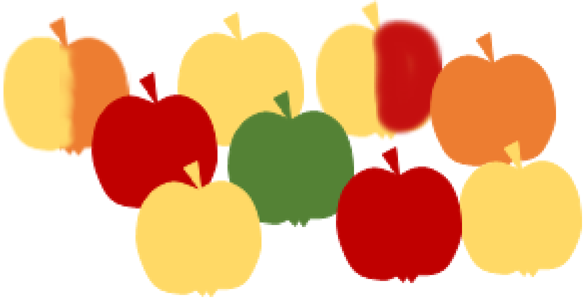
N. Andrienko, G. Andrienko, S. Miksch, H. Schumann, S. Wrobel (2020)
A theoretical model for pattern discovery in visual analytics.
Visual Informatics. https://doi.org/10.1016/j.visinf.2020.12.002

Definitions



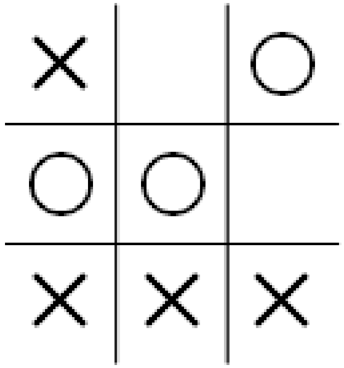
- The set of all *relationships* existing *between elements of a data component* is called the **organisation** of this data component.
- The **composition** of the overlay of a data distribution is the set of *instantiation relationships* between the elements of the overlay and their prototypes in the domain of the overlay.
- Arrangement relationships between elements of the overlay of a data distribution are the *relationships between the corresponding elements of the base*. The **arrangement** of the overlay of a data distribution is *the set of the arrangement relationships* between the overlay elements.
- The **variation** of the overlay of a distribution with respect to the base consists of the *domain-pertinent relationships between the overlay elements* (i.e., relationships belonging to the organisation of the overlay domain) considered *in connection to the arrangement relationships* between the overlay elements.

Patterns in data distributions (examples)



Yellow is the most frequent colour.

Yellow sometimes co-occurs with orange or red.



There are 4 crosses and 3 noughts.

All but one symbols make a triangle.

There are three crosses in a row.

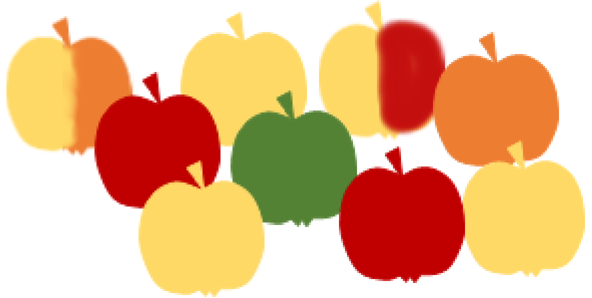
Patterns may refer to:

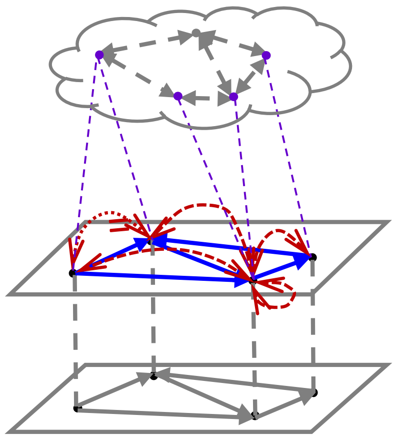
- overlay composition (frequencies of element occurrences)
- arrangement of overlay elements (co-occurrence, spatial shape, temporal sequence)
- overlay variation with respect to the arrangement (equivalence, differences, changes)

Phase sequence: new moon – first quarter – full moon – third quarter – new moon and so on.

Moon phases repeat every 4 weeks.

Patterns in data distributions (examples)



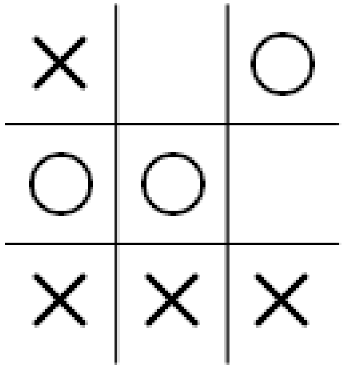


Yellow is the most frequent colour.

Yellow sometimes co-occurs with orange or red.

Relationships link multiple elements and multiple elementary connections into constructs that can be considered and described as single objects: groups, shapes, trends, periodic patterns, etc.

Treating multiple items as one unit means performing ***abstraction***.



There are 4 crosses and 3 noughts.

All but one symbols make a triangle.

There are three crosses in a row.

Phase sequence: new moon – first quarter – full moon – third quarter – new moon and so on.

Moon phases repeat every 4 weeks.

Definition of a pattern

- A **pattern in a data distribution** is a subset of the *relationships* involved in the composition, arrangement, or variation of the overlay over the base such that there exists an operation of *abstraction* allowing to treat this subset as a *unified whole*.
- An **abstracted data pattern** is a representation of an objective pattern as a unified whole regardless of the form, language, and medium of the representation.
  - In particular, it may be a representation in human's mind.

Pattern discovery by computers and by humans

- It is possible to create computer programs for pattern discovery, but this requires precise and detailed specifications and instructions:
  - what items can be put together (what they need to have in common or how they need to be related), how different they need to be from others, how big a group must be, …
  - Hence, any computer program can only find pre-specified types of patterns.
- Humans have an intrinsic capability to ***see*** various kinds of patterns without precise specifications given to them.
  - Human visual perception tends to unite multiple things into shapes and group similar things separating them from others.
  - Humans can find patterns in data when data items are represented visually.

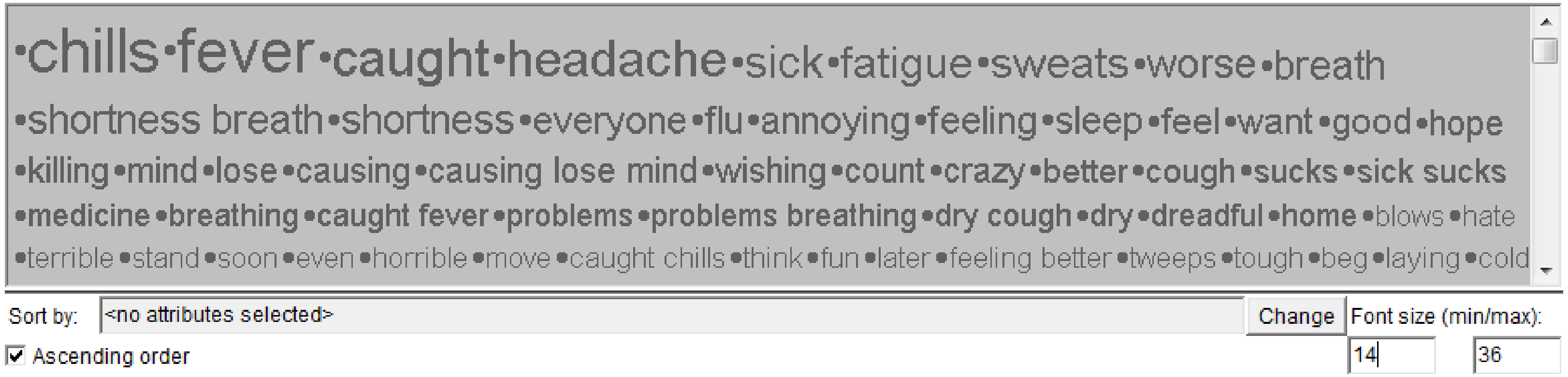
Relationships: visible and real

- Humans unite multiple visible things into patterns due to certain ***relationships*** between the things, such as proximity, similarity, alignment or other spatial arrangements.
- Patterns in data are made by relationships between data items, such as closeness and ordering (smaller/greater) of values or similarity of value combinations.
- For visual discovery of patterns existing in data, <u>relationships between data items need to be faithfully represented by relationships between visual items</u>.
  - Visual representation should enable perception of real relationships between data items and disable seeing of non-existing relationships.
  - This requirement underlies the <u>principles and rules of data visualisation</u>.

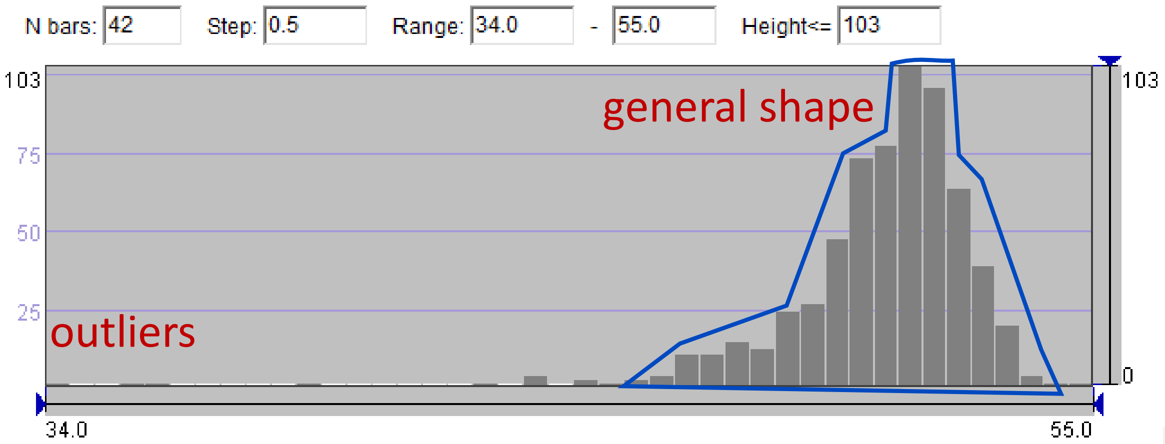
Pattern types

- <u>Composition</u> patterns
  - Patterns of frequency or probability distribution of numeric values: normal, skewed, long-tailed, …
  - Patterns of high or low frequency of categorical values or discrete entities
- <u>Arrangement</u> patterns: formed by relationships between base elements, distinguished according to the base relationships involved (e.g., ordering, distance, hierarchy, …)
  - Spatial arrangements: cluster, alignment, high or low density, …
  - Temporal arrangements: high or low temporal frequency, regular spacing, periodic re-occurrence, …
  - Co-occurrences: multiple overlay elements connected to the same base element
- <u>Variation</u> patterns: formed by relationships between base elements and between corresponding overlay elements; distinguished according to the base and overlay relationships involved
  - Variation over space: same/similar overlay elements at close locations, increase/decrease in some directions, …
  - Variation over time: trend, peak, constancy, …
  - Variation w.r.t. co-occurrences: which overlay elements tend to co-occur, which never co-occur, …

Pattern examples: frequency



Base: set of tweets
Overlay: set of words
Pattern types:
frequently occurring elements, absence of occurrence of particular elements

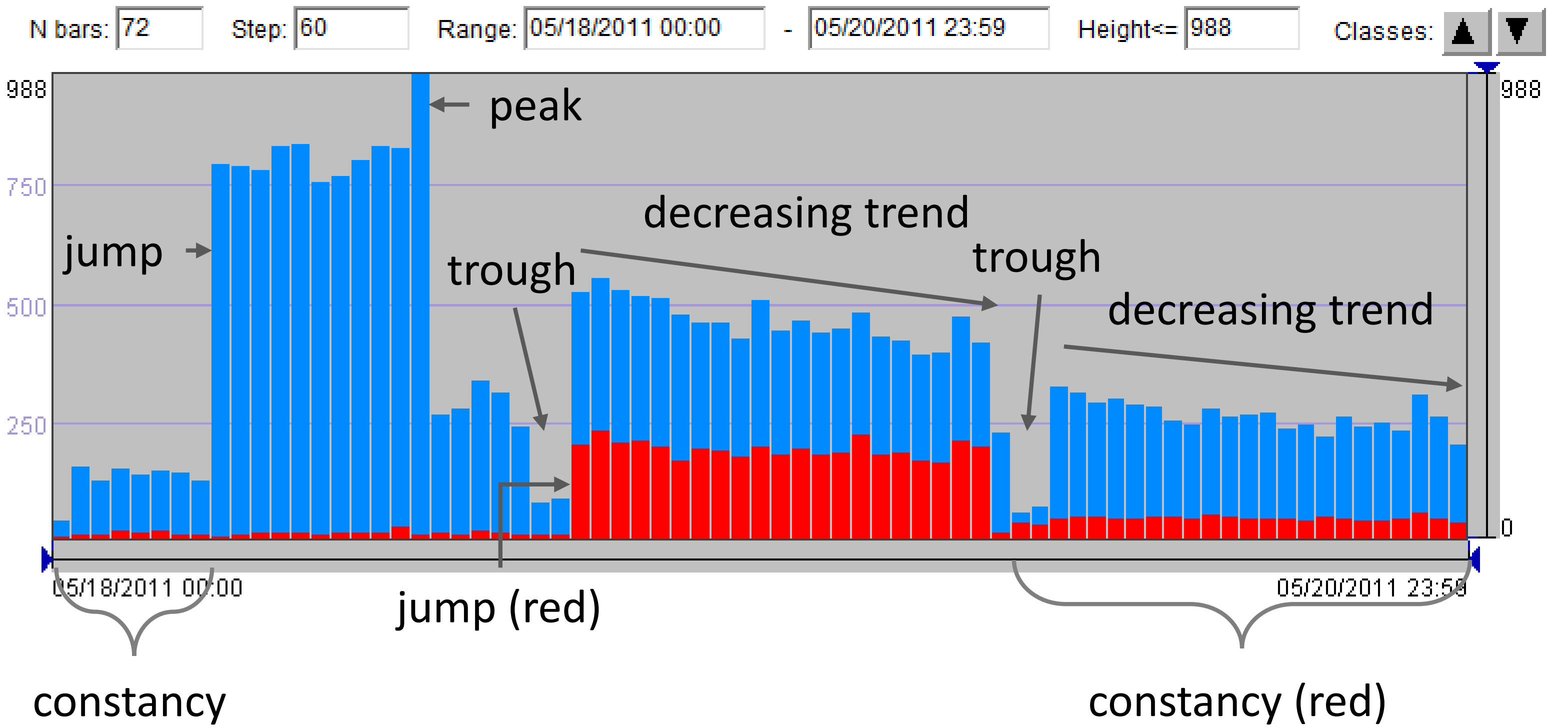


Base: set of wards in London
Overlay: set of values of a numeric attribute (% of female inhabitants)
Pattern types:
uniform or non-uniform frequency, interval(s) of most frequent values, prevalence of high/low/medium values, …

Andrienko, N., Andrienko, G., Fuchs, G., Slingsby, A., Turkay, C., Wrobel, S (2020): Visual Analytics for Data Scientists. Springer https://doi.org/10.1007/978-3-030-56146-8

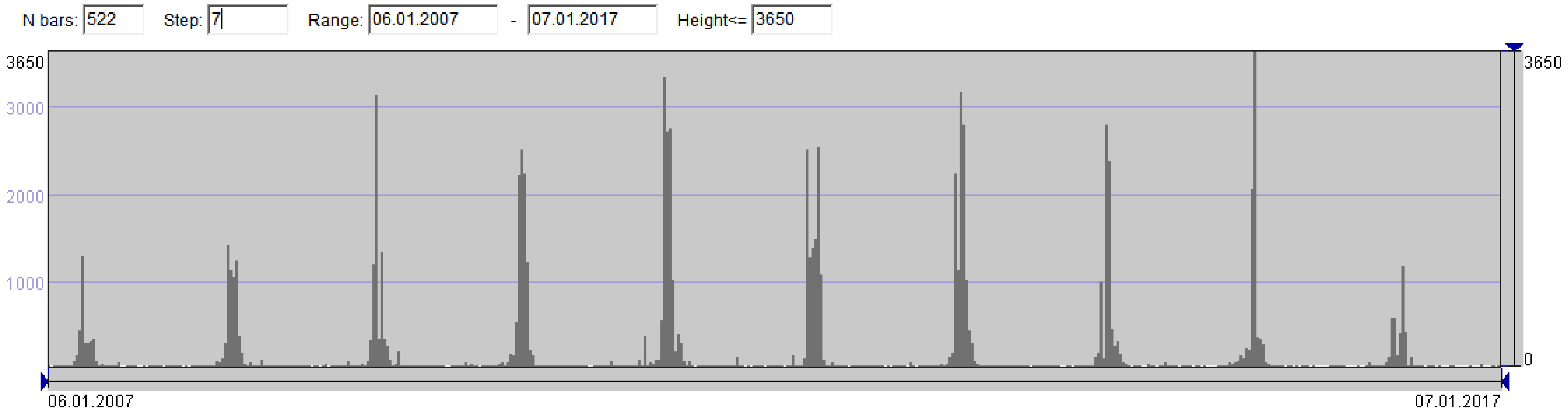
Pattern examples: temporal arrangement and variation

Epidemic outbreak development

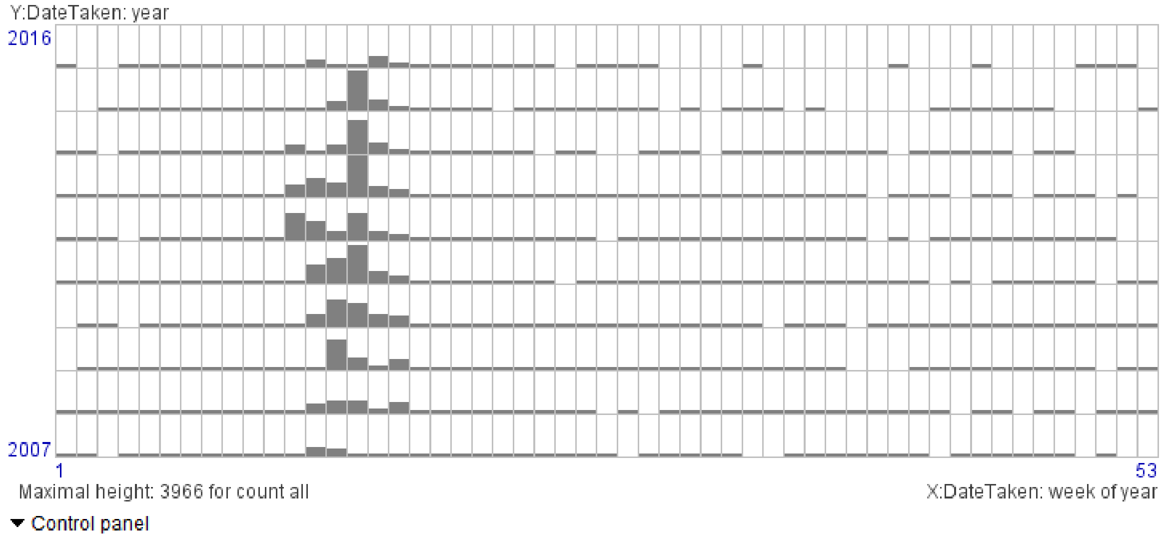


Pattern example: periodic temporal arrangement

Photos of cherry blossoming published in flickr

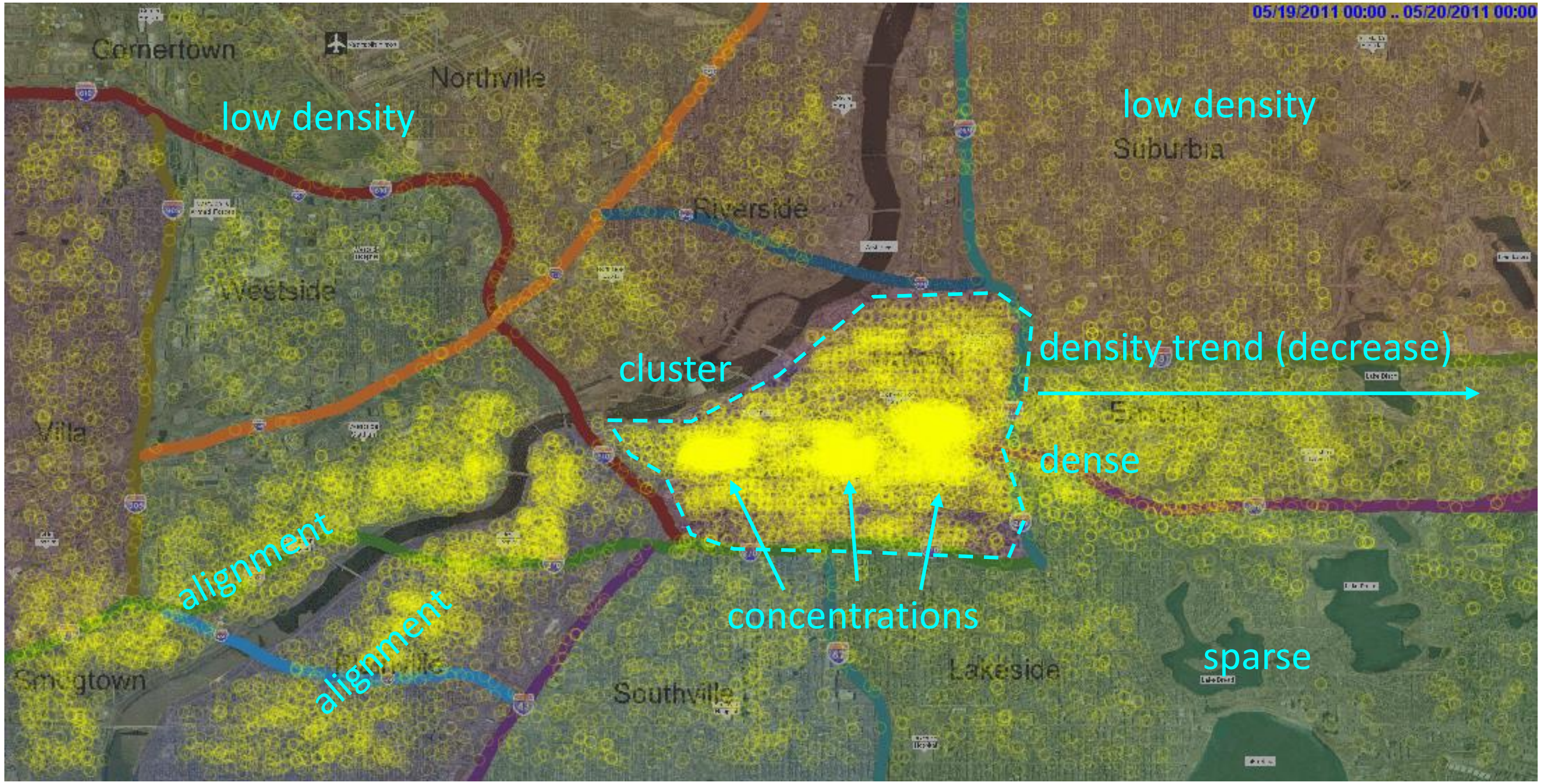


2d time histogram
Rows: years
Columns: weeks
Bar sizes:
counts of taken photos

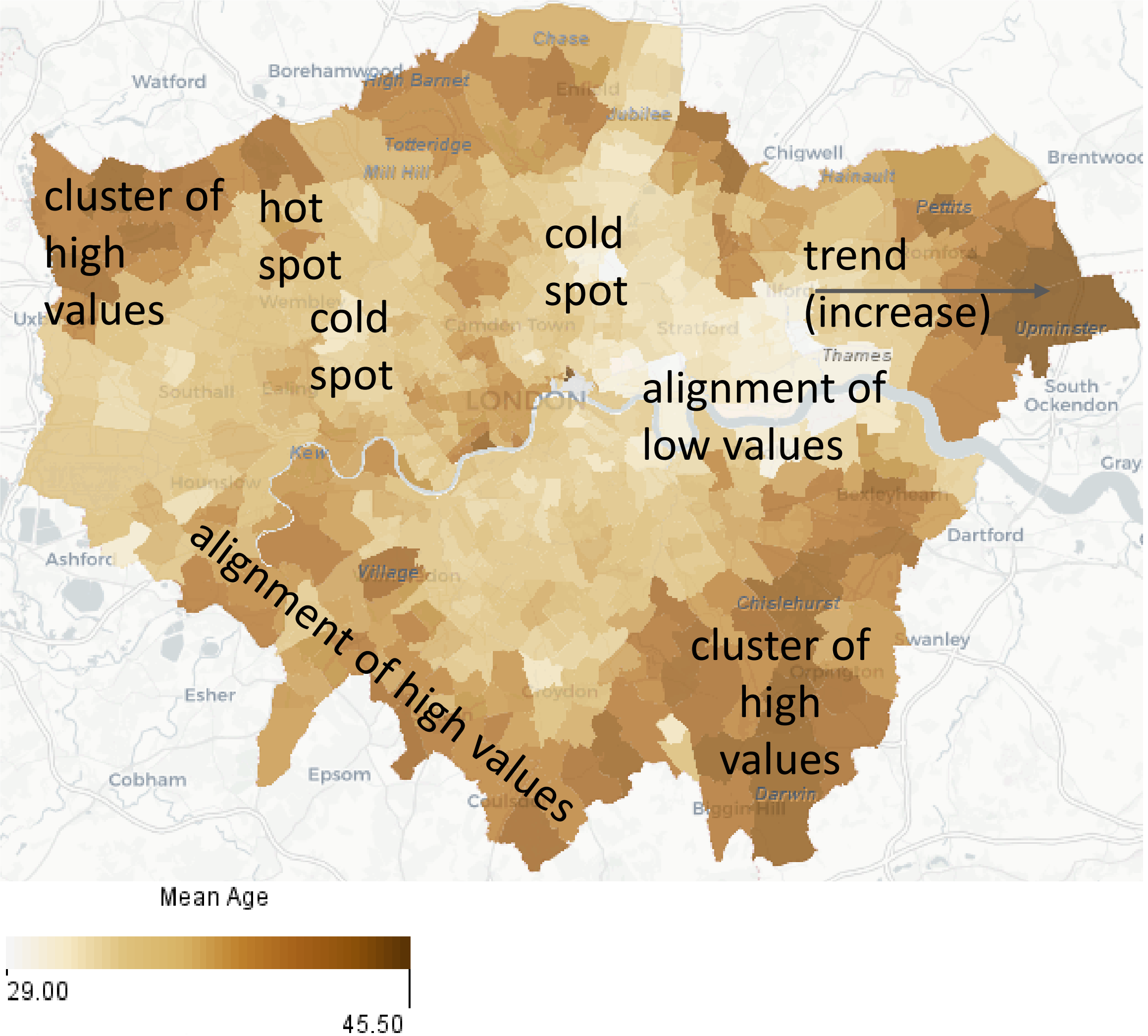


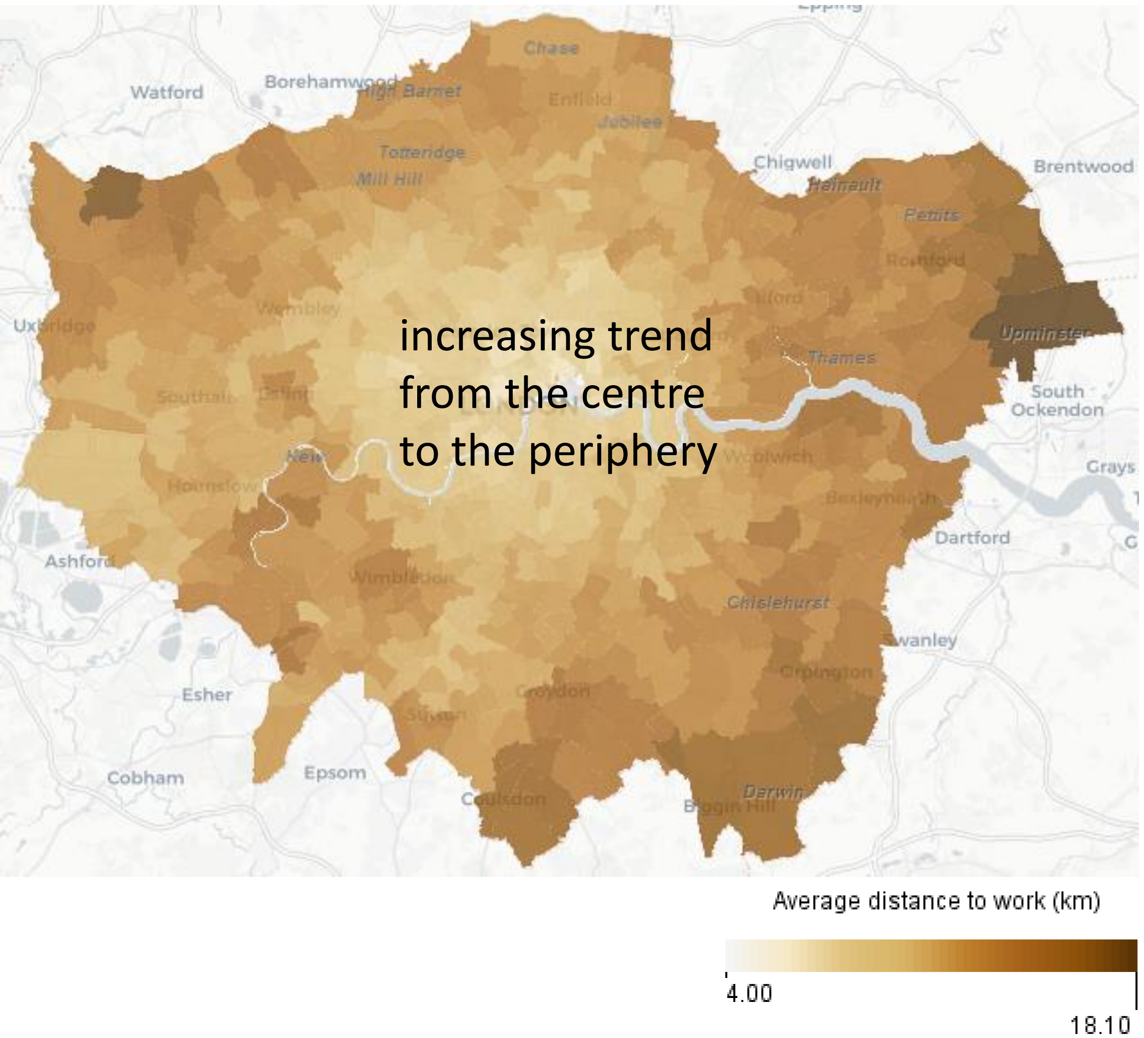
Pattern examples: spatial arrangements

Spatial distribution of tweets related to epidemic outbreak

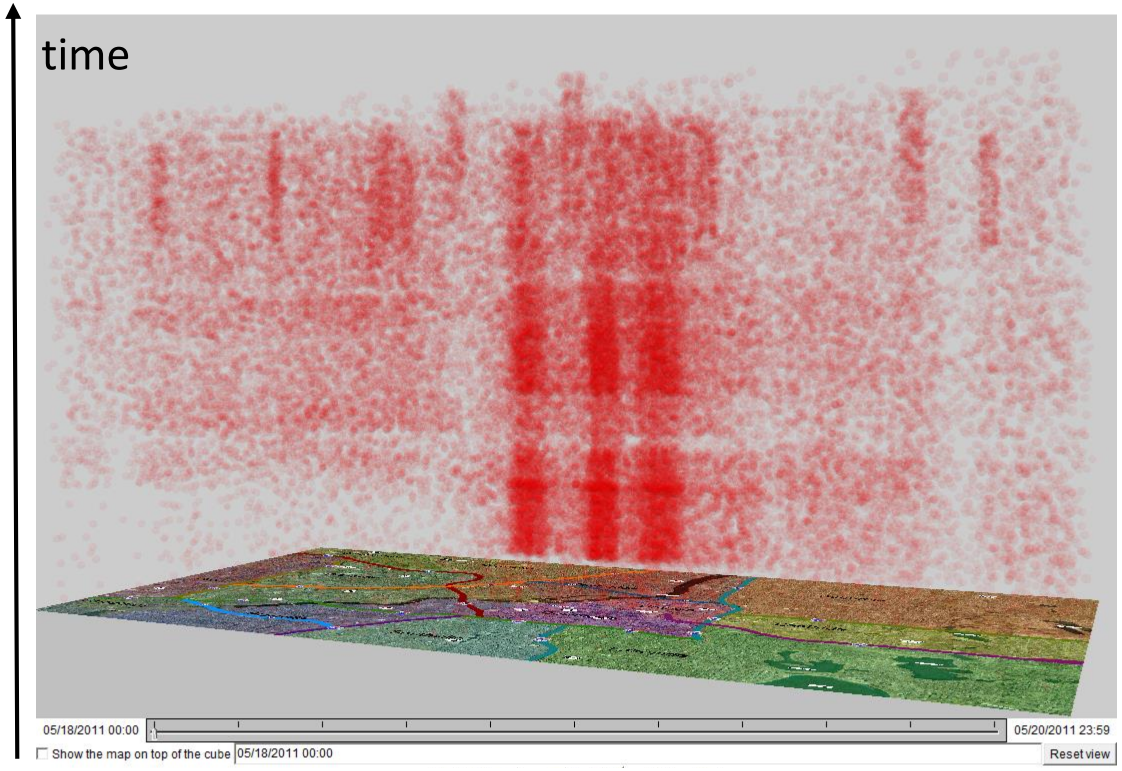


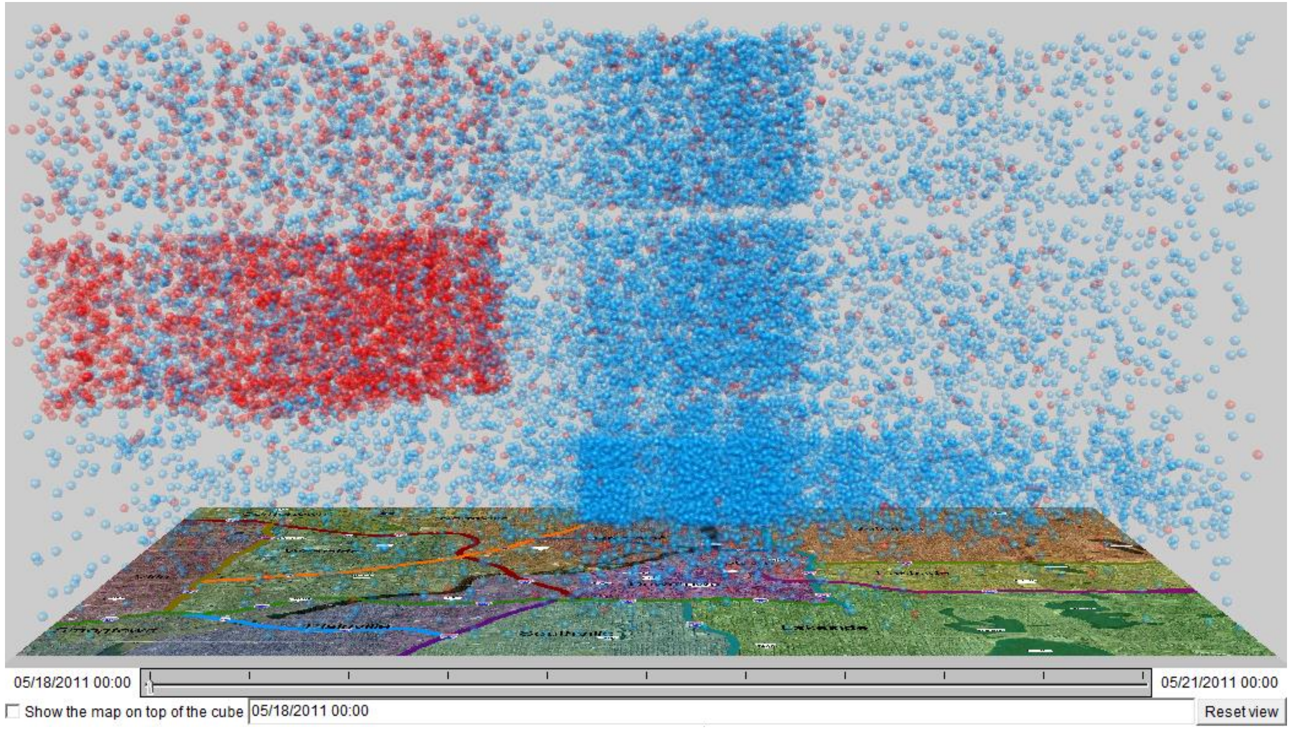
Pattern examples: spatial variation





Pattern examples: spatio-temporal arrangements and variation





Complex base of the distribution: combination (Cartesian product) of space and time. Each element of the base is a combination of a spatial location and a time unit.

Patterns in a spatio-temporal distribution: periods of existence of spatial patterns, trends in changes of spatial patterns over time, repetition of similar spatial patterns, …

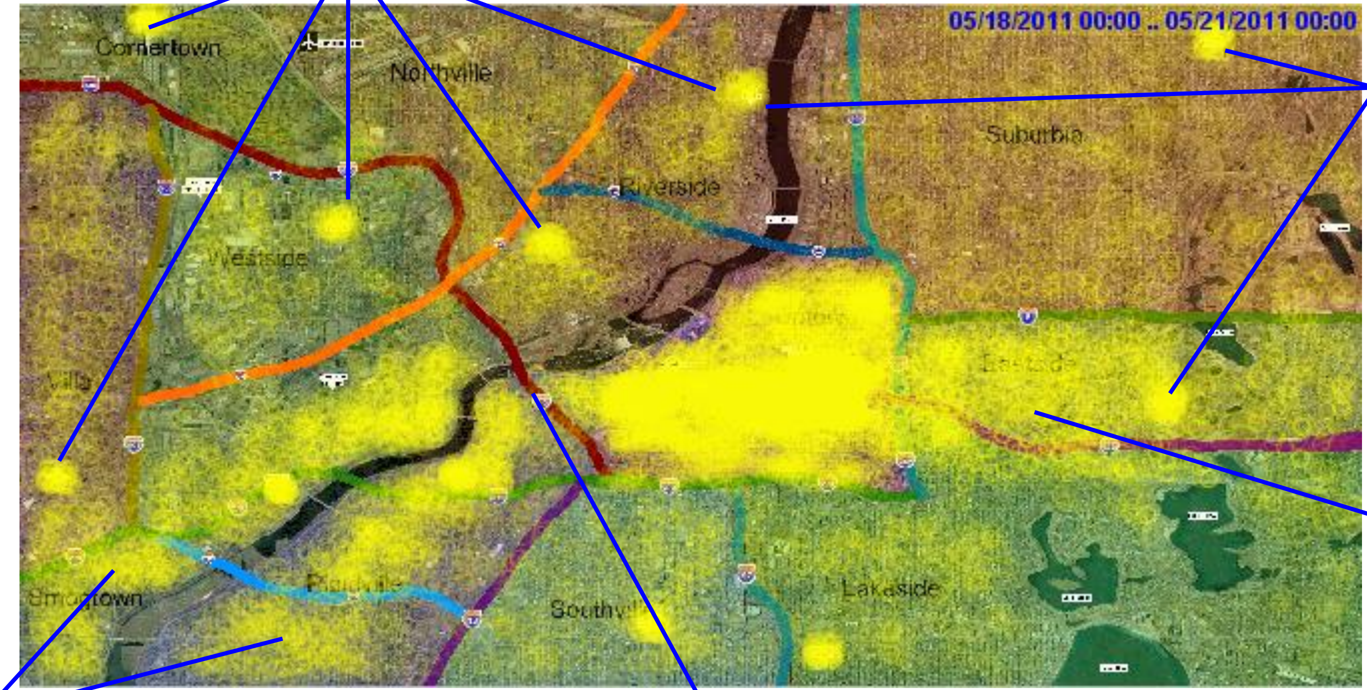
Relationships between patterns

- **Arrangement** relationships between bases of patterns in the base of the overall data distribution
  - e.g., adjacency in space, sequence in time
- Containment: An objective pattern X **contains** (or **includes**) another objective pattern Y when the base of X includes the base of Y.
  - X: super-pattern; Y: sub-pattern
- Similarity: Two or more objective patterns are **similar** if they can be represented by the same abstracted pattern.
- **Repetition**: a super-pattern contains two or more *similar* sub-patterns with non-overlapping bases.
- **Cross-overlay relationships** between patterns existing in distributions of distinct overlays over a common base consist of relationships between the bases of the patterns.

N. Andrienko, G. Andrienko, S. Miksch, H. Schumann, S. Wrobel (2020)
A theoretical model for pattern discovery in visual analytics.
Visual Informatics. https://doi.org/10.1016/j.visinf.2020.12.002

Examples of relationships between patterns

Repetition of multiple **similar** clusters (all are small, roundish, and very dense).



Cross-overlay relationship: the small roundish clusters are located at hospitals.

A spatial cluster **contains** a sub-cluster of somewhat lower density extended in the eastern direction.

Two spatial clusters stretch *in parallel*.

Three spatial clusters *nearly touch* each other at one spot.

Spatial **arrangement relationships** between pattern bases

Analytical operations on patterns

Operations on internal contents:

- Characterise: derive synoptic features of patterns (e.g., statistical summaries, spatial outlines)
- Aggregate: represent as a single element of data
- Refine: divide into parts to represent more precisely (i.e., decrease the abstraction level)

Comparing contents of two or more patterns:

- Compare relationships: determine commonalities and differences of compositions, arrangements, variations
- Group by similarity of contents
- Unify: represent similar patterns by a common abstracted pattern (i.e., increase the abstraction level)

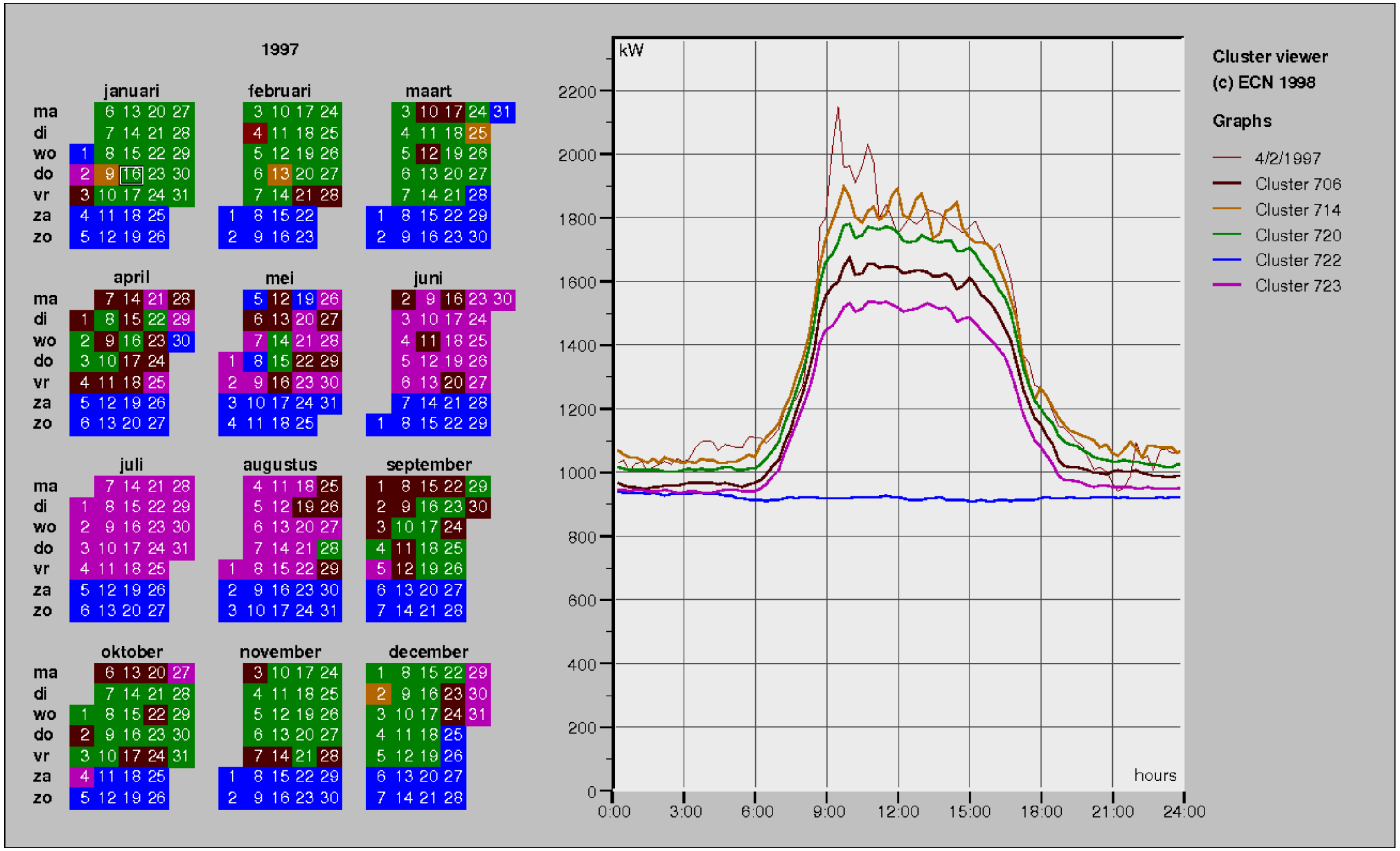
Operations involving relationships to external elements within the same distribution:

- Relate to context: determine relationships of the pattern base and overlay to the remainder of the distribution
- Relate patterns: determine relationships between patterns within the same distribution
- Unite patterns: create super-patterns from several patterns and relationships between them

Operations involving connections of pattern base or overlay to elements of other data components:

- Extract connected elements for elements of pattern base or overlay
- Characterise using connections: derive synoptic features of patterns by summarising connected elements
- Identify cross-overlay relationships: determine relationships between patterns in different distributions with a common base

Examples of pattern operations



van Wijk, J.J., van Selow, E.R. (1999)
Cluster and calendar based visualization of time series data.
In: Proc. IEEE Symposium on Information Visualization (InfoVis), pp. 4–9.

Compare variation patterns if different days

Group patterns: create clusters of similar patterns of daily variation

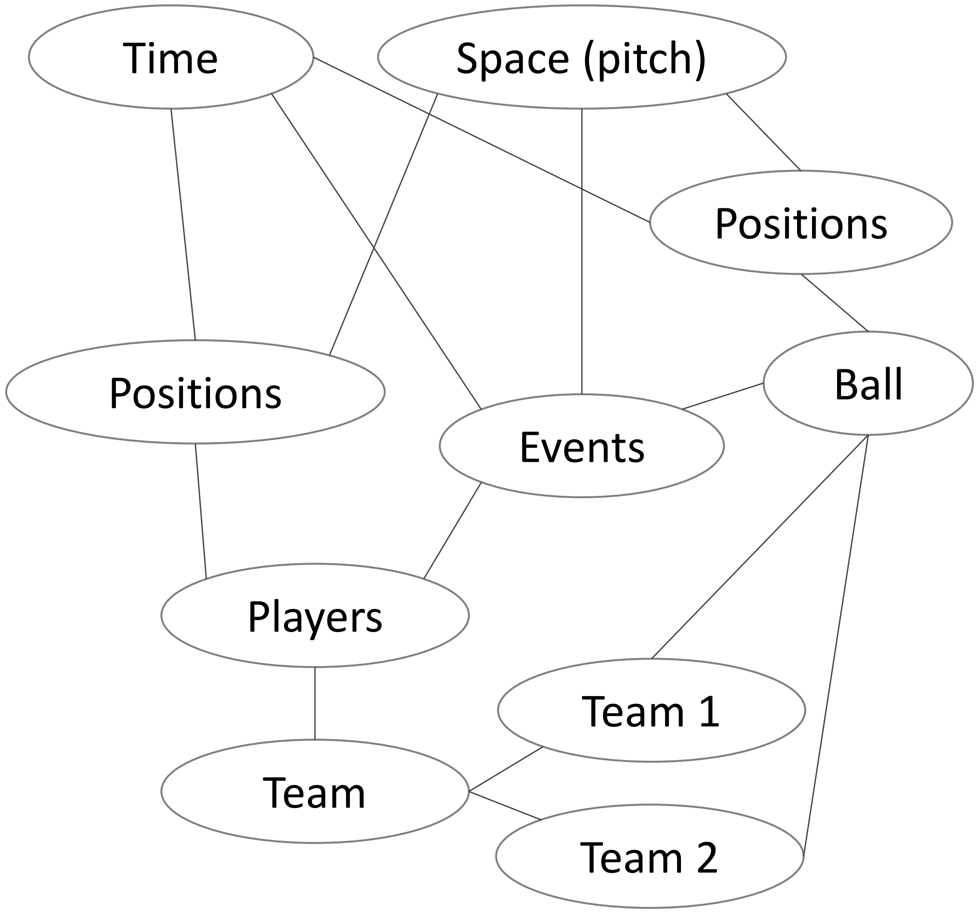
Unify patterns: represent clusters of similar daily patterns by common abstracted patterns

Aggregate patterns: treat unified patterns as data elements and study their distribution over the year

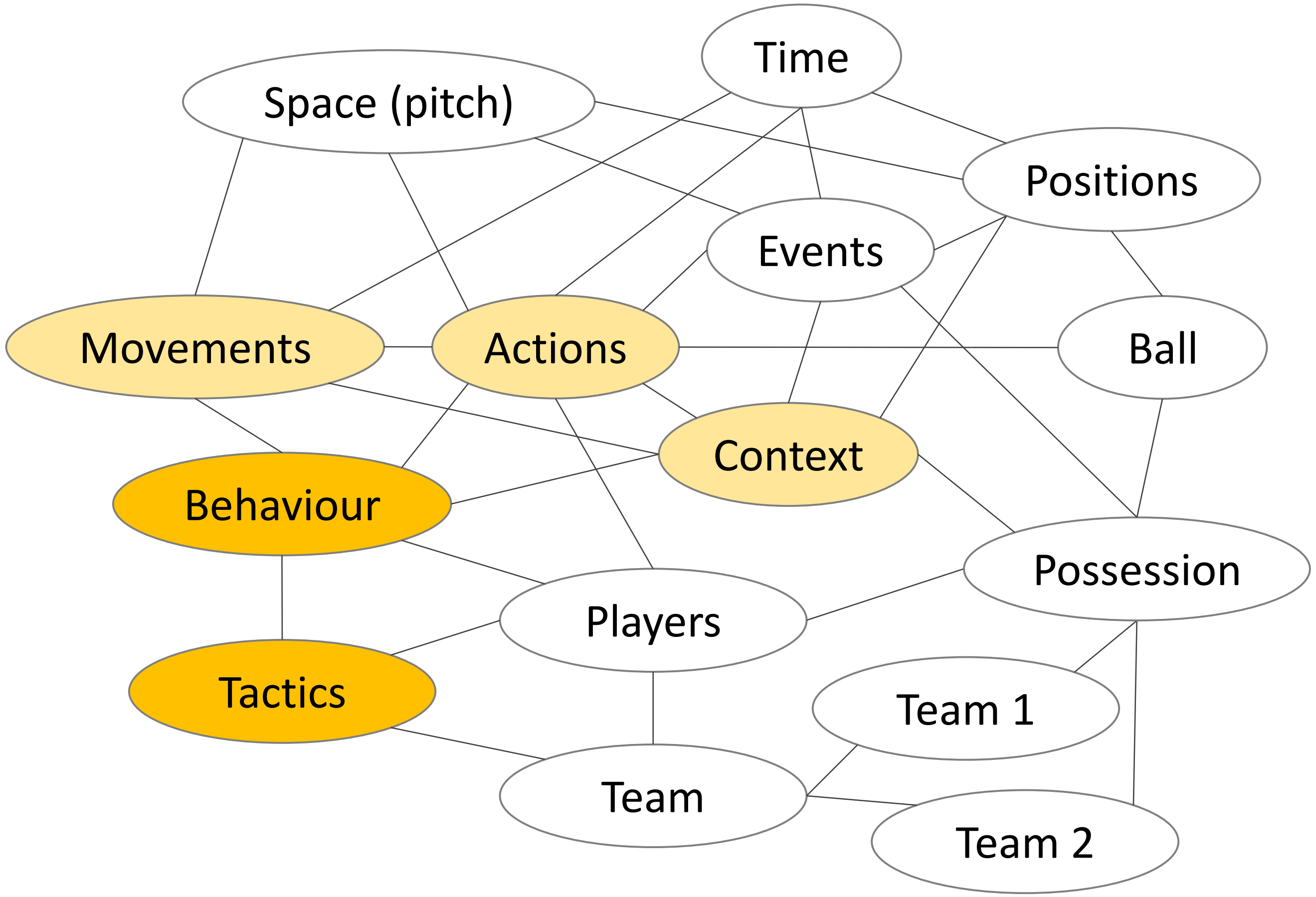
Requirements to visualisation and analysis techniques

- Since patterns are formed by relationships between data elements, the relationships must be properly taken into account in the analysis.
  - Particularly, visualisation must fulfil the principle of correspondence:
    - Visualisation must faithfully show existing relationships.
    - Visualisation must not provoke seeing non-existent relationships, to preclude generation of false abstracted patterns.
- Since patterns need to be considered and represented holistically, analytical techniques need to support unification and abstraction.
  - Particularly, visualisation needs to facilitate perceptual unification of multiple elements (principle of unification).
- Tools for data analysis need to enable pattern operations.

What about football?

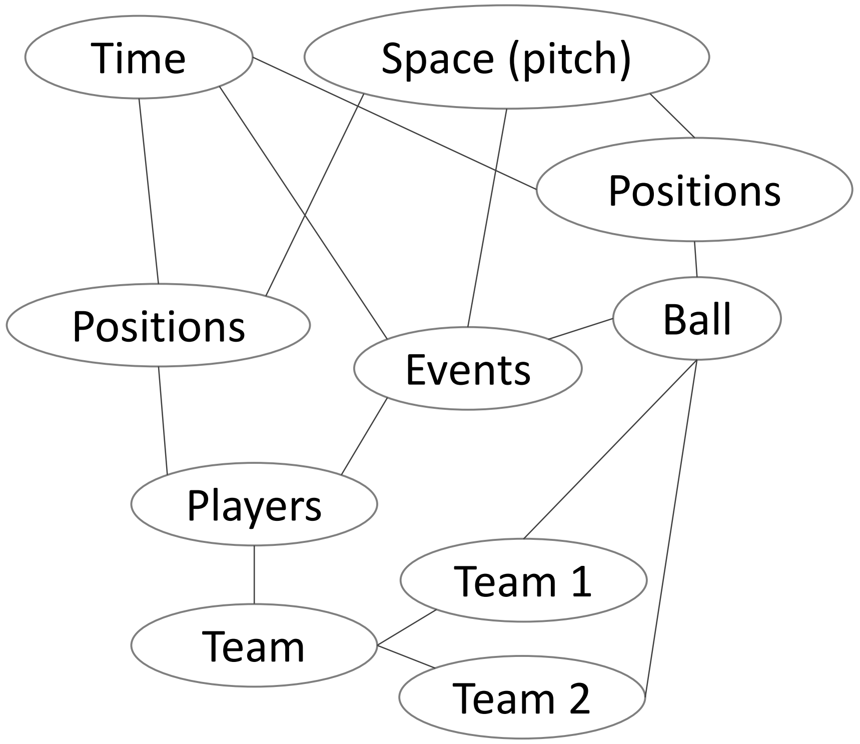


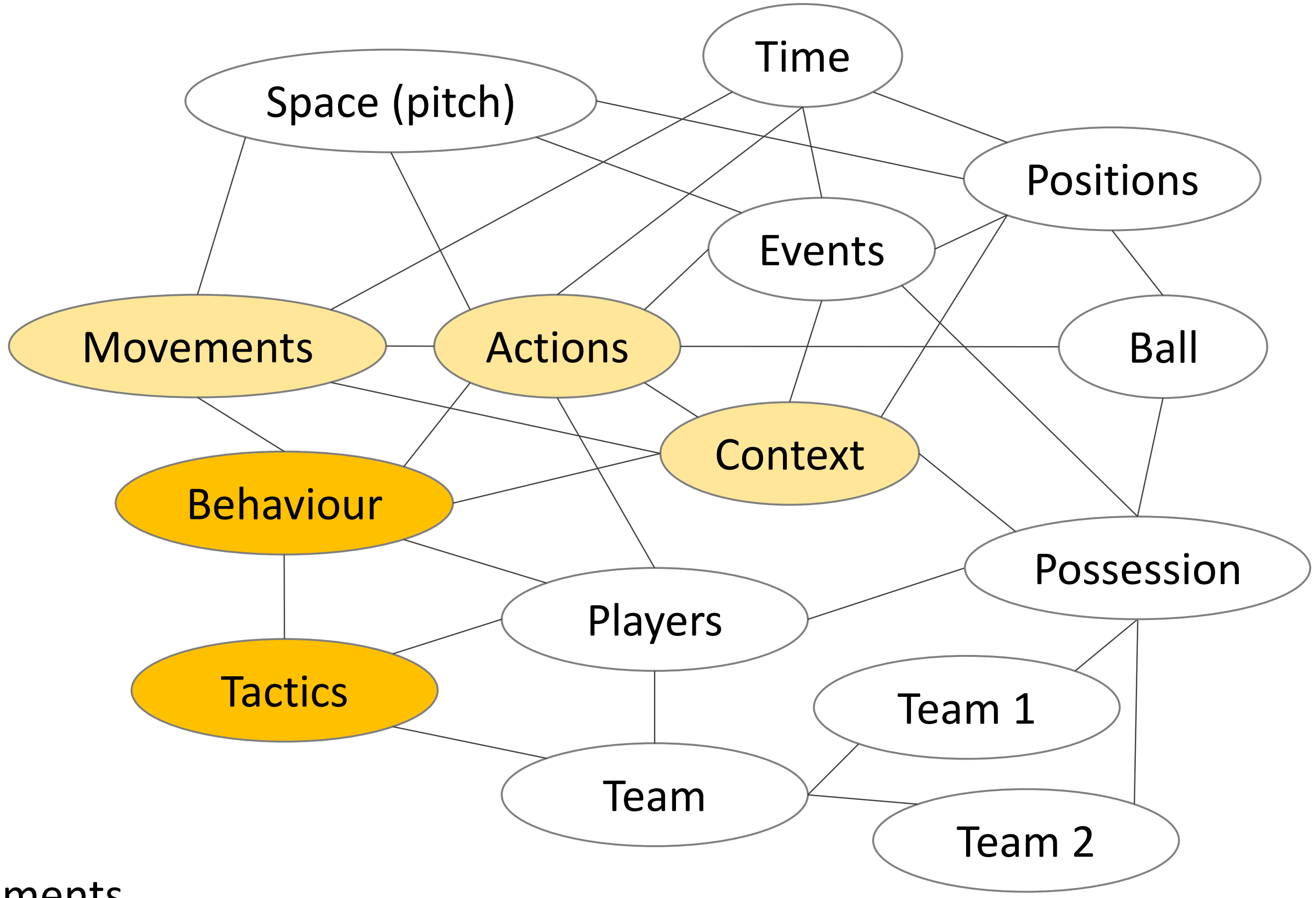
Components of data



What we need to reconstruct and model*

* represent in a generalised manner

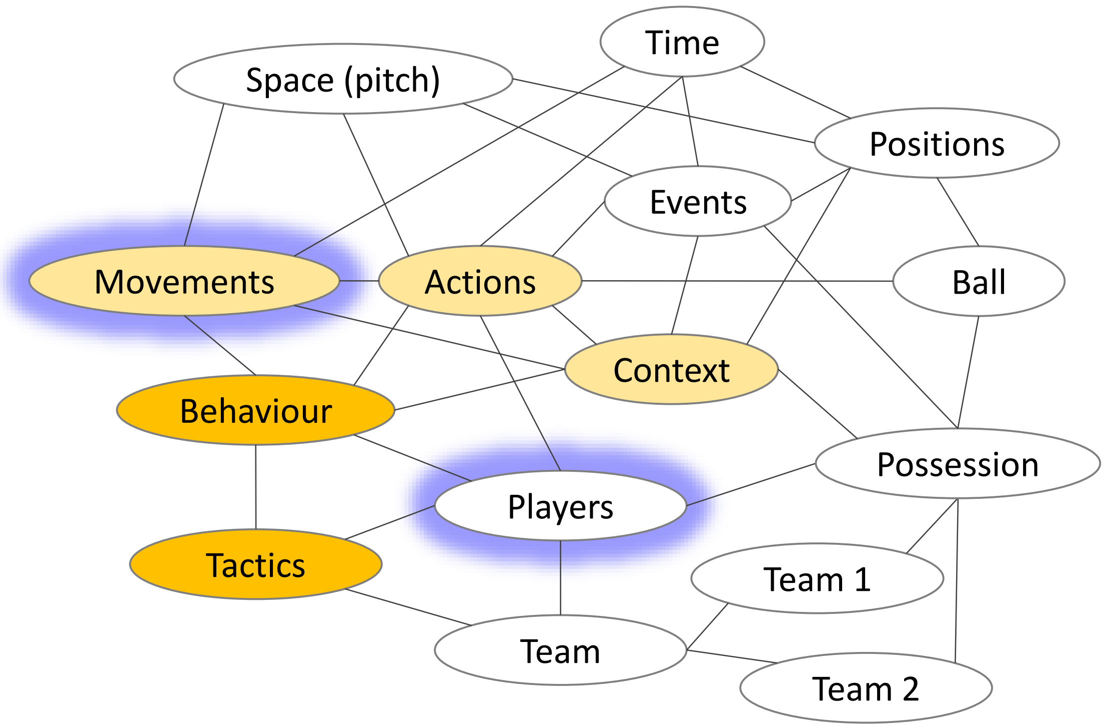




Players' positions (time, space) → players' movements
Events (time, space) → players' actions (time, space); e.g., passes, shots
Players' positions (time, space) + ball possession (team) + ball positions (time, space) + opponents' positions (time, space) → players' actions (time, space); e.g., pressure on opponents, goal coverage
Players' movements (*context*) + players' actions (*context*) → players' behaviours
Players' behaviours (team) → team tactics

Modelling players' movements

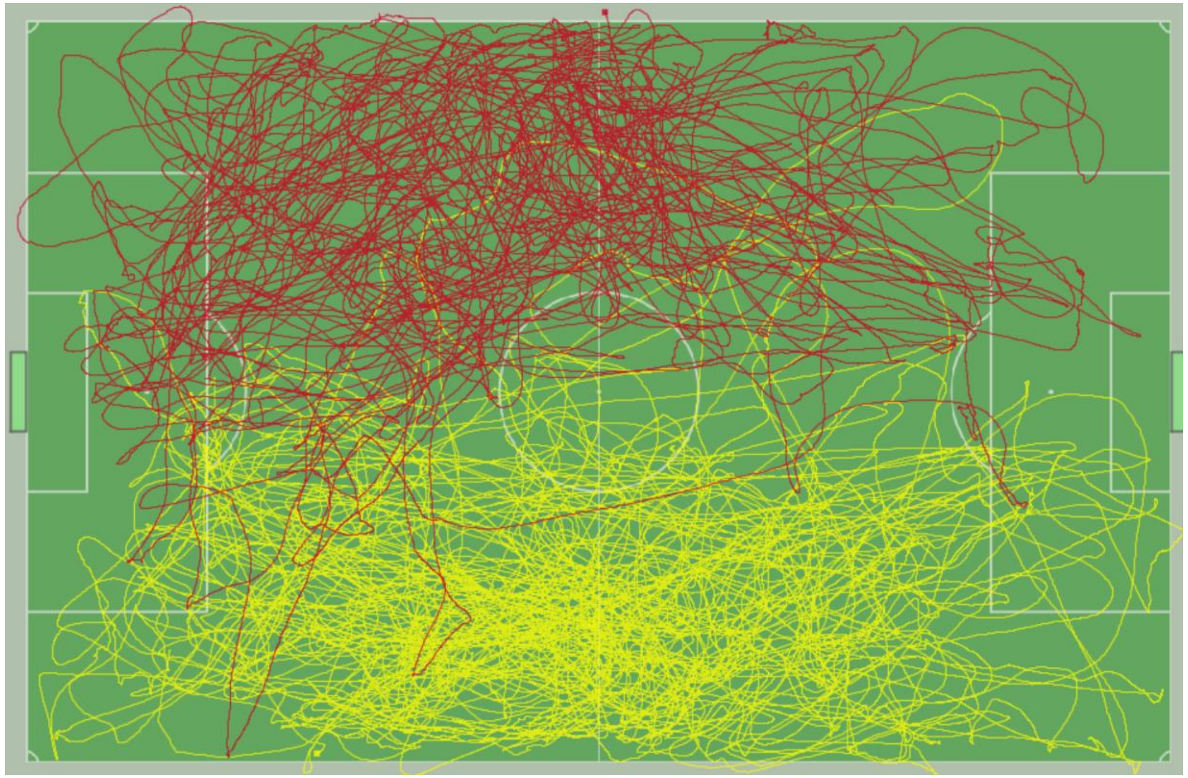
From time-stamped position records to movement patterns



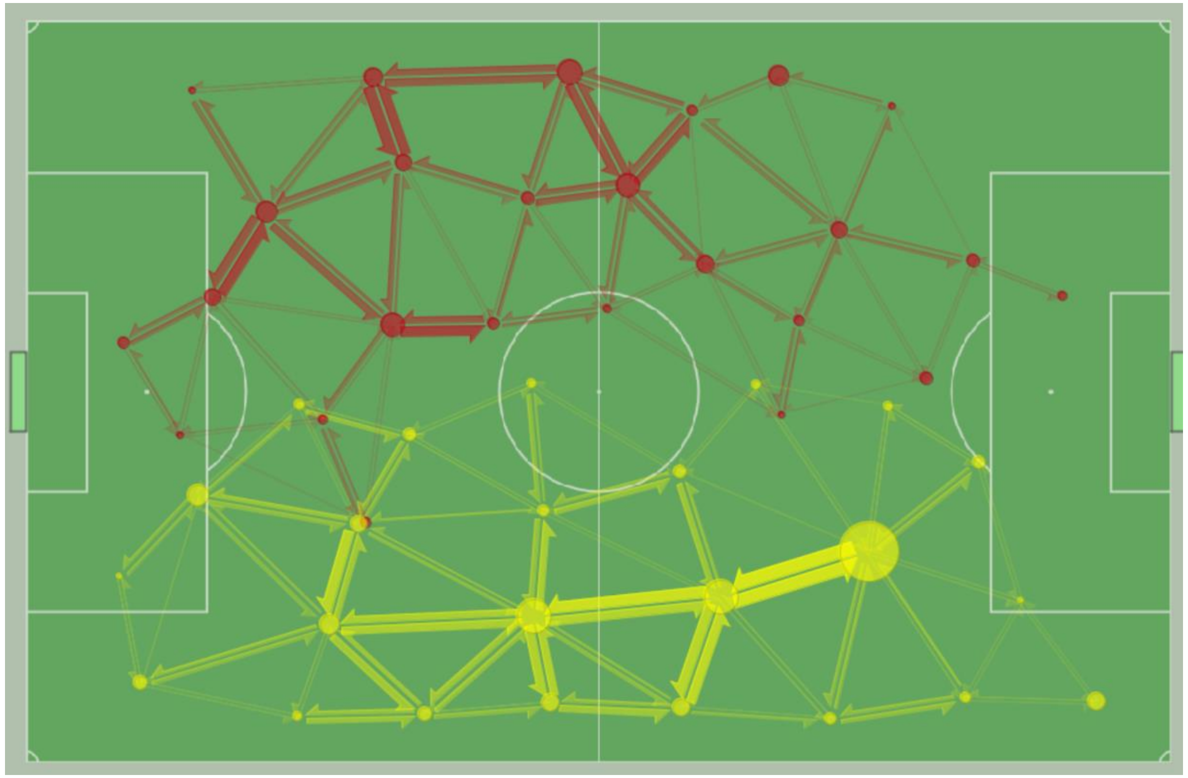
Unification: players' positions (time, space) $\rightarrow$ players' movements (trajectories).
The level of abstraction is too low, so that the patterns are too complex and hard to understand.

Approach to increase the abstraction level:

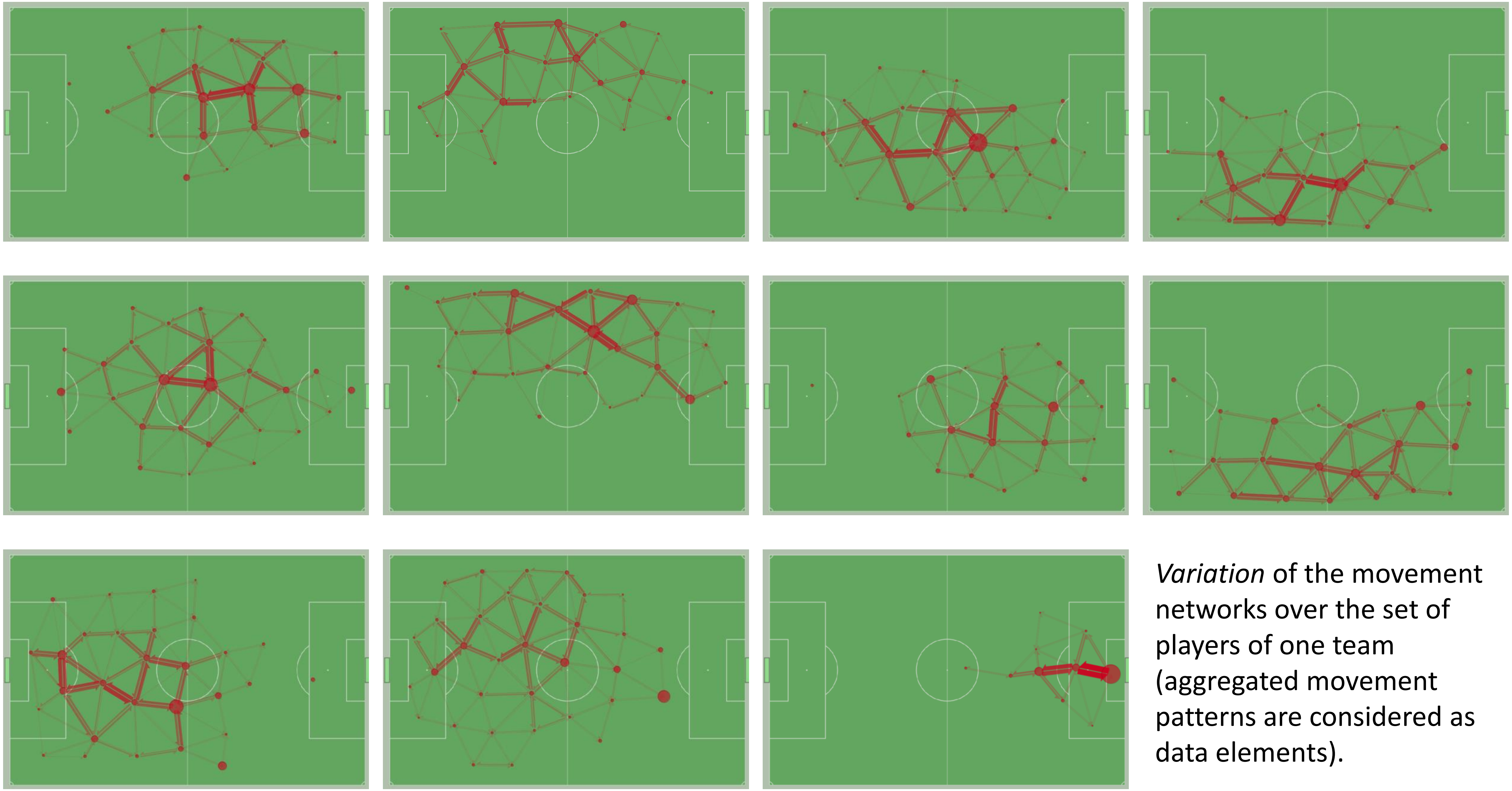
- group points into areas *(exploit spatial distance relationships)*;
- replace trajectory segments by transitions between areas *(exploit temporal ordering relationships)*;
- aggregate transitions *(exploit commonality of origins and destinations)*



Example: trajectories of two selected players from different teams.

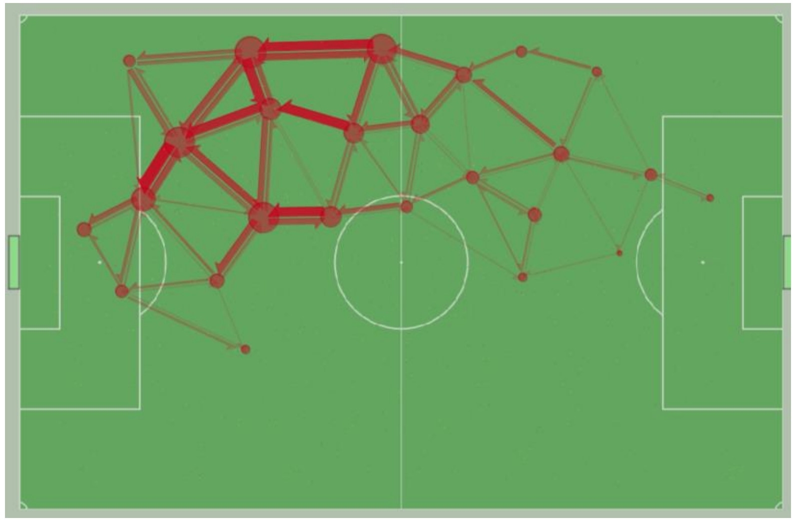


Operation: aggregate a pattern. Result: a spatial network.

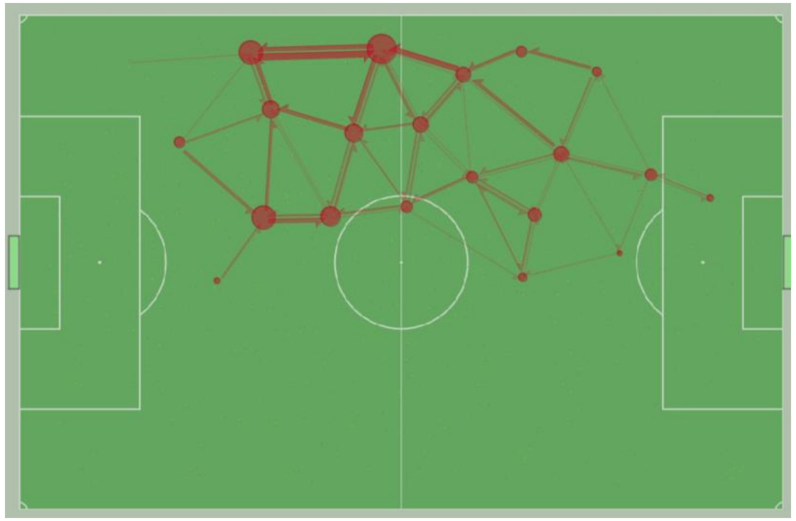


Variation of the movement networks over the set of players of one team (aggregated movement patterns are considered as data elements).

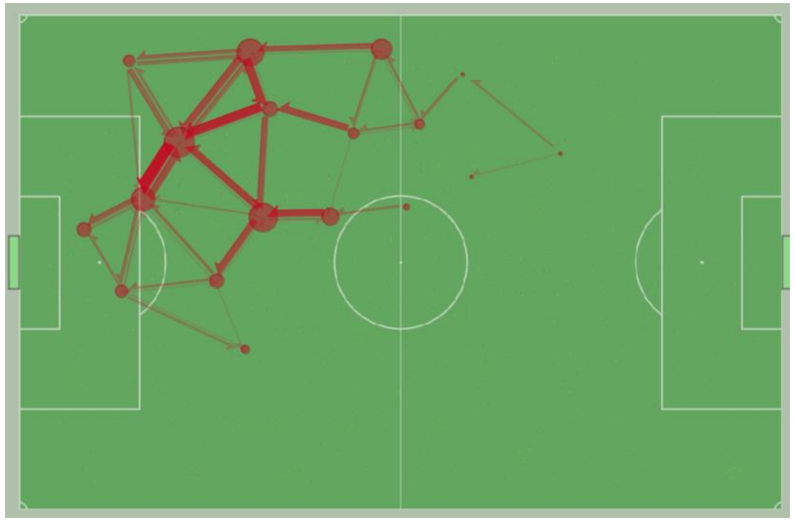
Player’s movements in different *contexts*



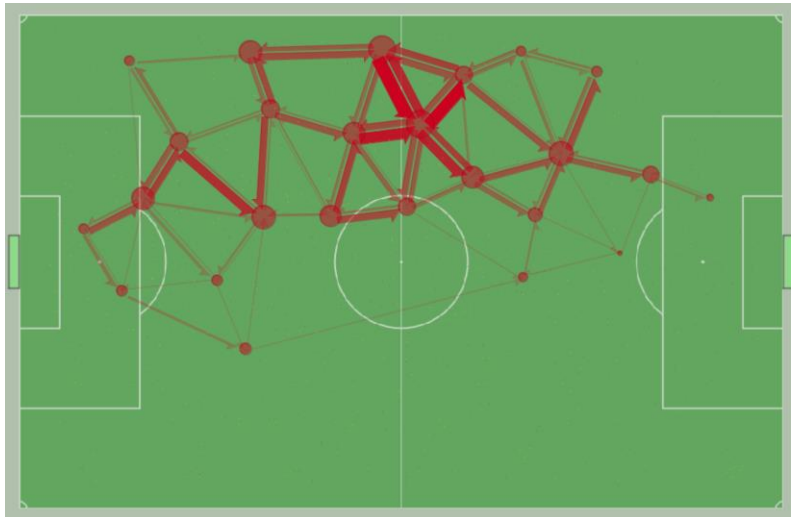
When his team possesses the ball



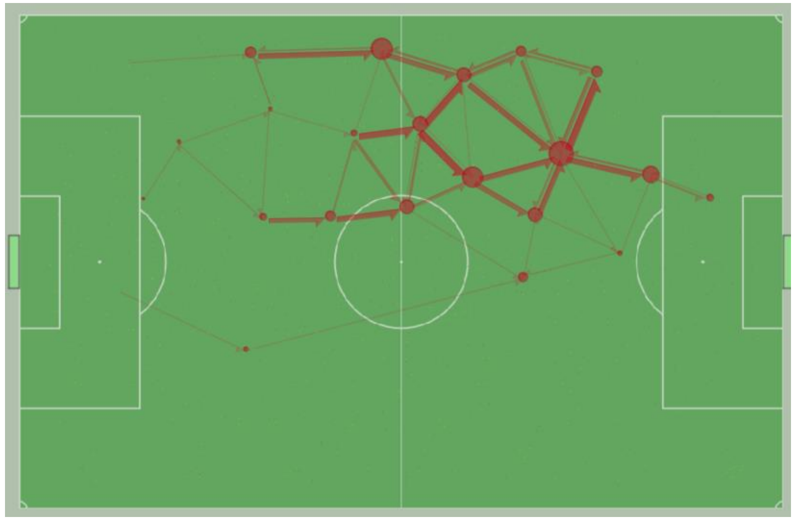
When the ball is on the own team’s side



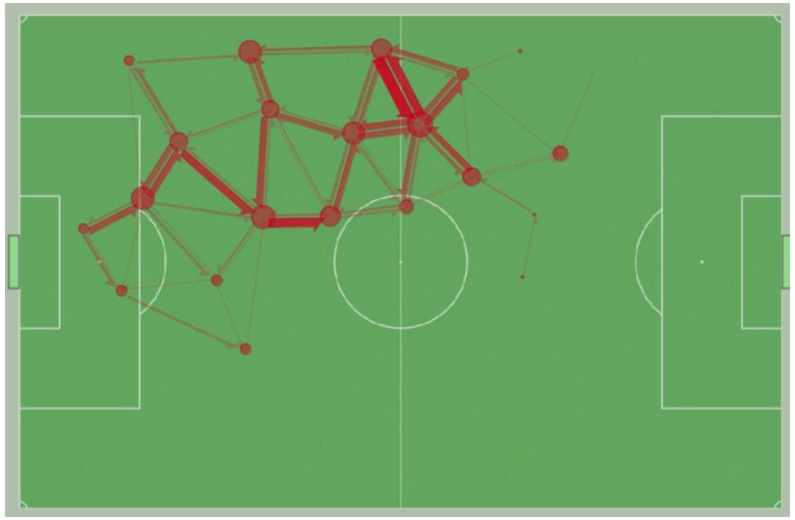
When the ball is on the opponents’ side



When the opponents possess the ball



When the ball is on the own team’s side

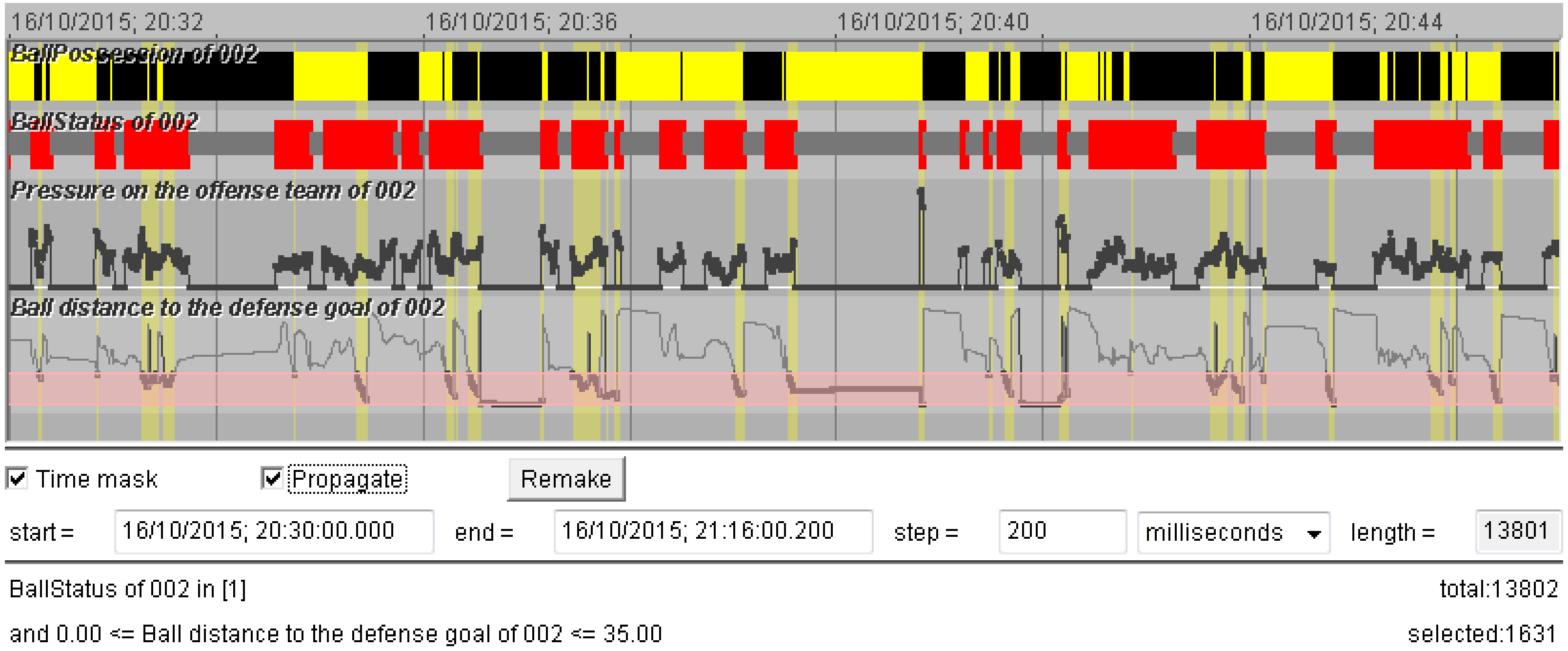


When the ball is on the opponents’ side

Spatio-temporal *context*

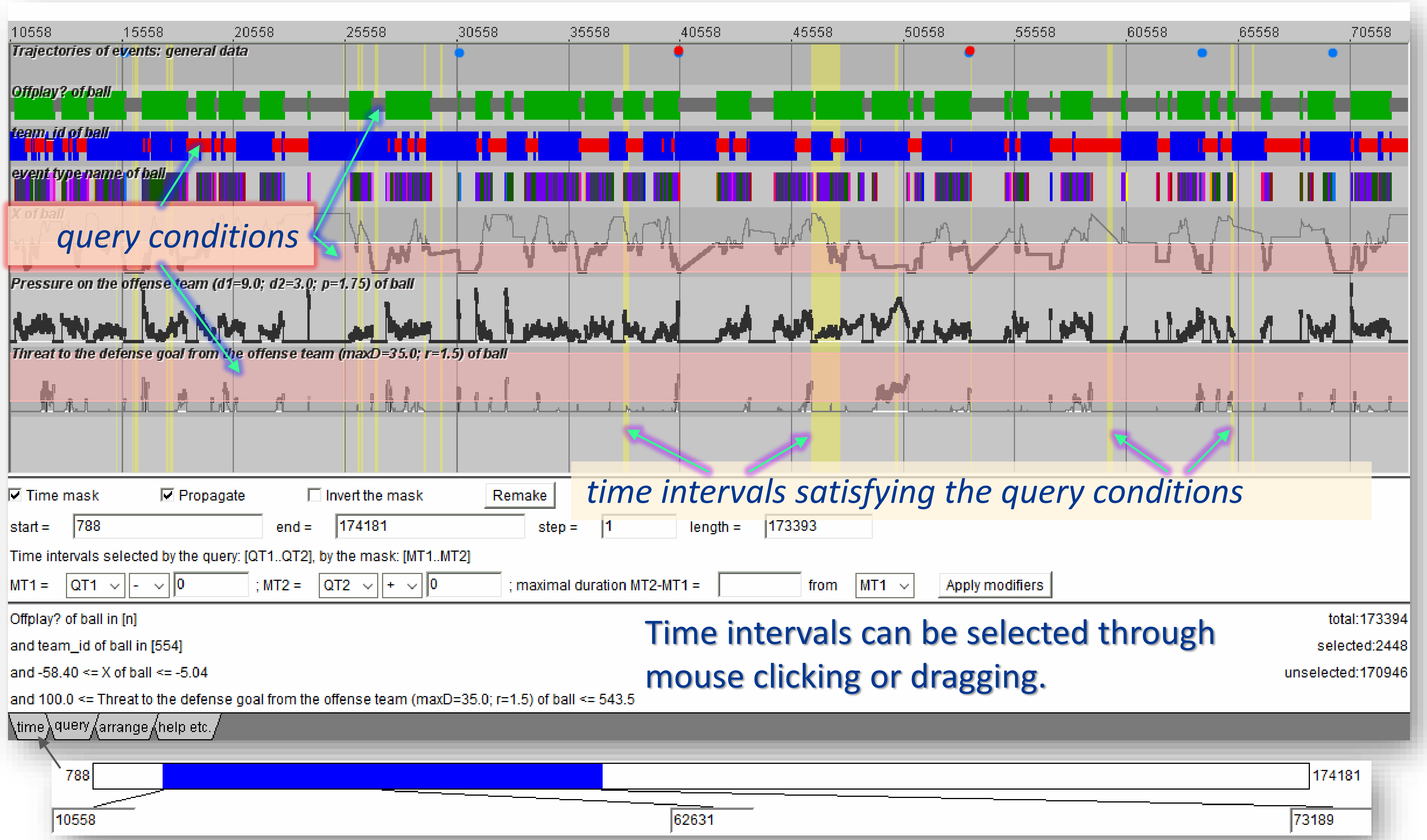
- The context of players’ movements and actions includes various kinds of circumstances existing at the current time moment.
  - Ball possession, ball position, positions and actions of other players, current score, remaining time to the break or to the game end, time since last ball possession change, …
- The context involves multiple data components *distributed over a common base*: time.
- Time is also the base for players’ movements.
- To understand players’ behaviours (= ways of acting depending on circumstances), we need to analyse *cross-overlay relationships* between the temporal distributions of different data components.

A tool to explore cross-overlay relationships over time: Time Mask

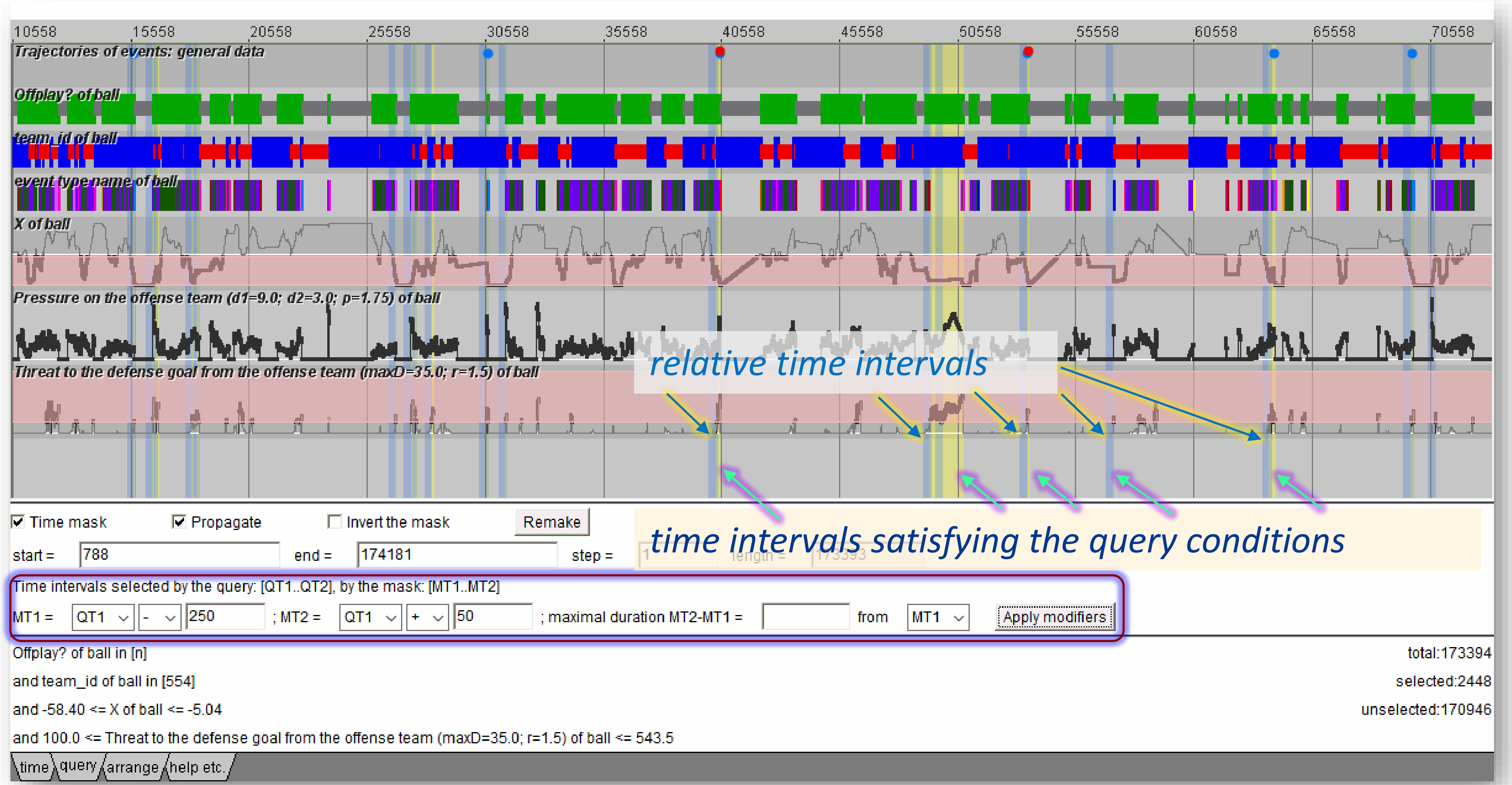


N. Andrienko, G. Andrienko, E. Camossi, C. Claramunt, J.M. Cordero Garcia, G. Fuchs, M. Hadzagic, A.-L. Jousselme, C. Ray, D. Scarlatti, G. Vouros (2017)
Visual exploration of movement and event data with interactive time masks.
Visual Informatics 1(1): 25-39, https://doi.org/10.1016/j.visinf.2017.01.004.

Specifying query conditions to define a context of interest



Selection of relative times w.r.t. the query



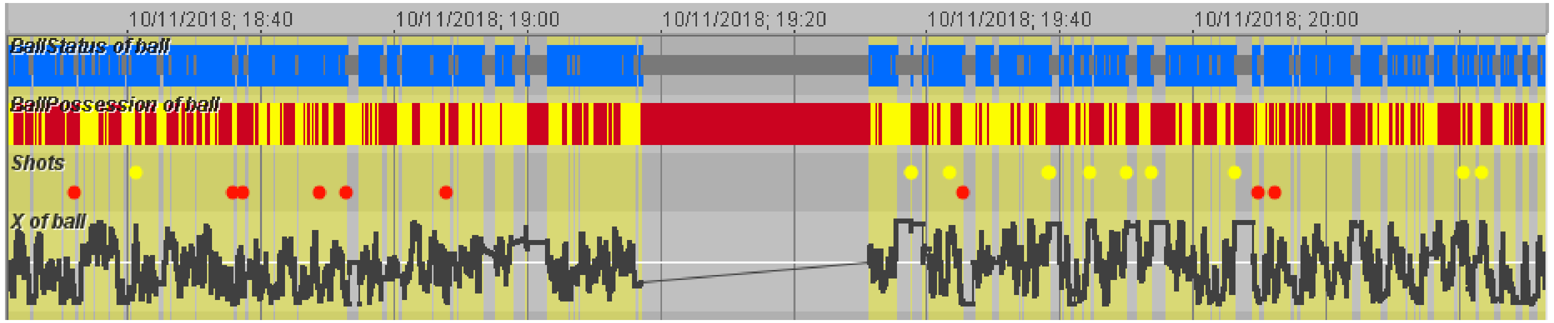
Time Query

- Basic primitive:
  Select time moments by conditions representing game context expressed through attributes of the ball, players, and the teams, and game events
- Unite the selected time moments into time intervals, or episodes
  - Be precise: allow only intervals that satisfy given conditions (e.g. duration)
  - Be tolerant: ignore breaking episodes by a few unselected time moments
- When appropriate, shift or extend / restrict the selected intervals in time

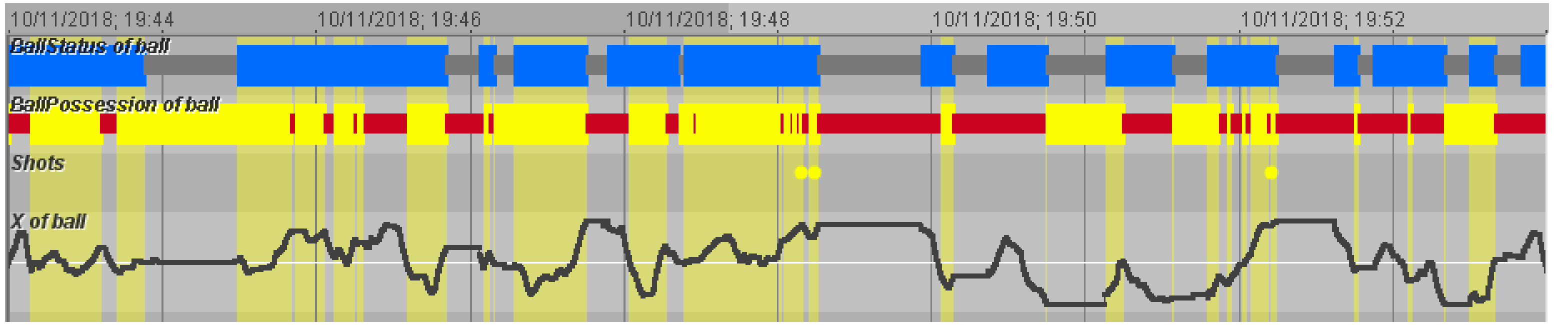
G. Andrienko, N. Andrienko, G. Anzer, P. Bauer, G. Budziak, G. Fuchs, D. Hecker, H. Weber, and S. Wrobel (2019)
Constructing Spaces and Times for Tactical Analysis in Football.
IEEE Transactions on Visualization and Computer Graphics, https://doi.org/10.1109/TVCG.2019.2952129.

Time query: example

- out of play excluded: 102 episodes / 97159 frames

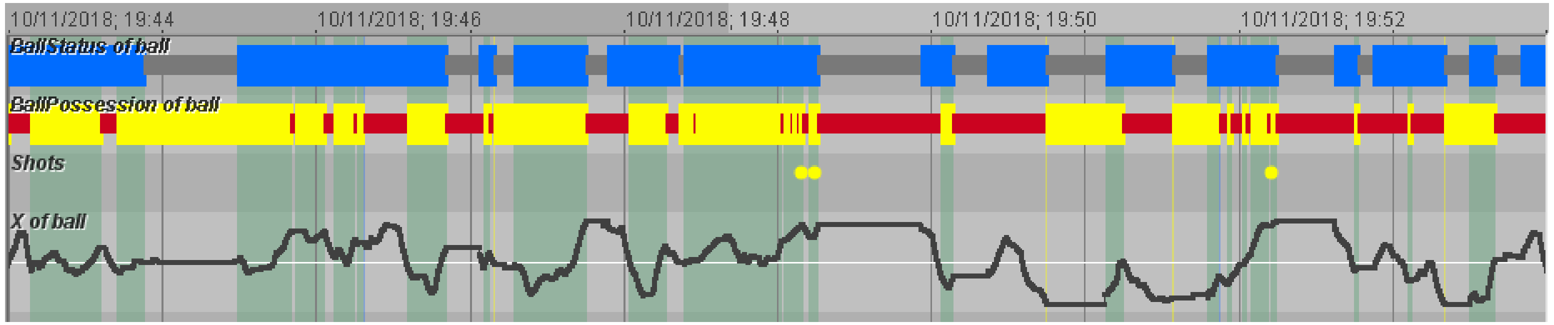


- ball possession by the “yellow” team (time zoom to 15 minutes): 230 / 43607 frames

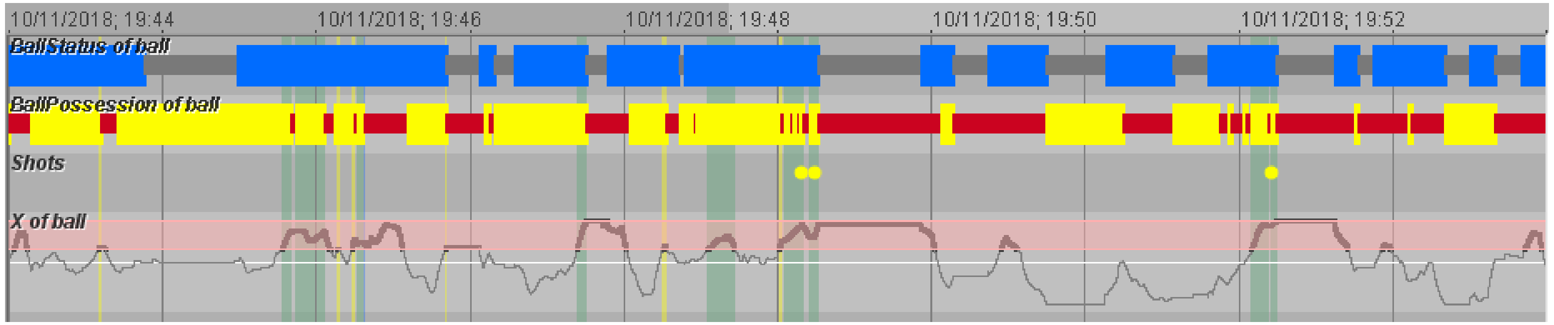


Time query: example (continued)

- episodes shorter than 1 second ignored: 206 episodes / 43291 frames

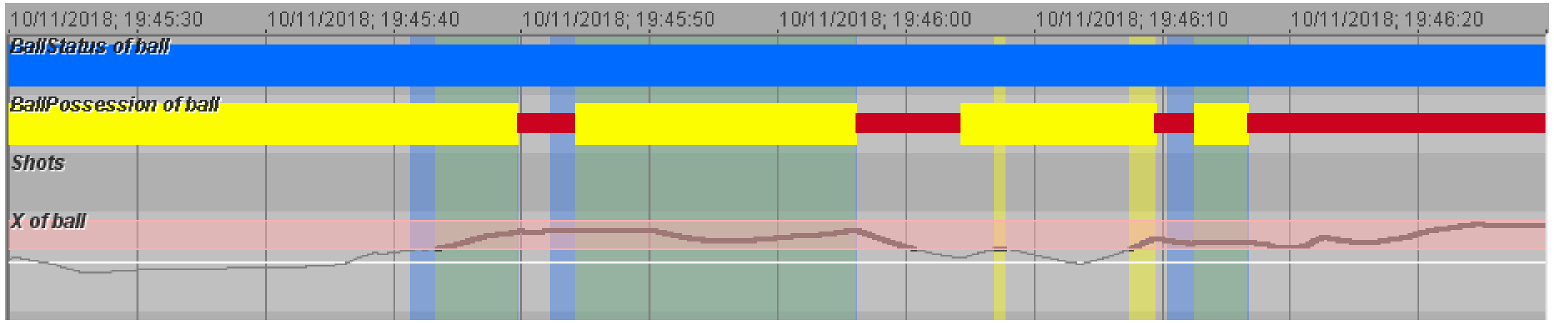


- the ball is in the attacking third of BVB: 60 episodes / 6341 frames

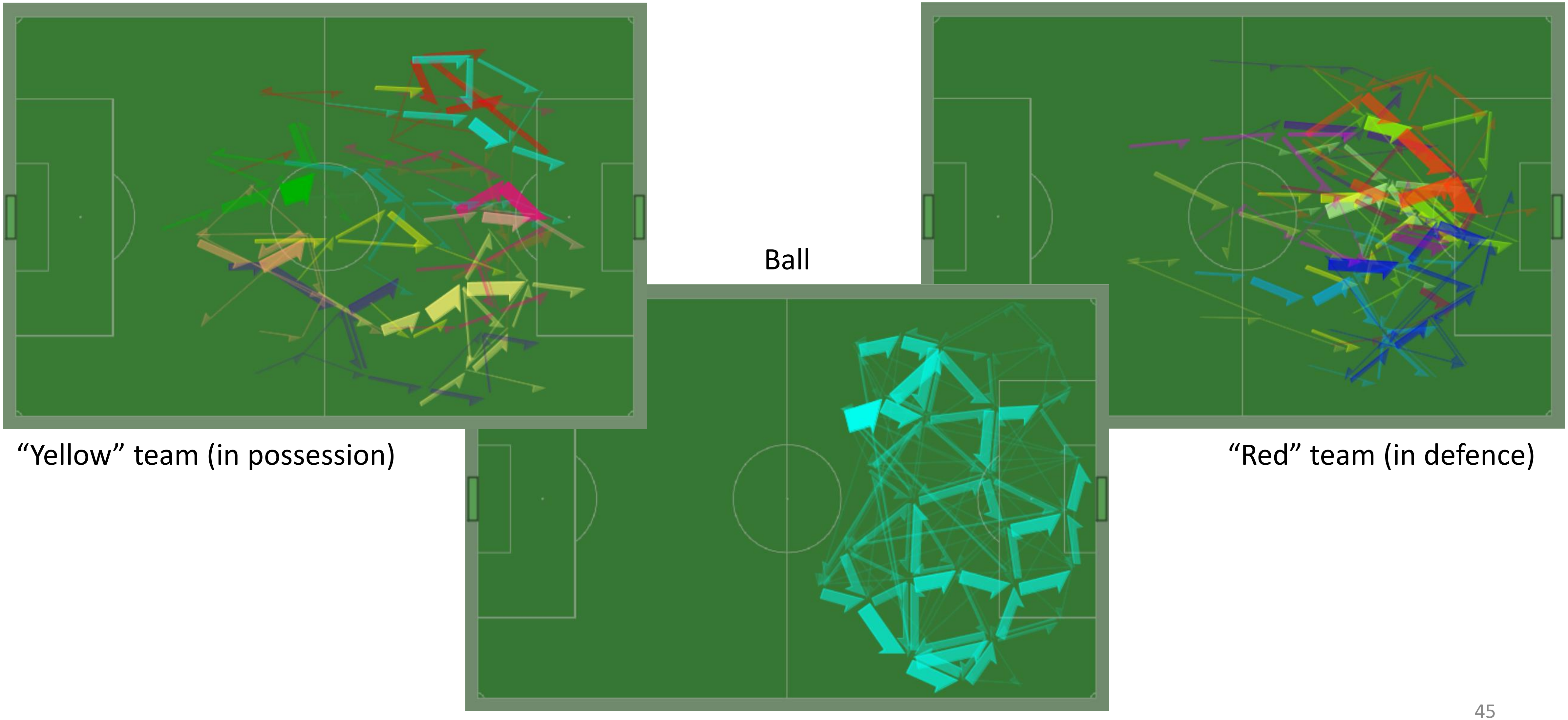


Time query: example (continued)

- Relative intervals: add 1 second before the selected episodes:
  60 episodes / 7841 frames



Abstracted movement patterns in this group of situations



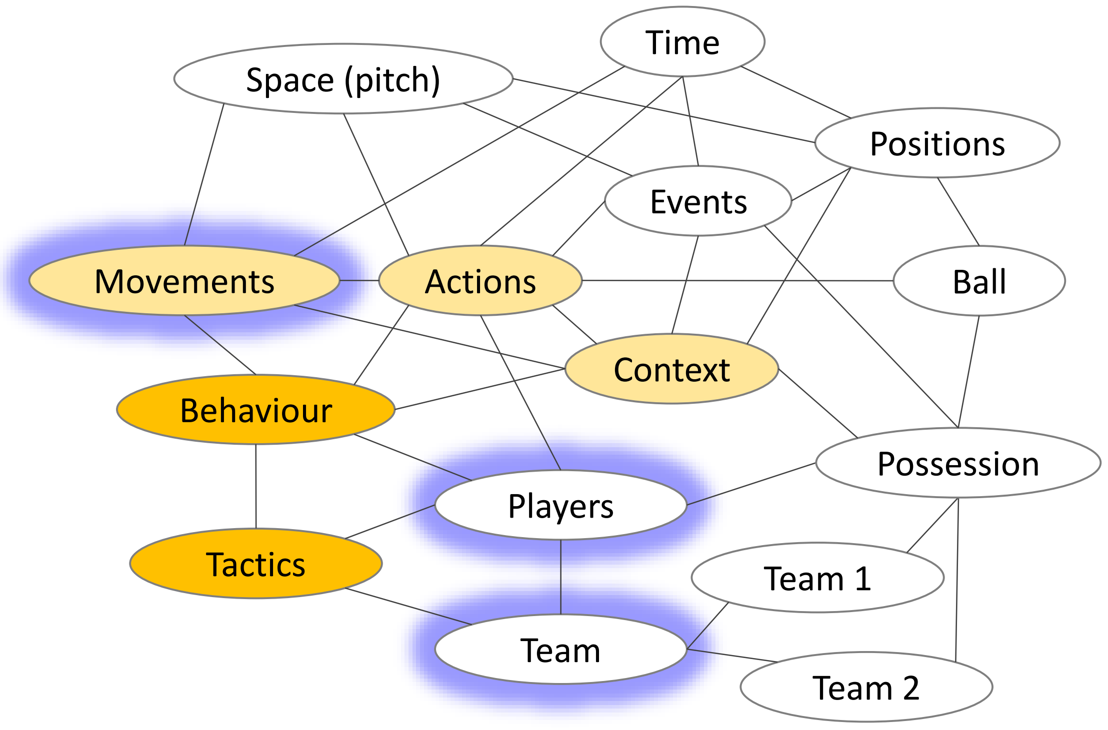
“Yellow” team (in possession)

Ball

“Red” team (in defence)

Players' relative arrangement and movements in teams

Distribution base: team space

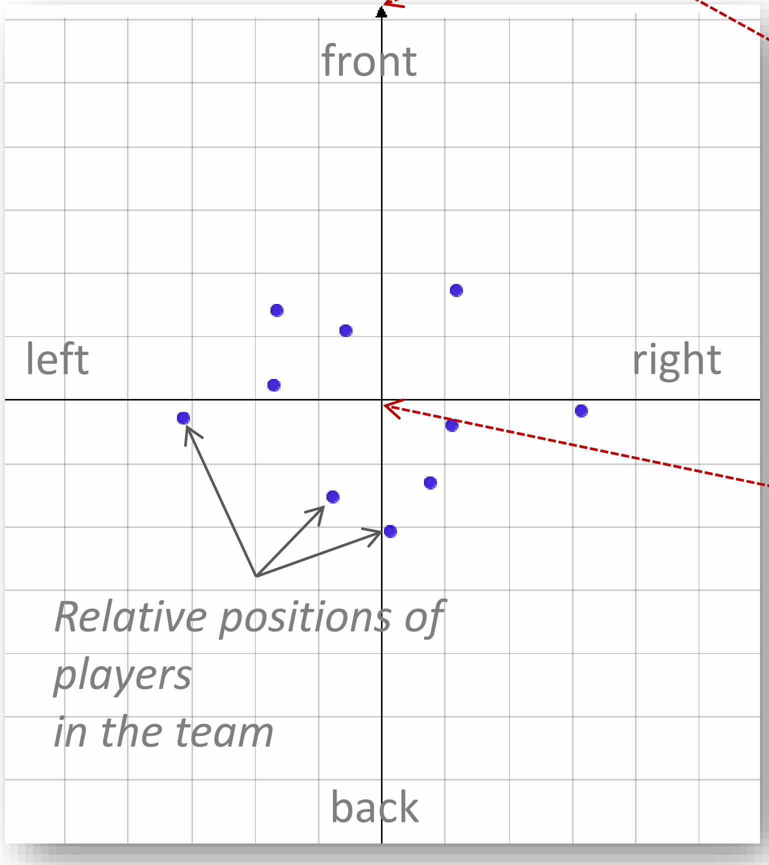


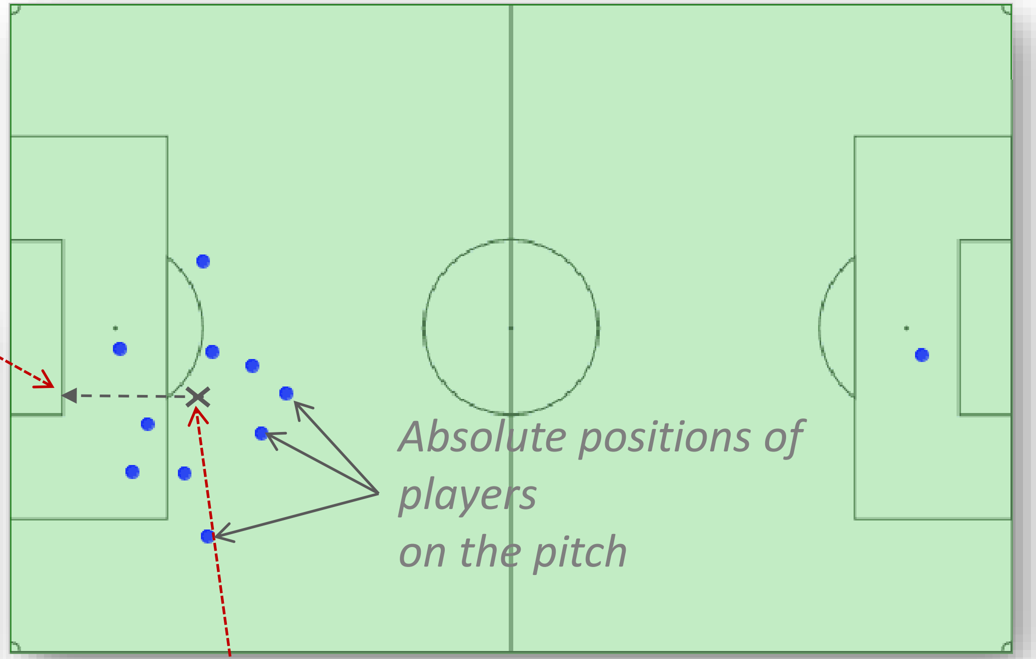
Football teams’ formations

Players generally strive to keep the arrangement while moving but vary it depending on the context.

Team space

Direction towards the opponents' goal

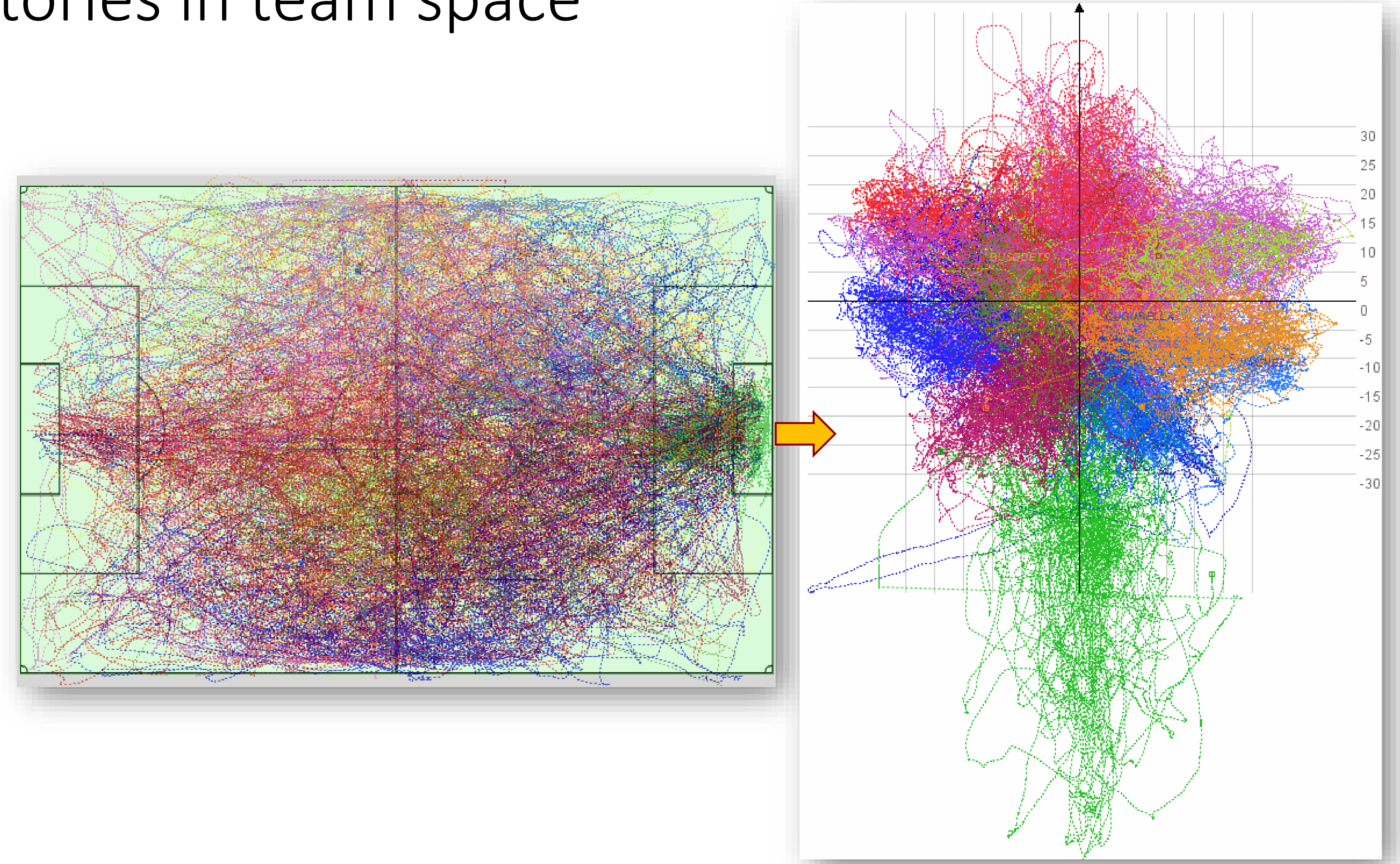




Team's centre
(average position
excluding outliers)

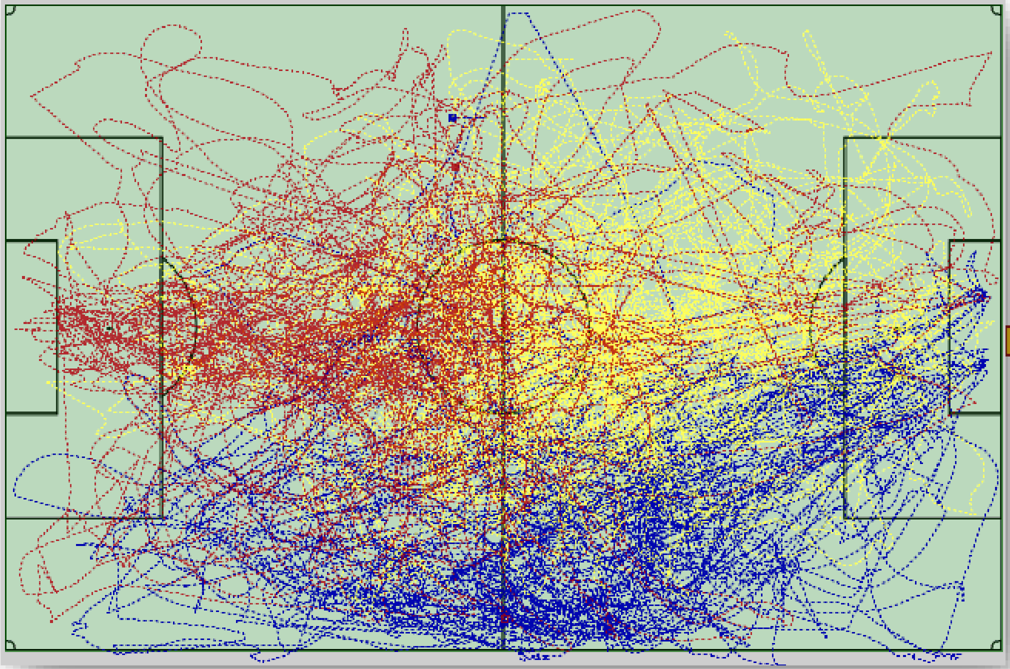
This transformation is done for each time step.
Result: trajectories representing relative movements within the team.

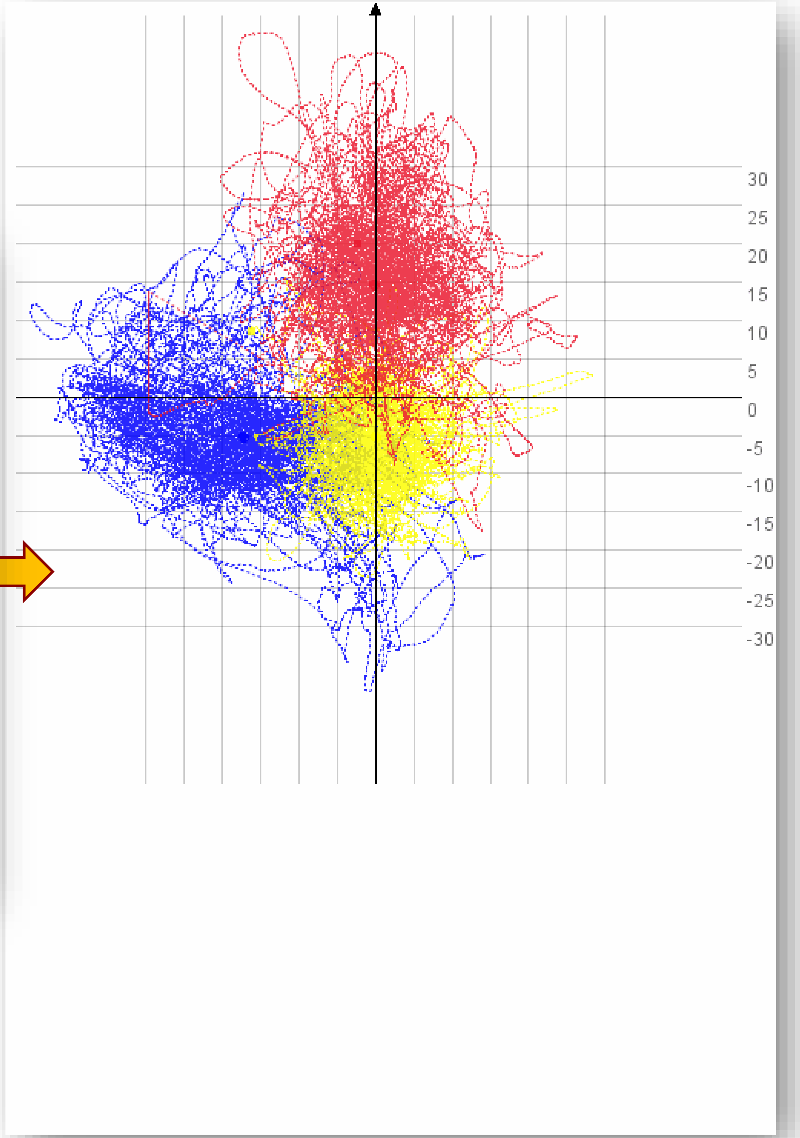
Trajectories in team space



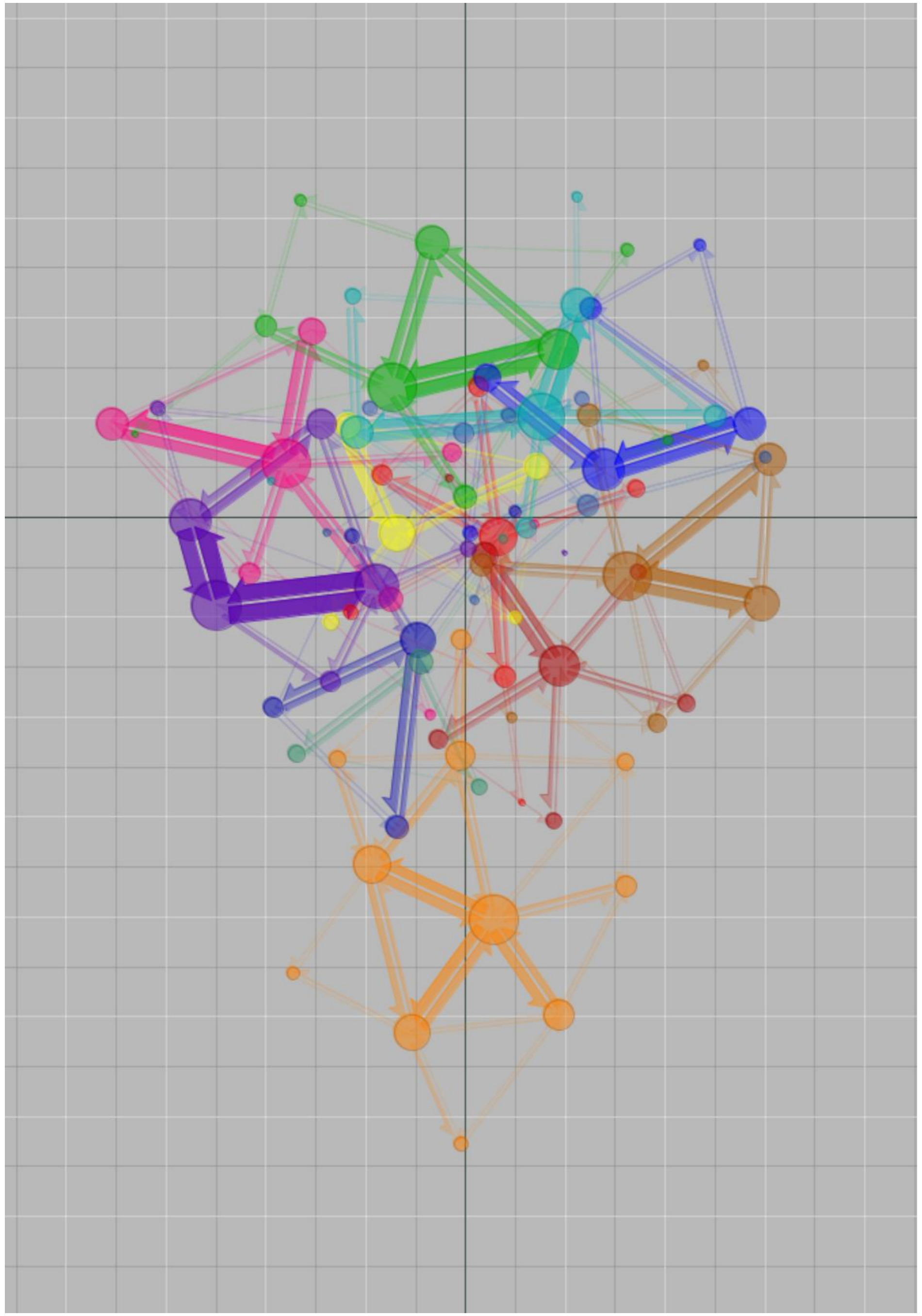
Trajectories in team space

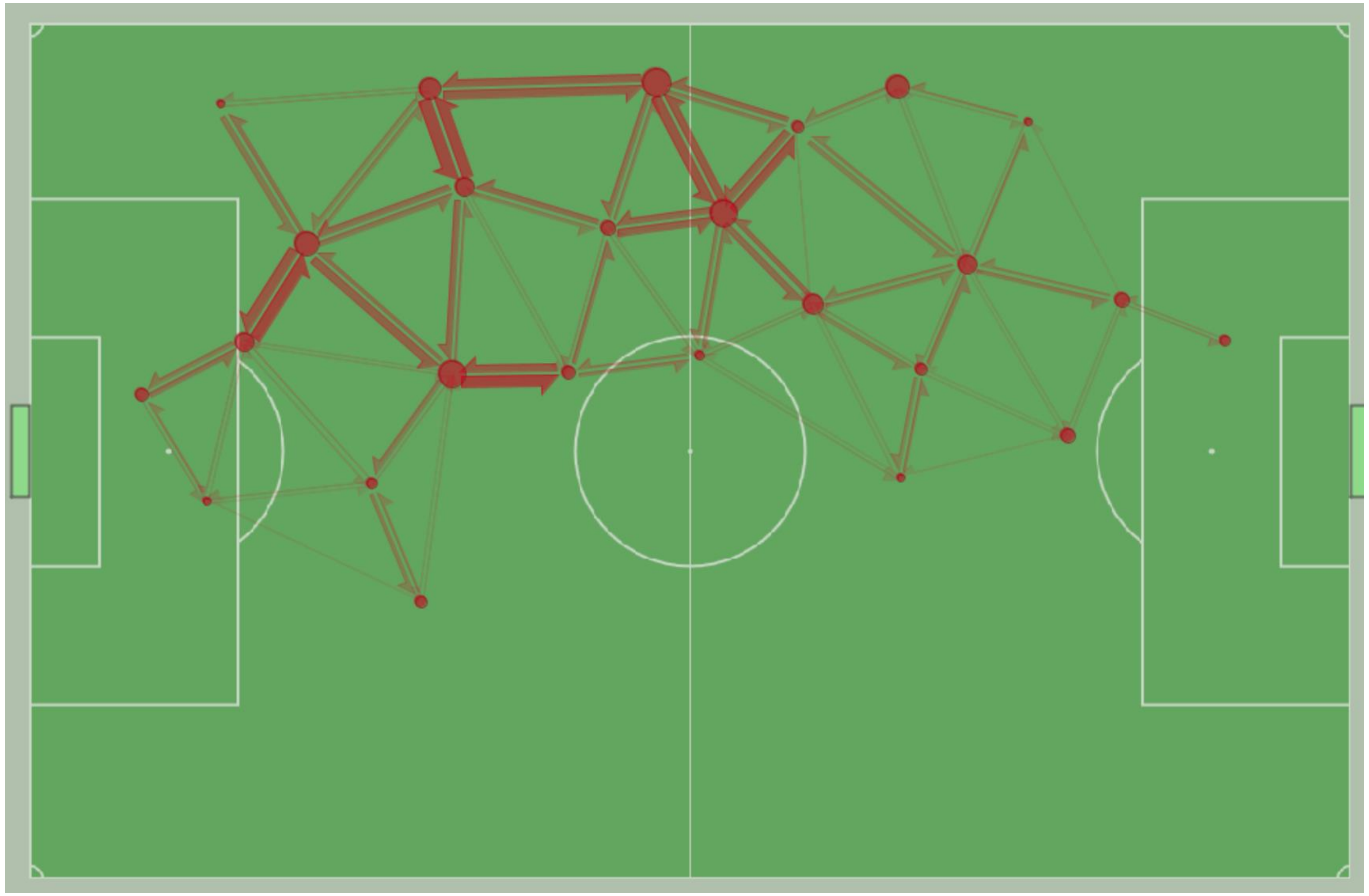
3 players selected



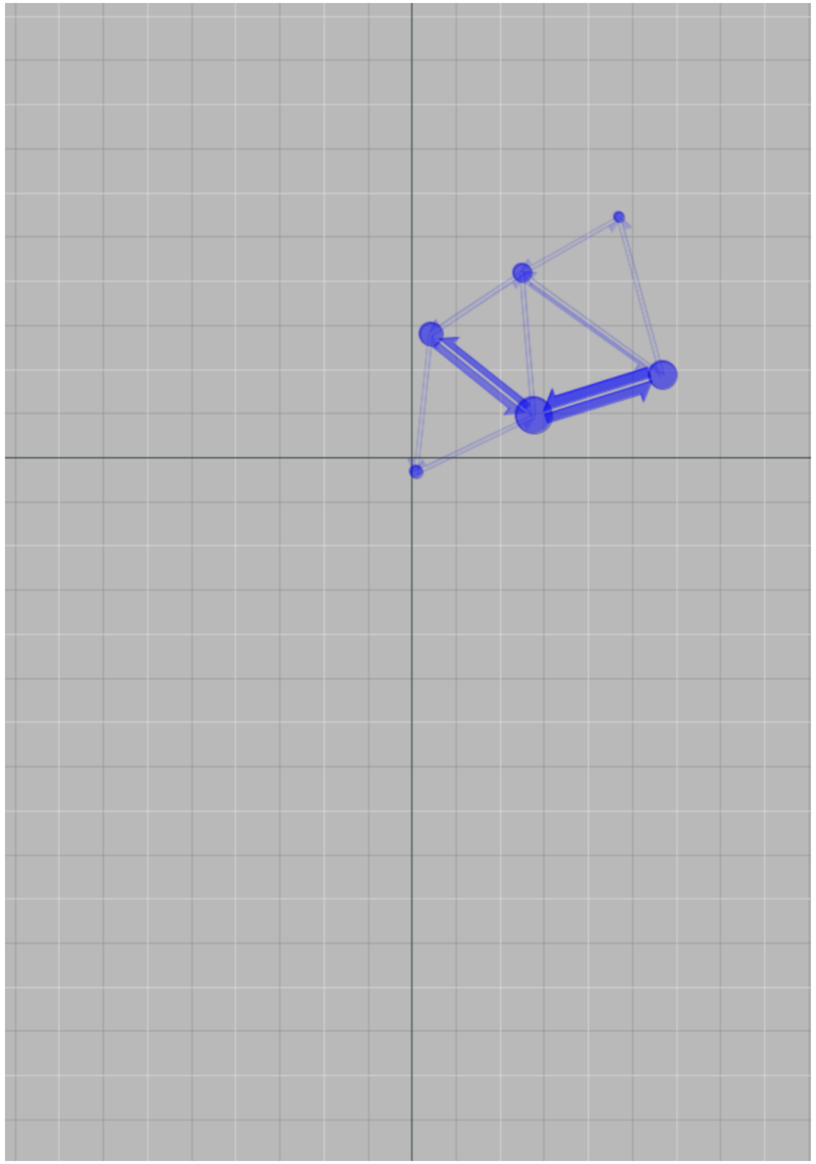


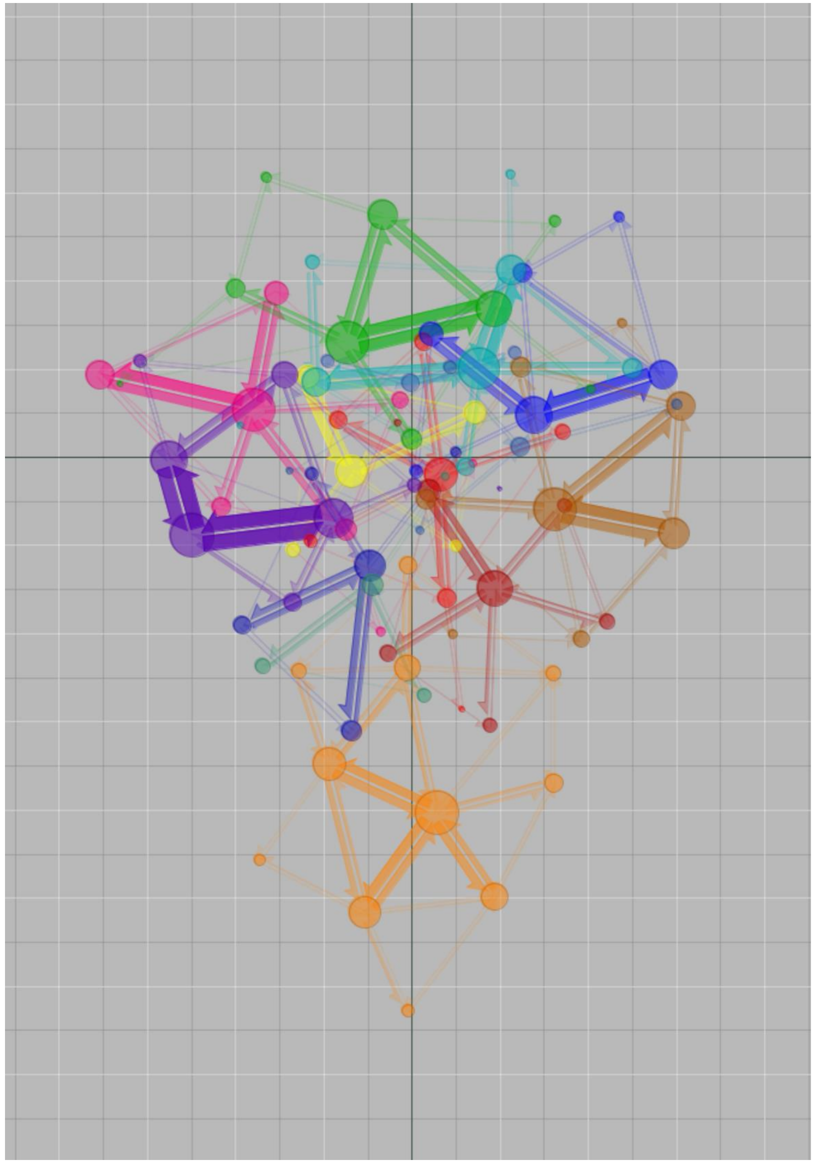
Players’ movement networks in their team’s space



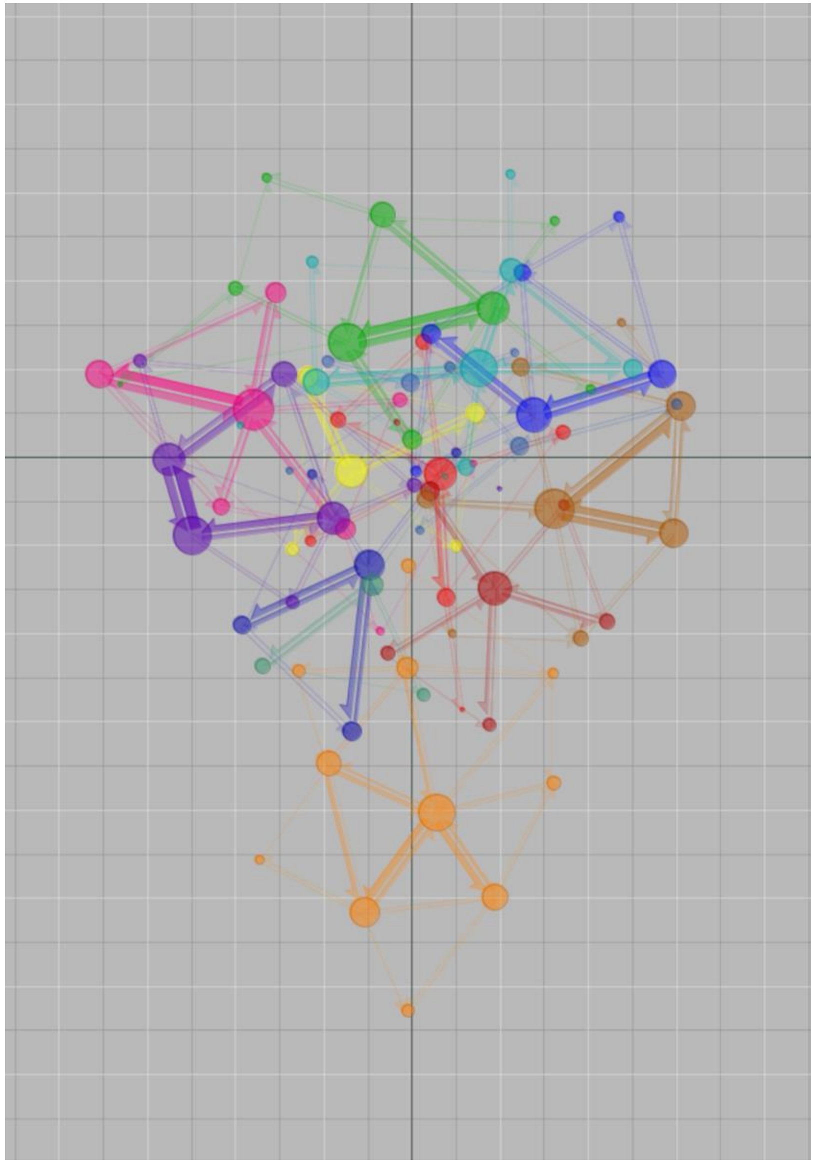


=

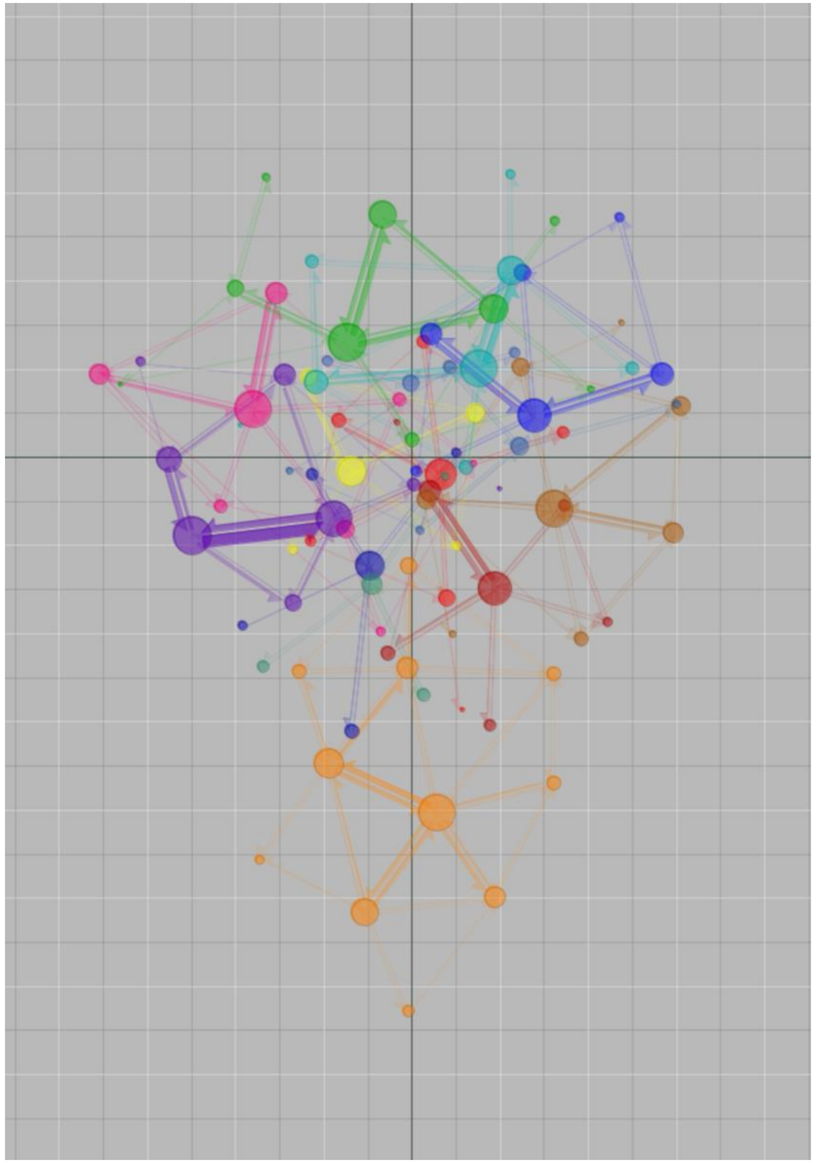




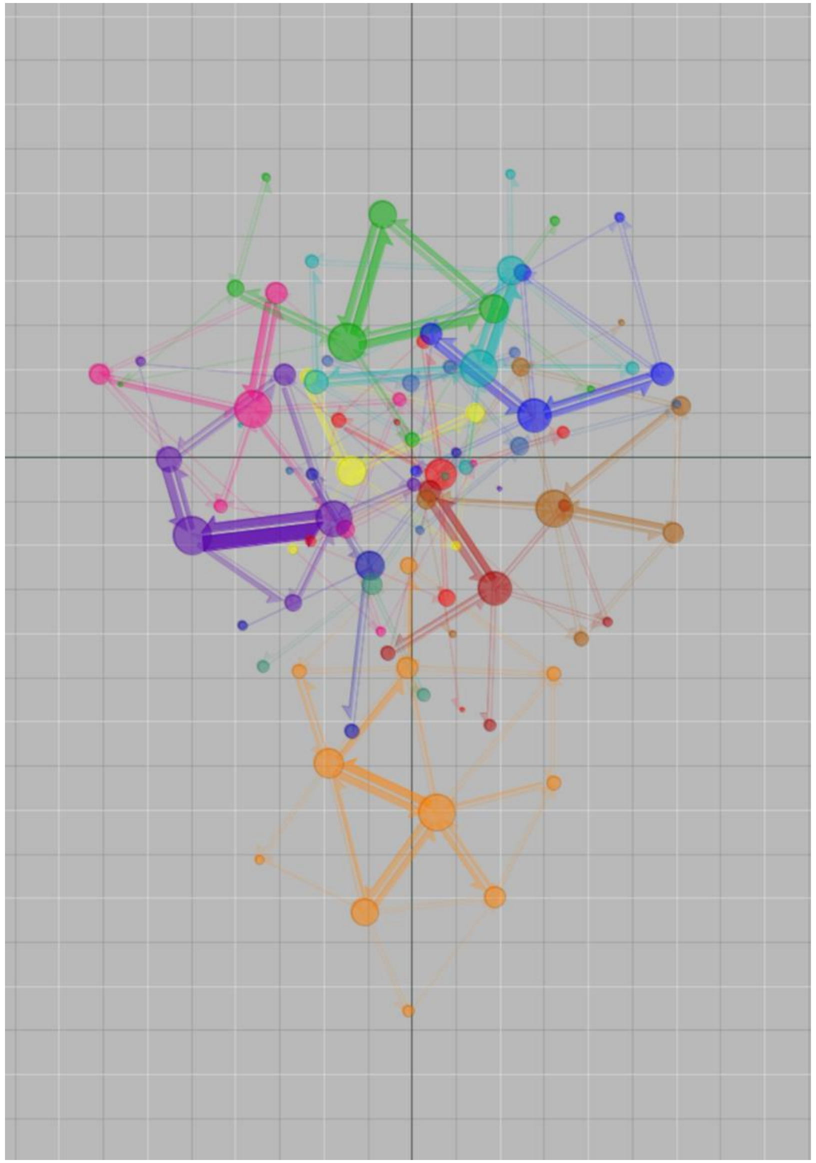
Whole networks



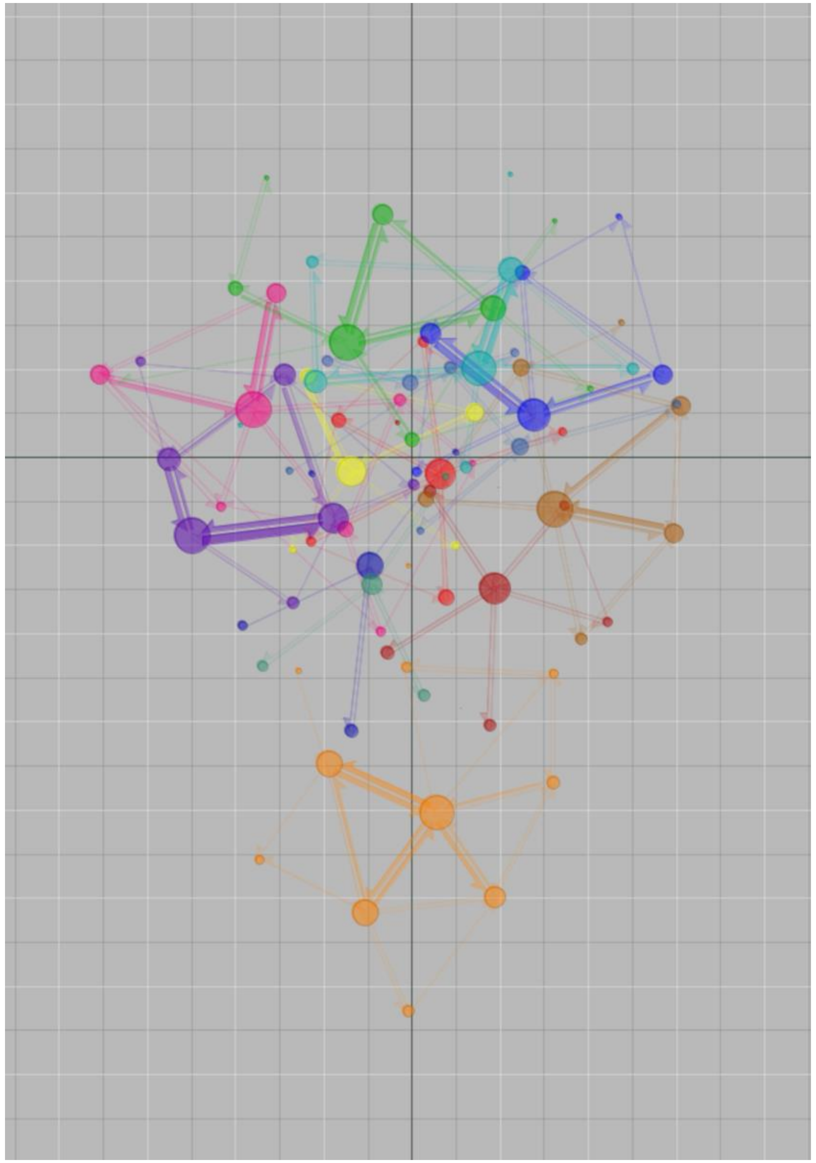
Own ball possession



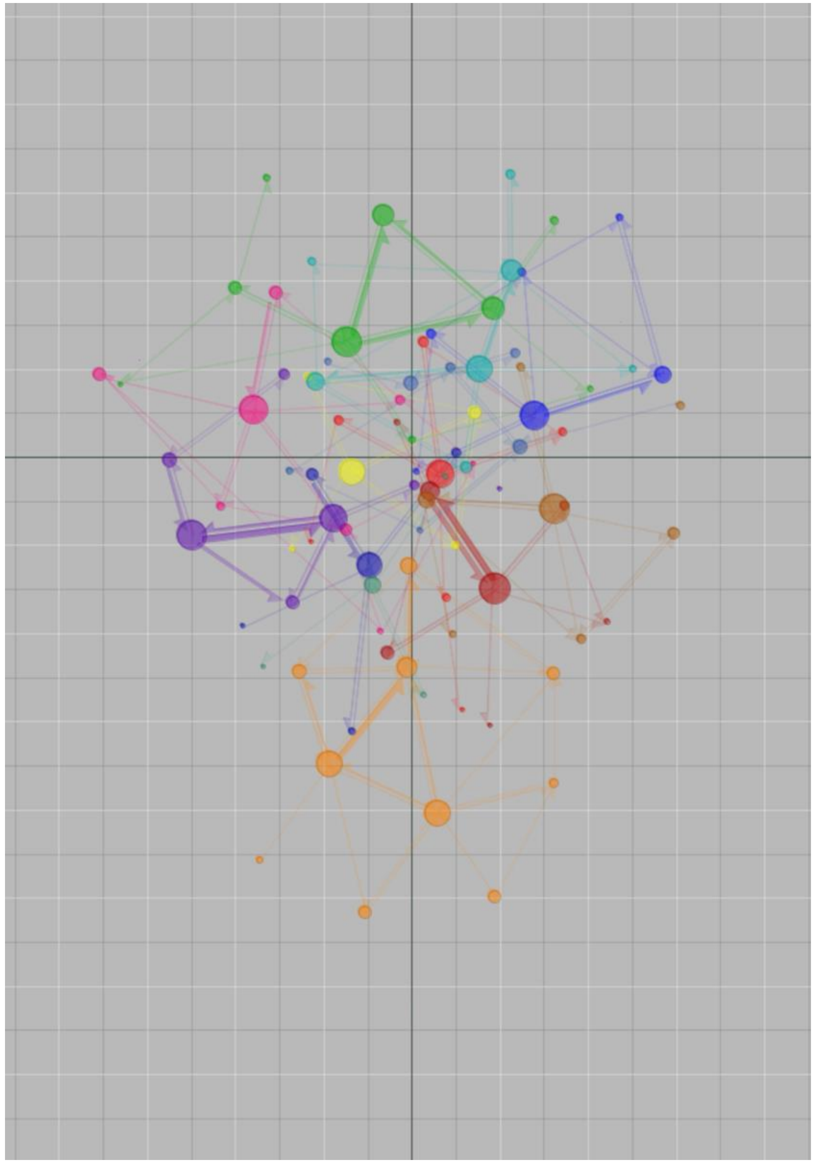
Opponents’ ball possession



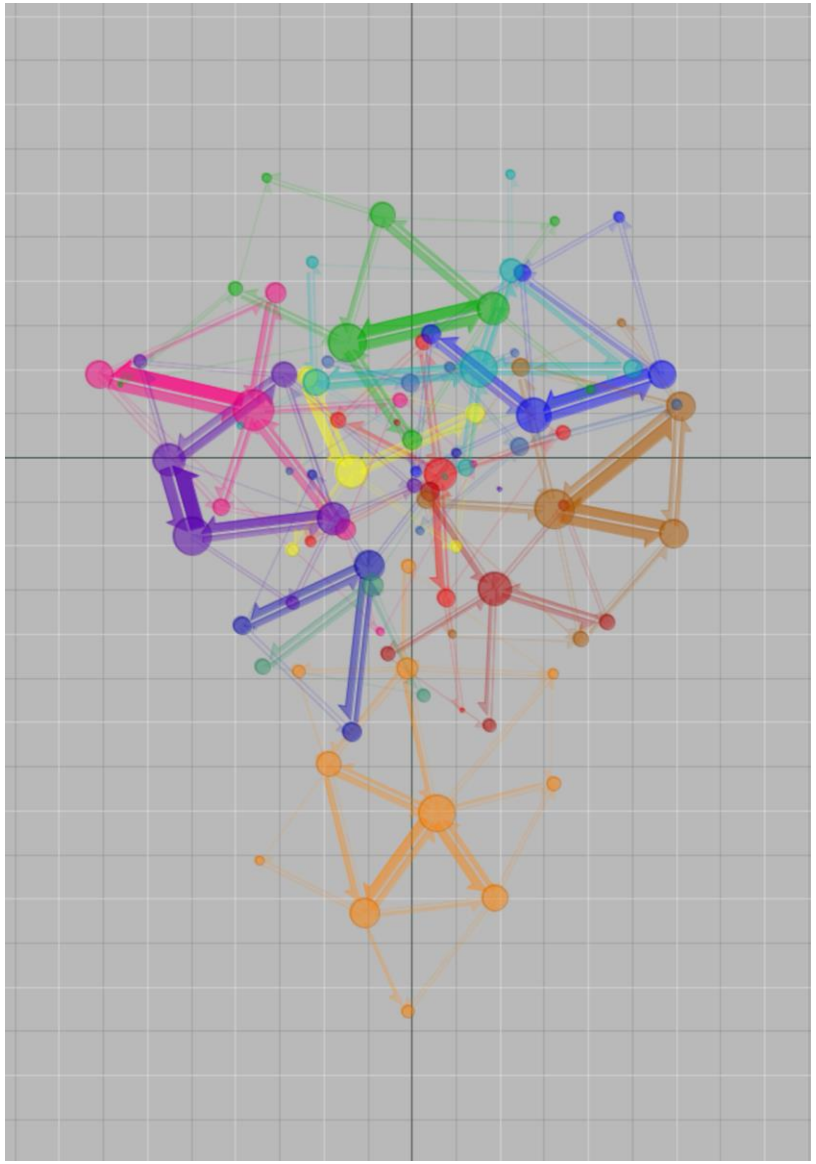
Opponents’ ball possession



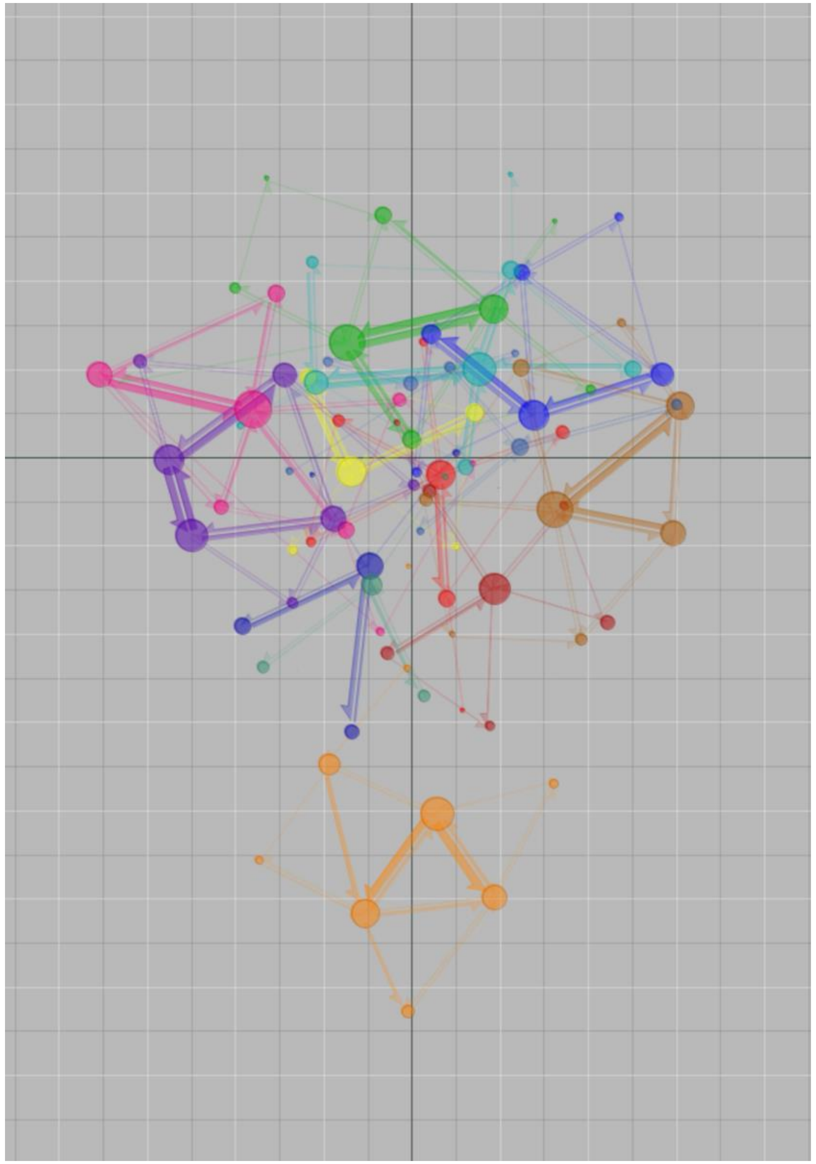
Opponents’ ball possession
Ball on the opponents’ side



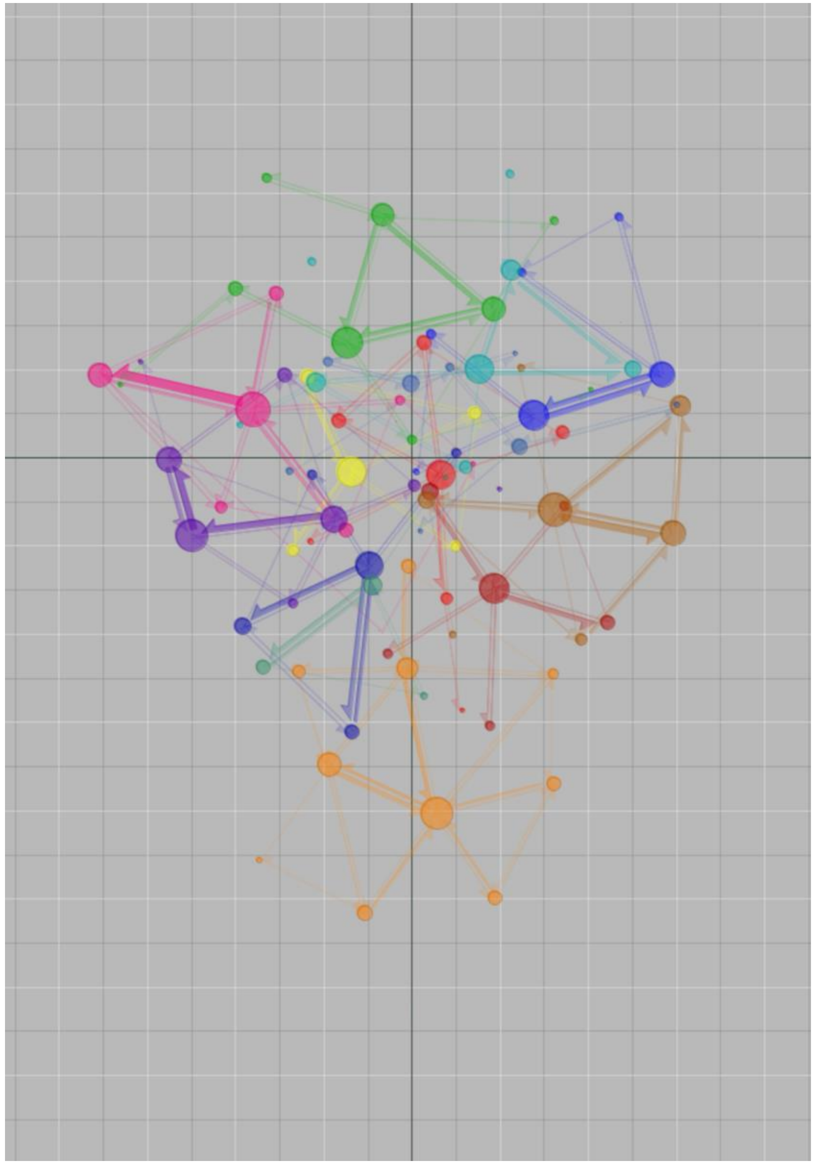
Opponents’ ball possession
Ball on the own side



Own ball possession



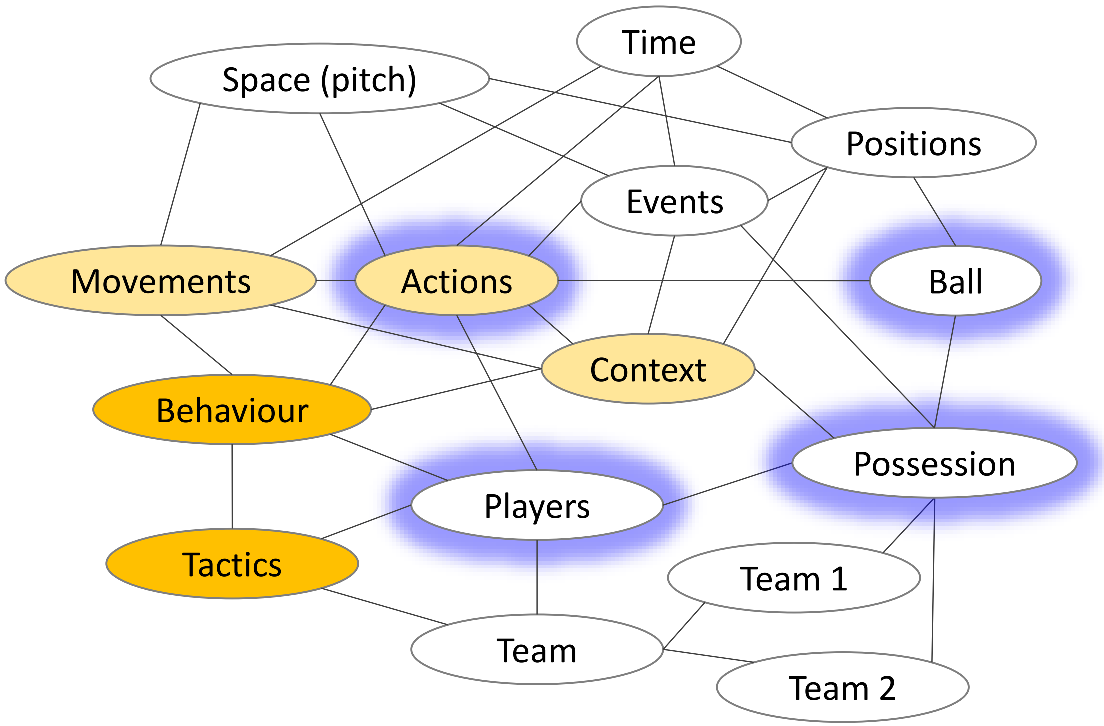
Own ball possession
Ball on the opponents’ side

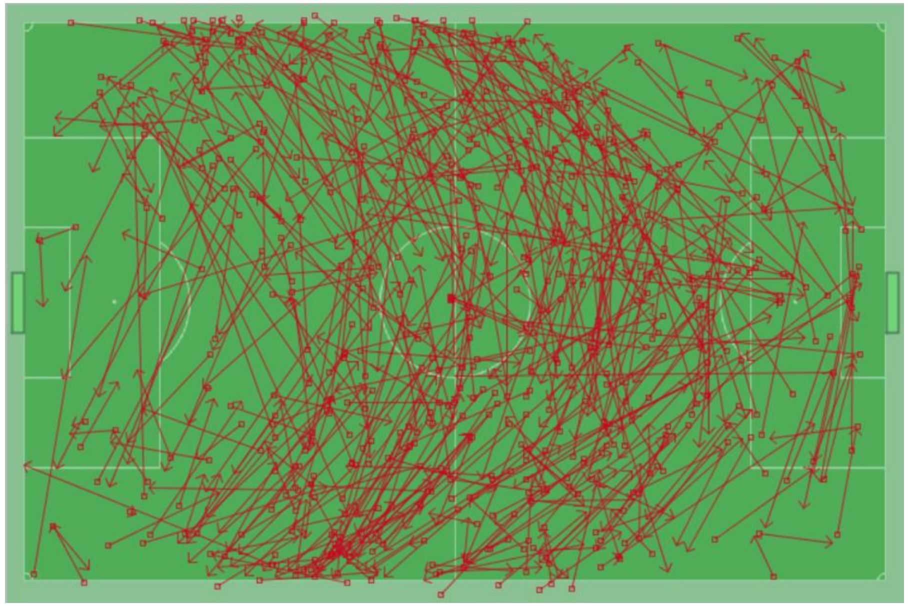
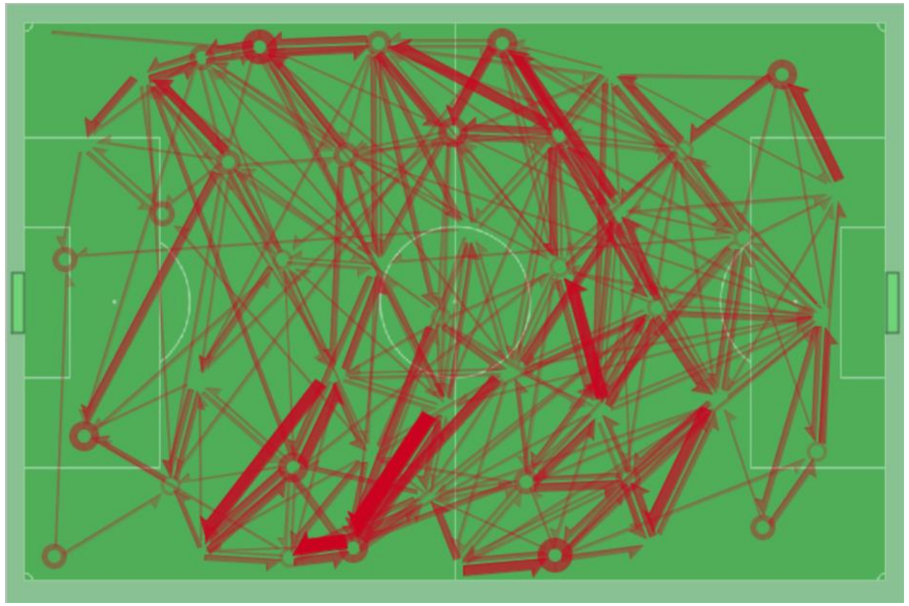
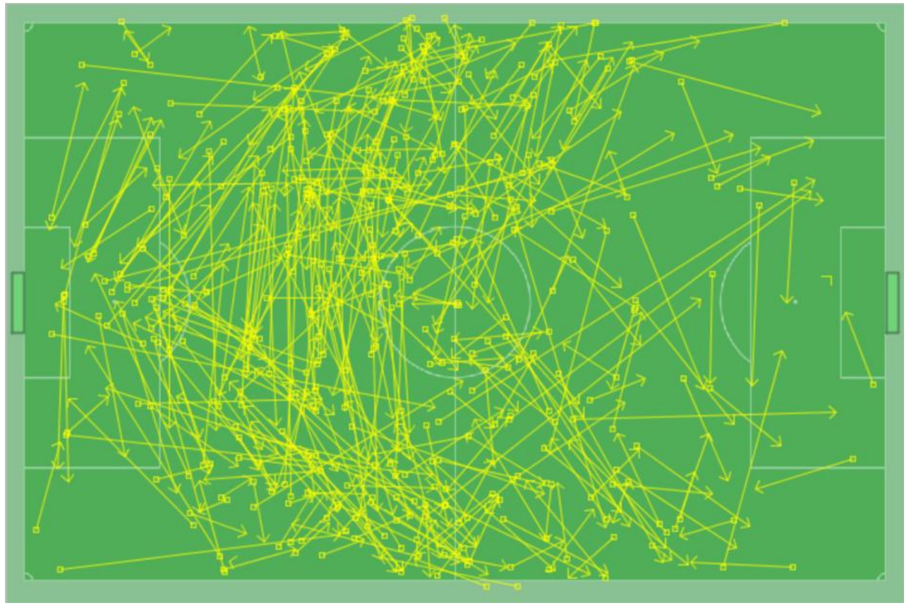
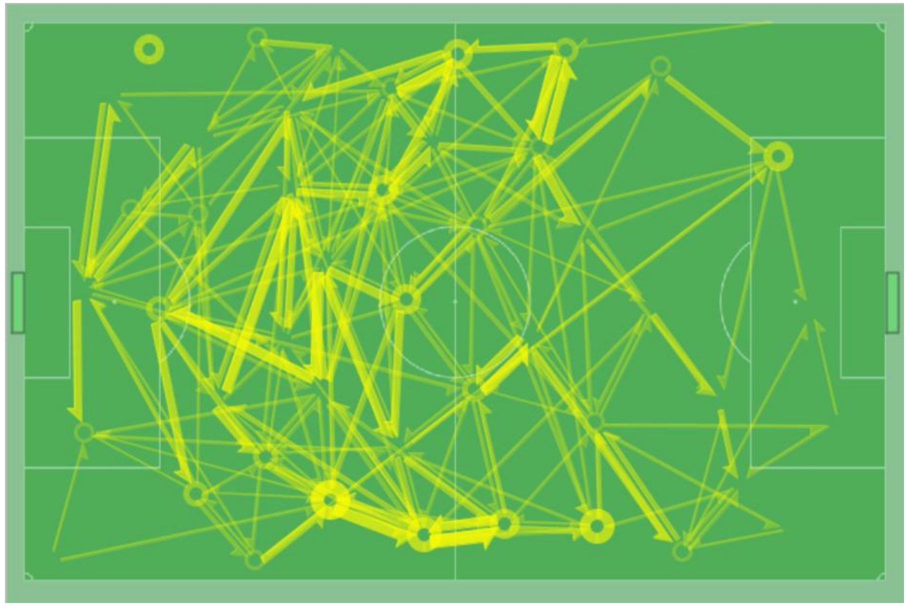


Own ball possession
Ball on the own side

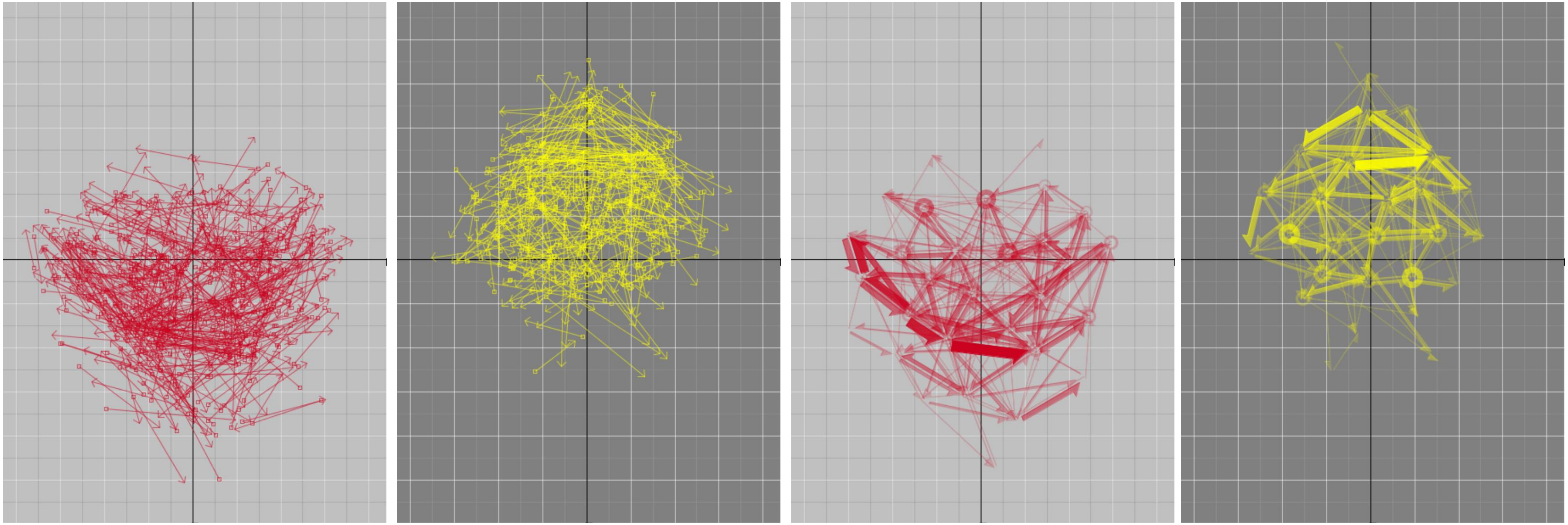
Players’ actions: passes

Spatial patterns of passes

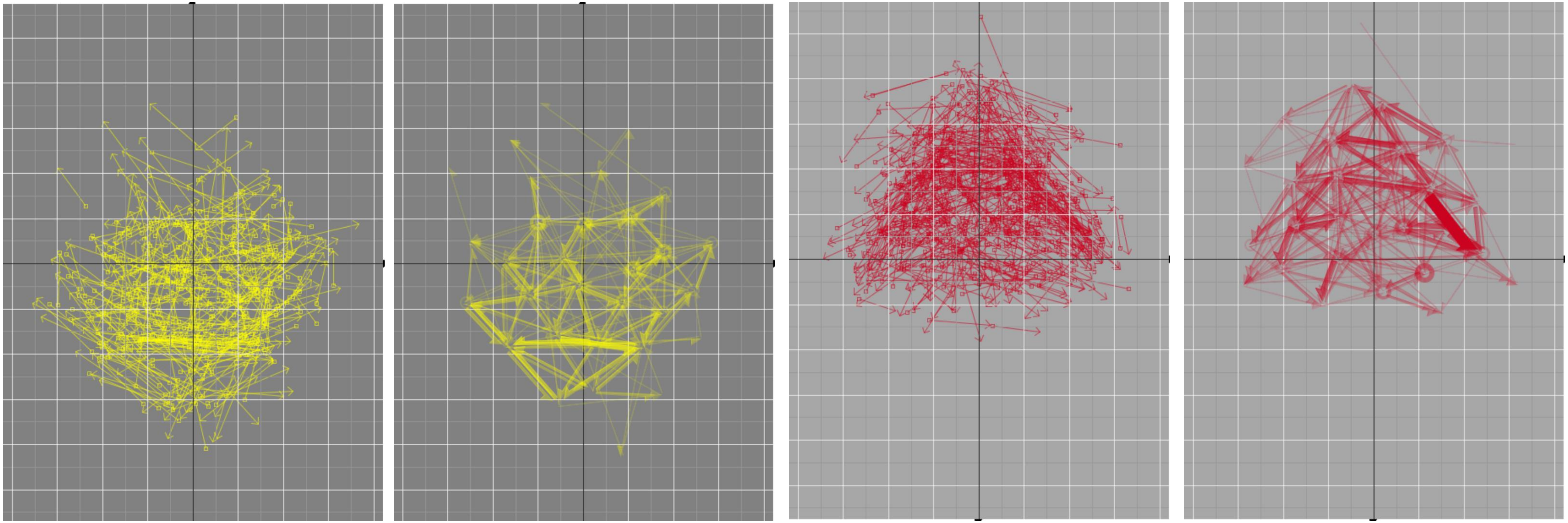


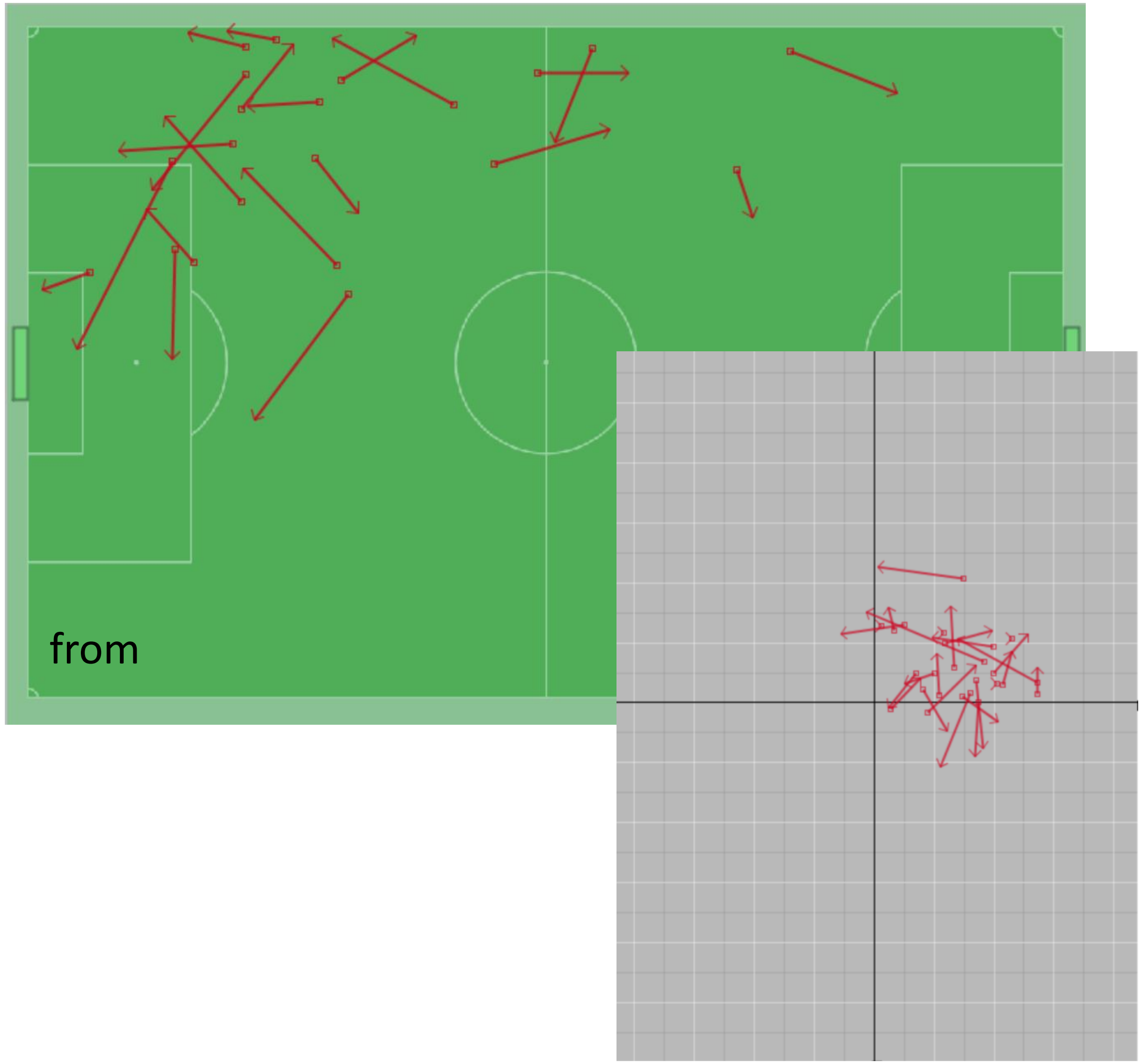
Passes of the two teams in the team space of the red team

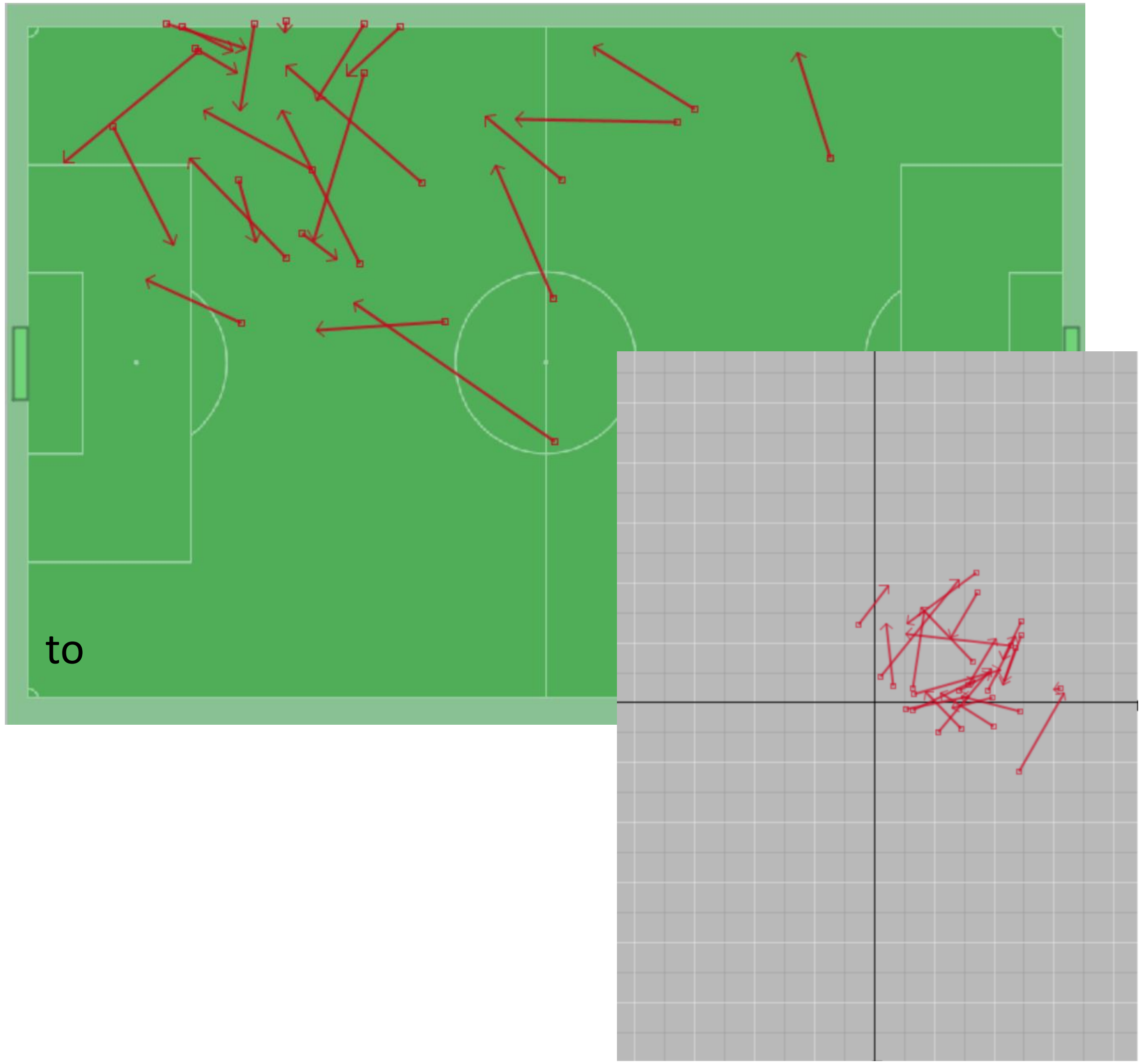


Passes of the two teams in the space of the yellow team



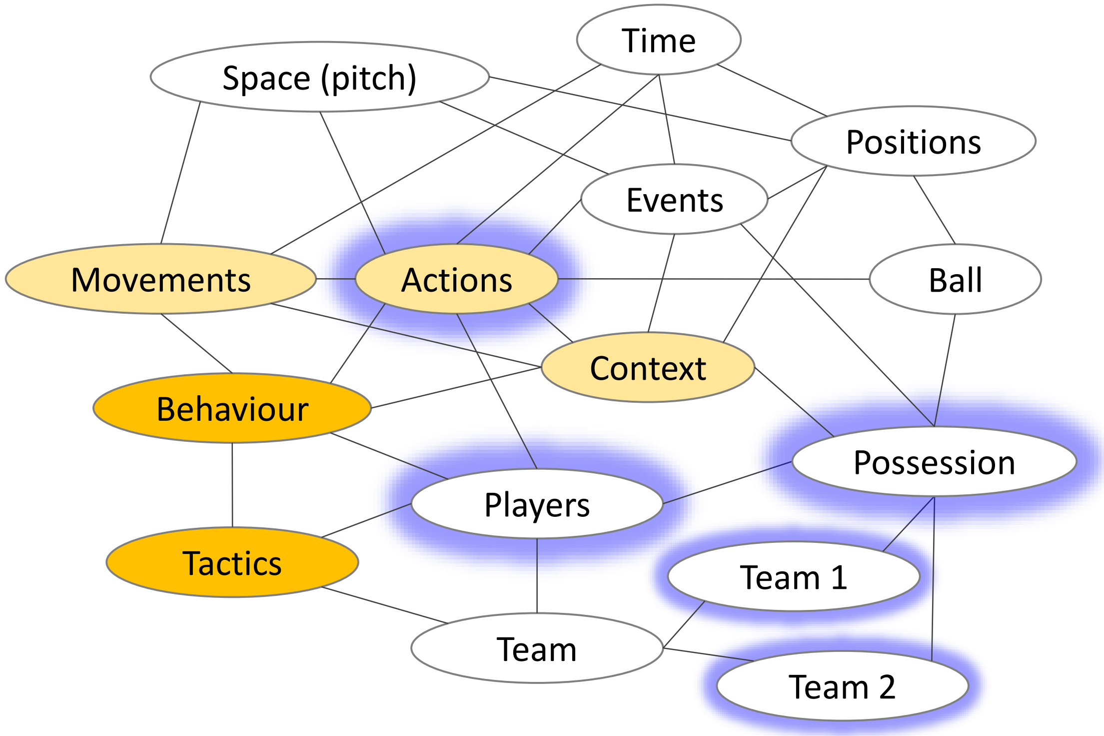
Passes from/to a selected player



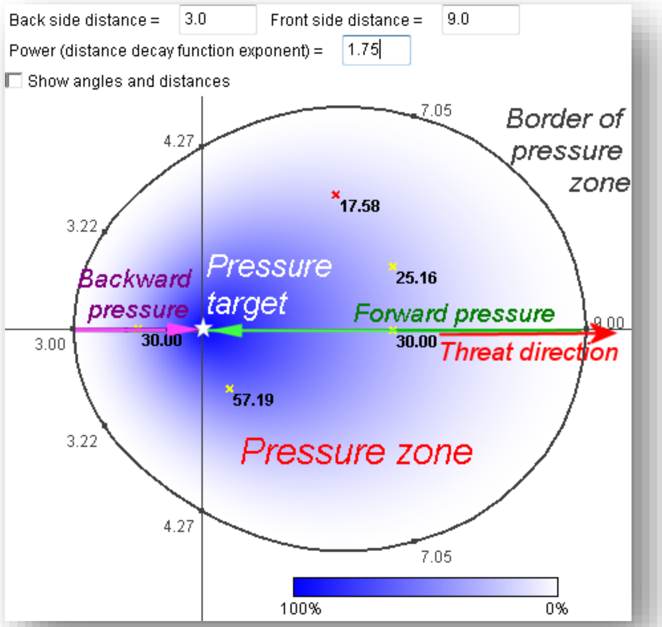


Players’ actions in defence: pressure

Restricting opponents’ possibilities to move and act



Pressure assessment

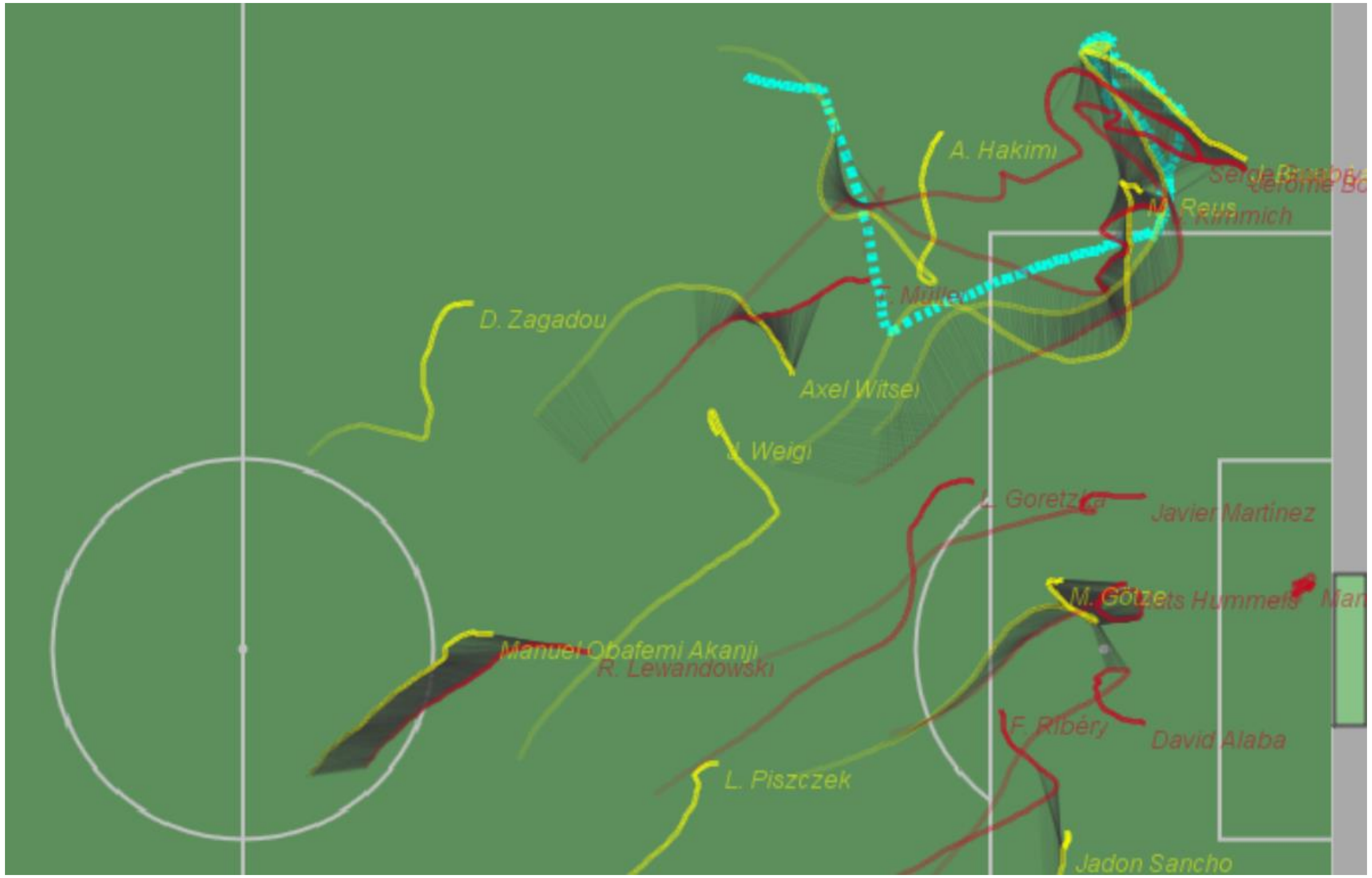


G. Andrienko, N. Andrienko, G. Budziak, J. Dykes, G. Fuchs, T. von Landesberger, H. Weber

Visual Analysis of Pressure in Football

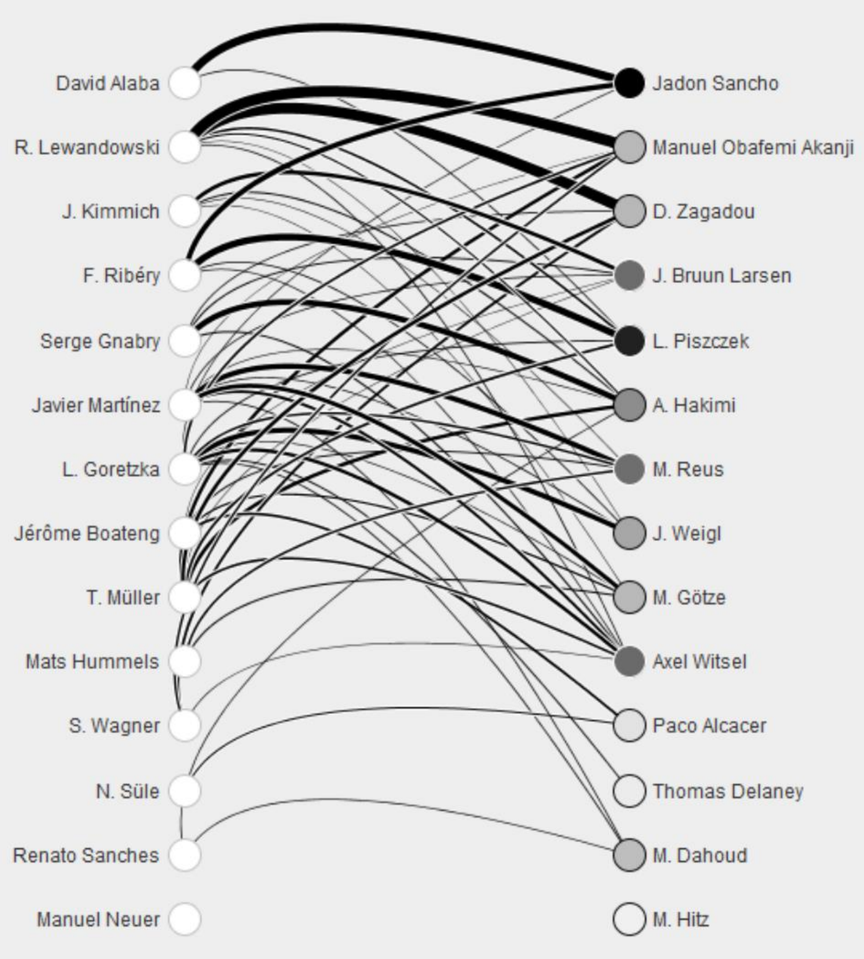
Data Mining and Knowledge Discovery
31(6): 1793-1839, 2017

http://dx.doi.org/10.1007/s10618-017-0513-2

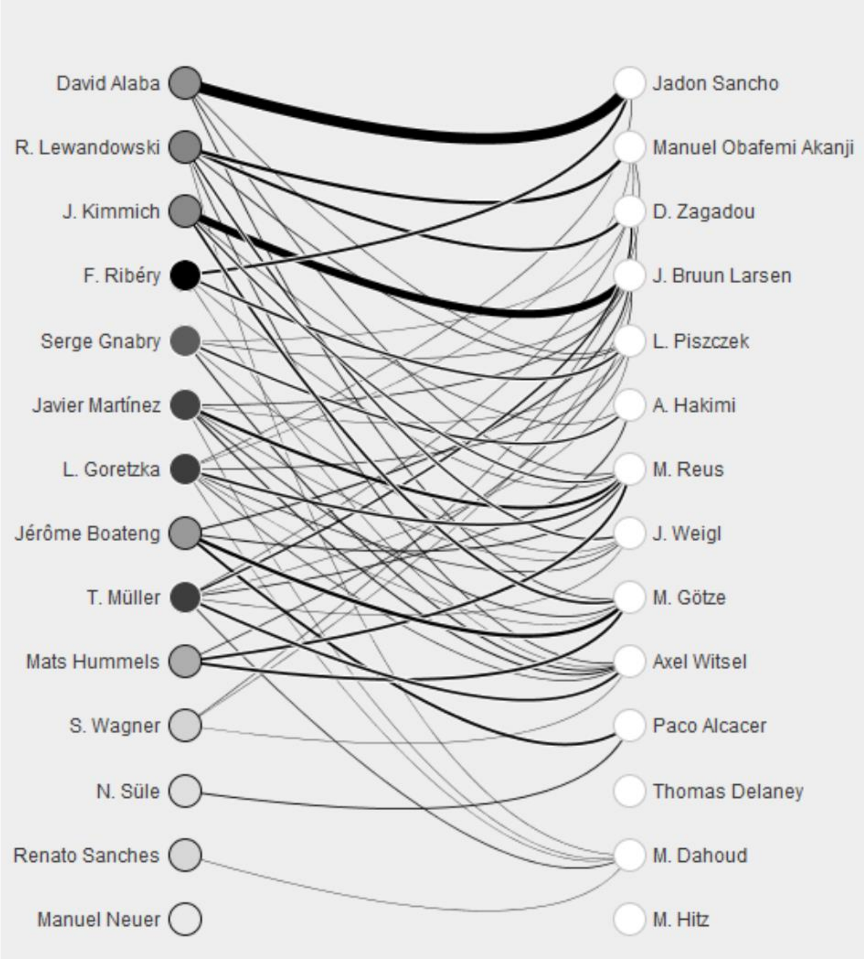


Pressure of players of the yellow team onto opponents in a selected episode

Pressure graphs: distribution of pressure over players

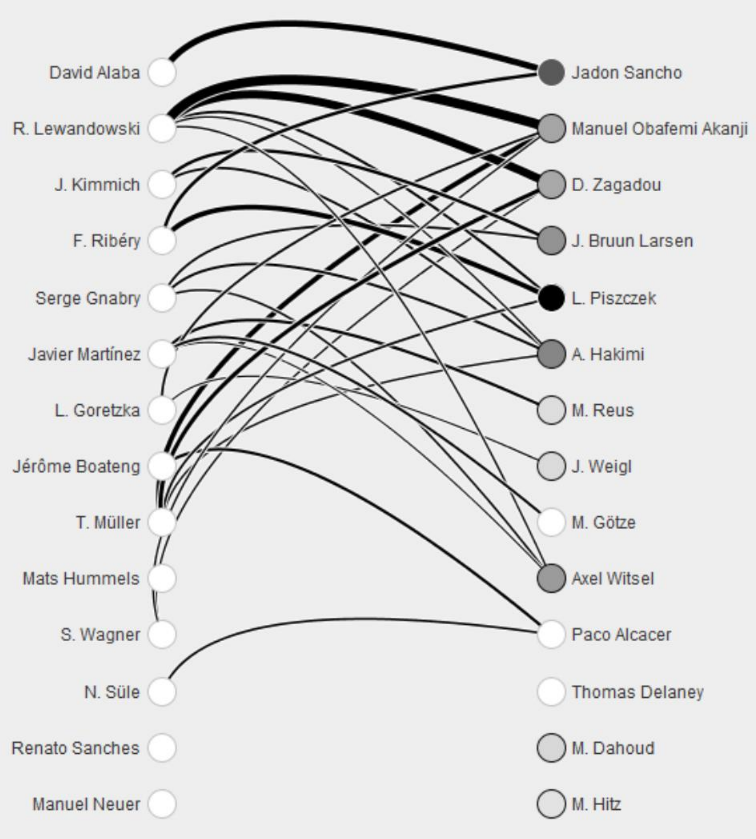


Yellow on red
53552 frames; 55.12% time

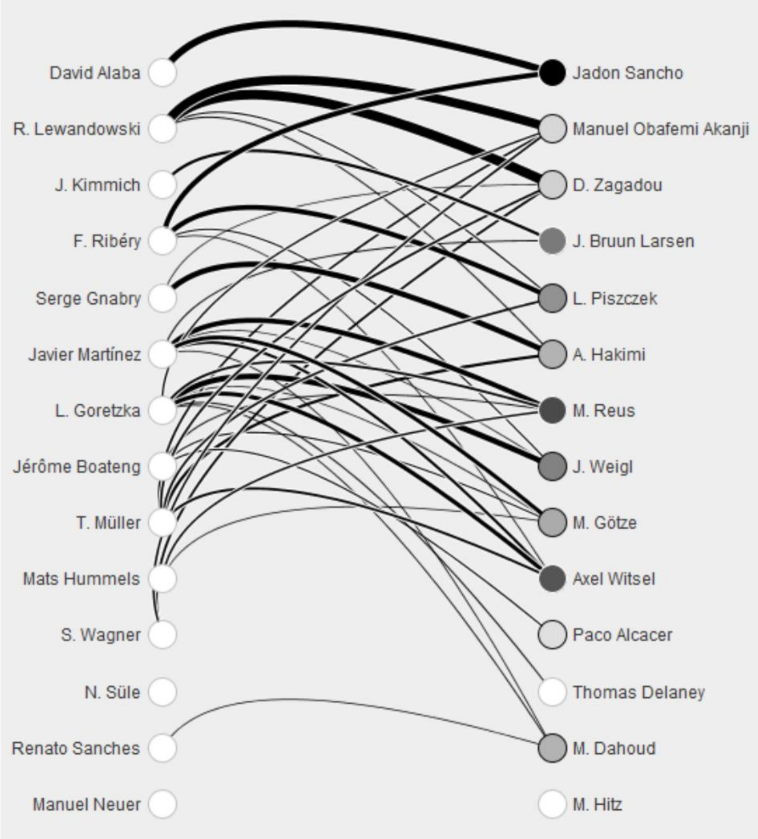


Red on yellow
43607 frames; 44.88% time

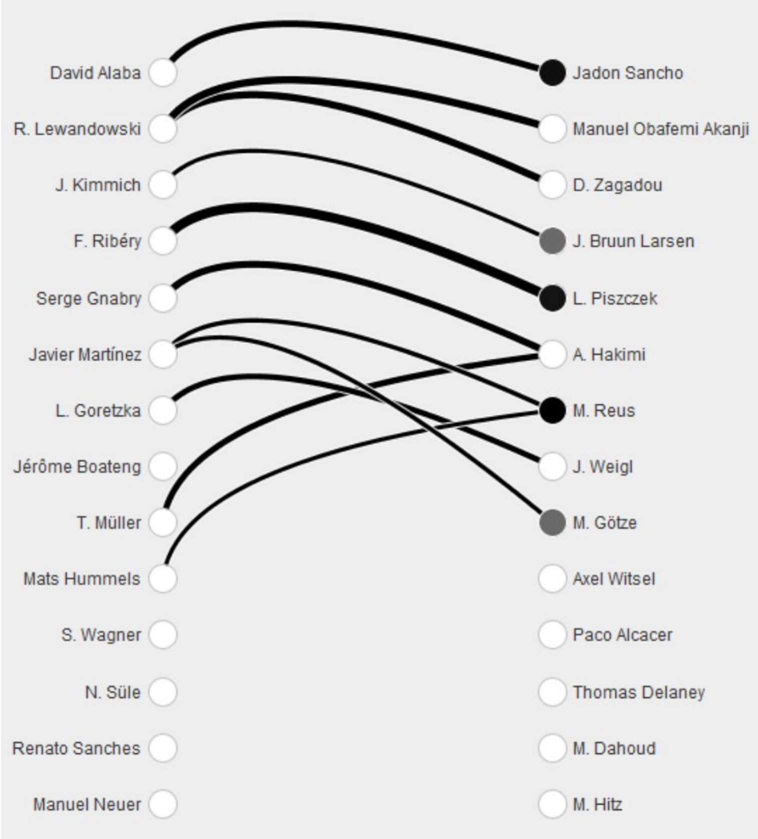
Yellow team's pressure depending on ball position



Own third
11466 frames
11.8% time

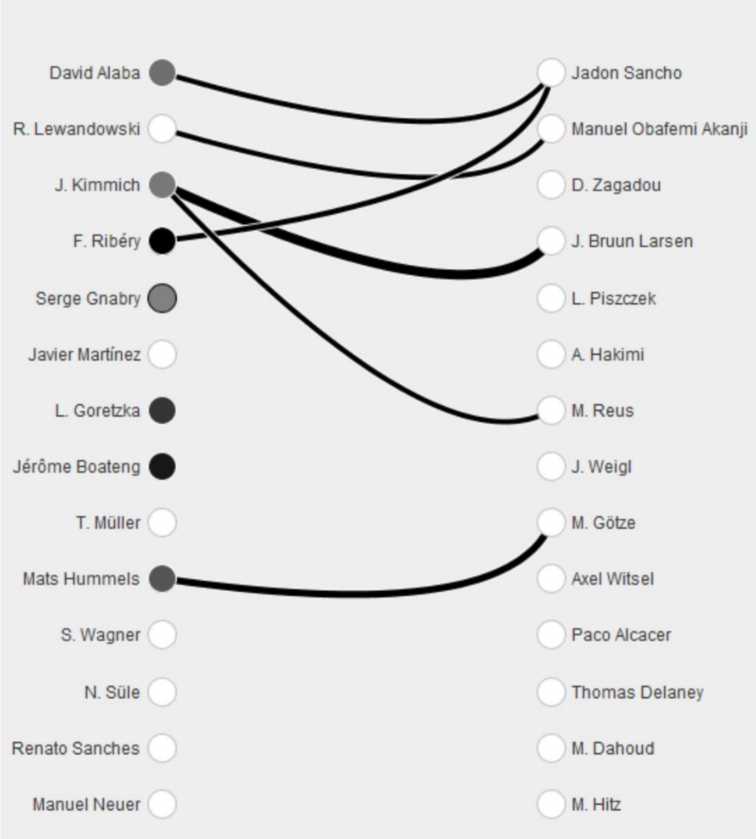


Central third
31847 frames
32.8% time

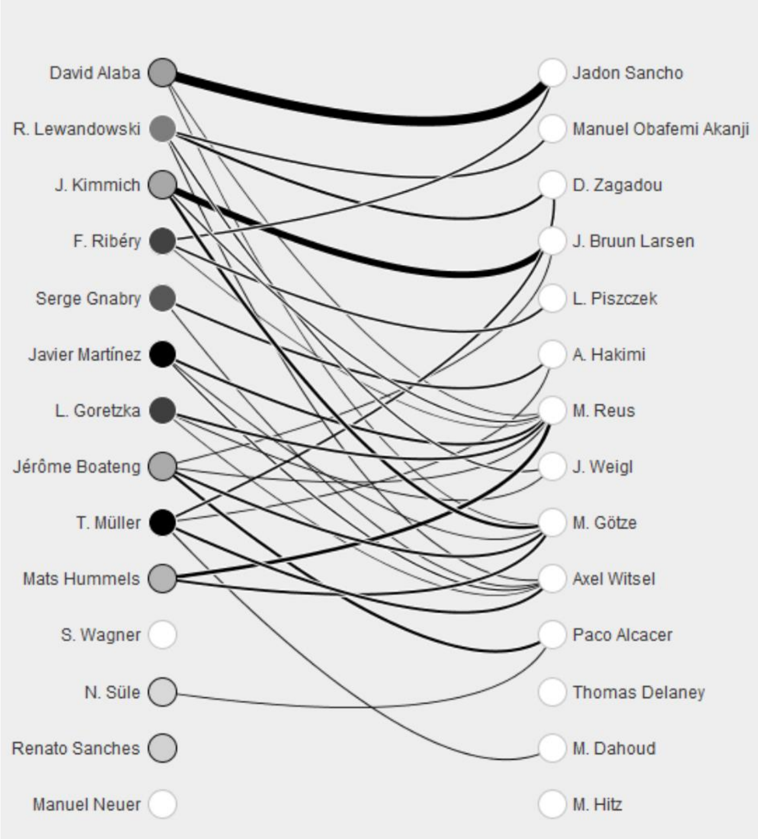


Opponents' third
10252 frames
10.55% time

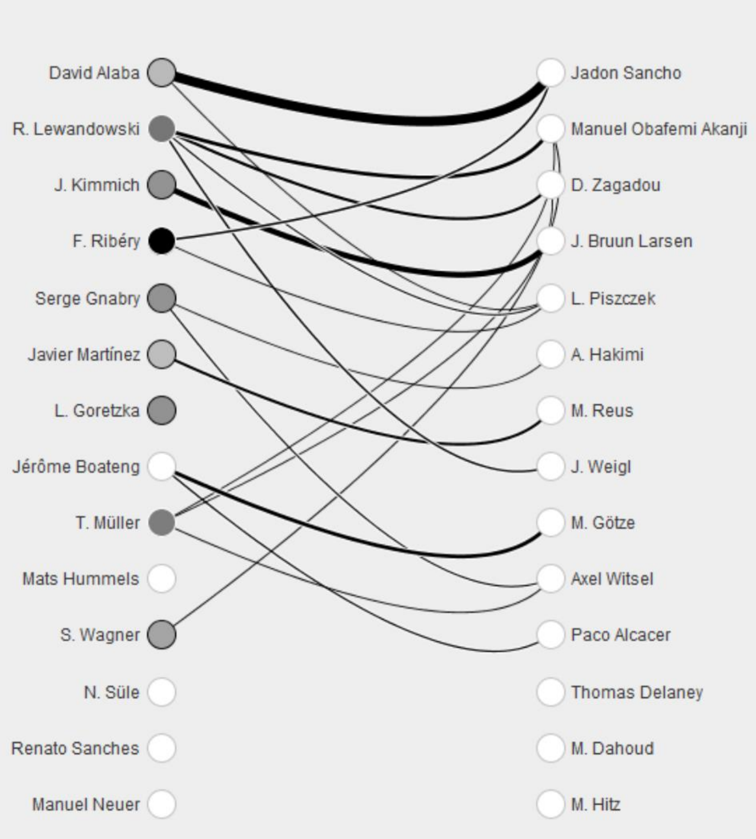
Red team's pressure depending on the ball position



Own third
4966 frames
5% time

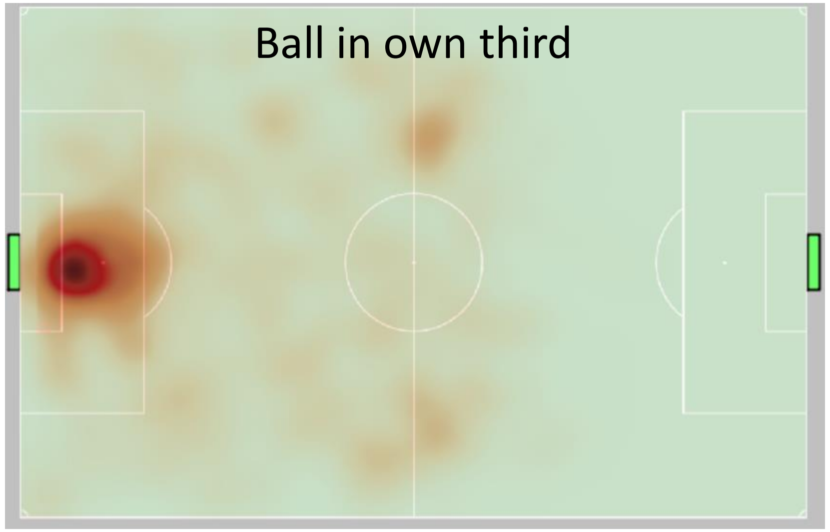


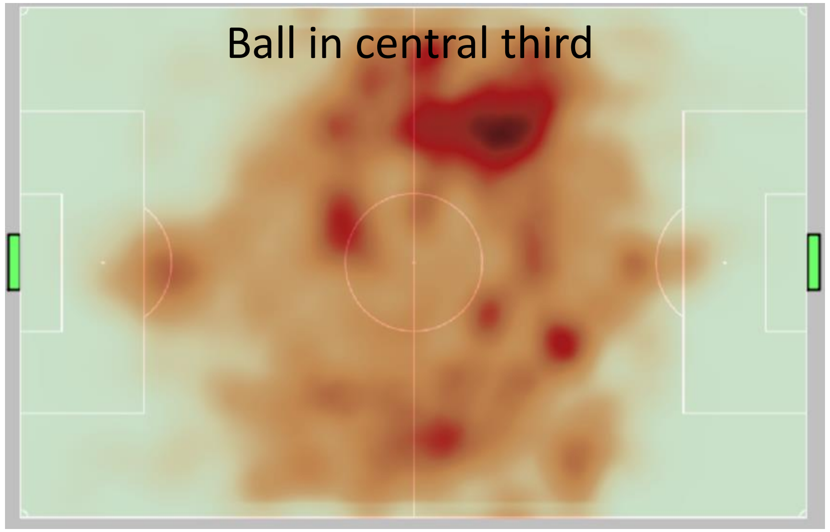
Central third
25641 frames
26.4% time

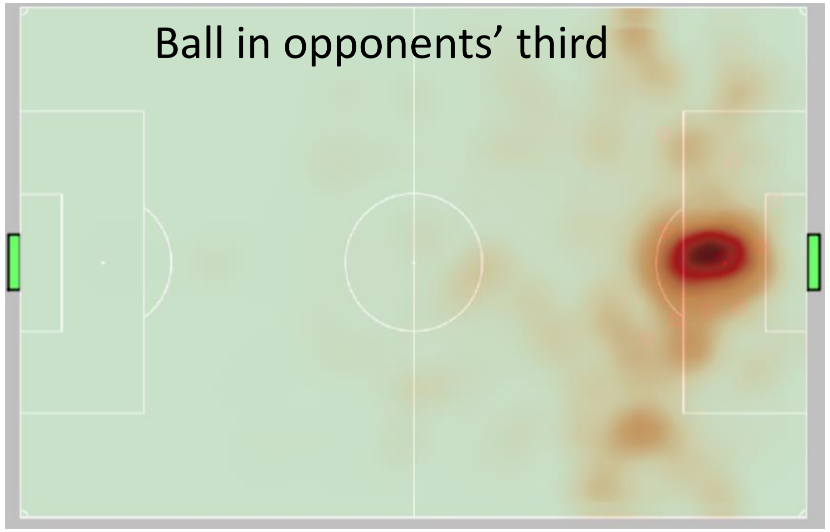


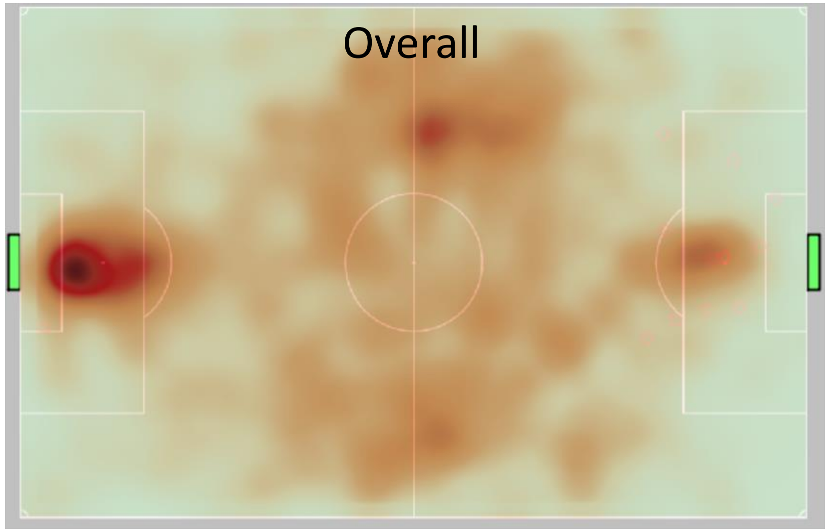
Opponents' third
13013 frames
13.4% time

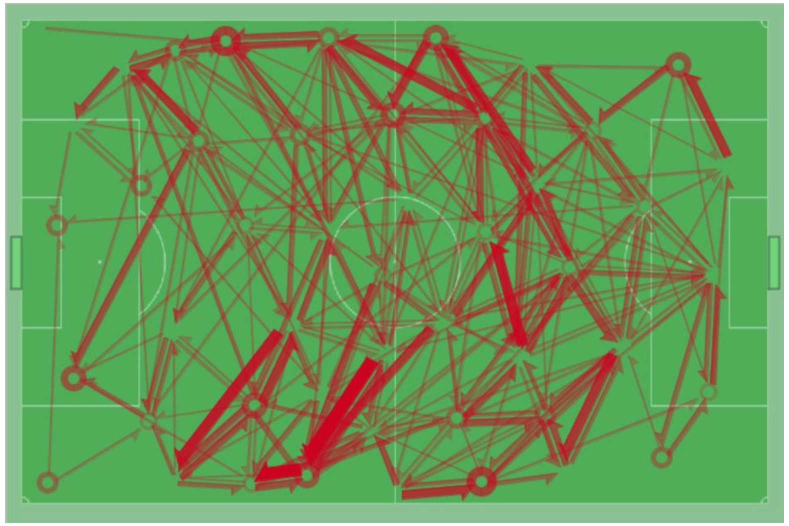
Spatial distribution of the yellow team’s pressure





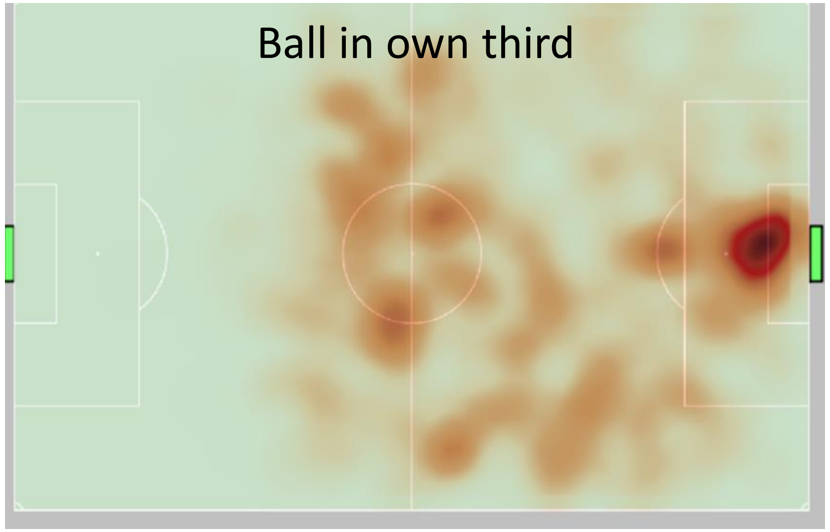


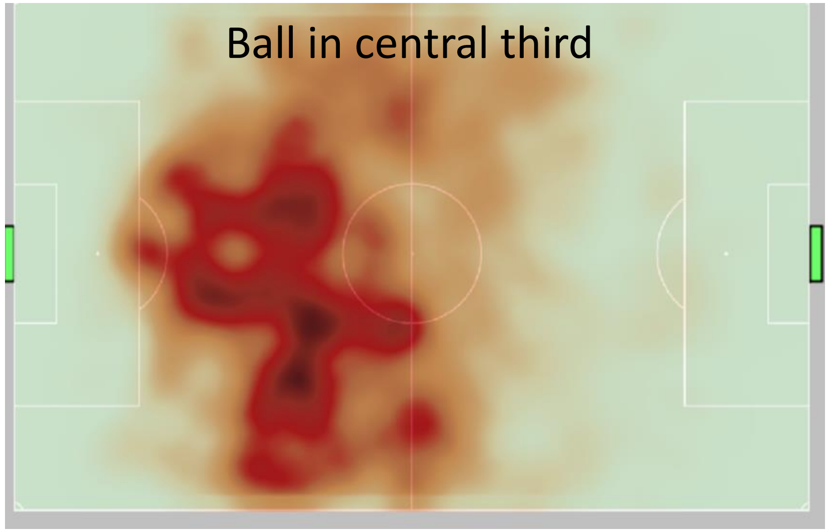


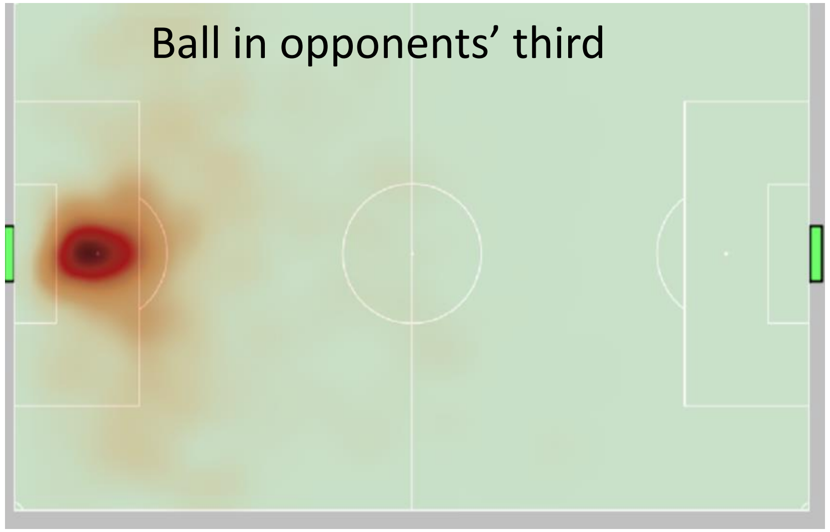


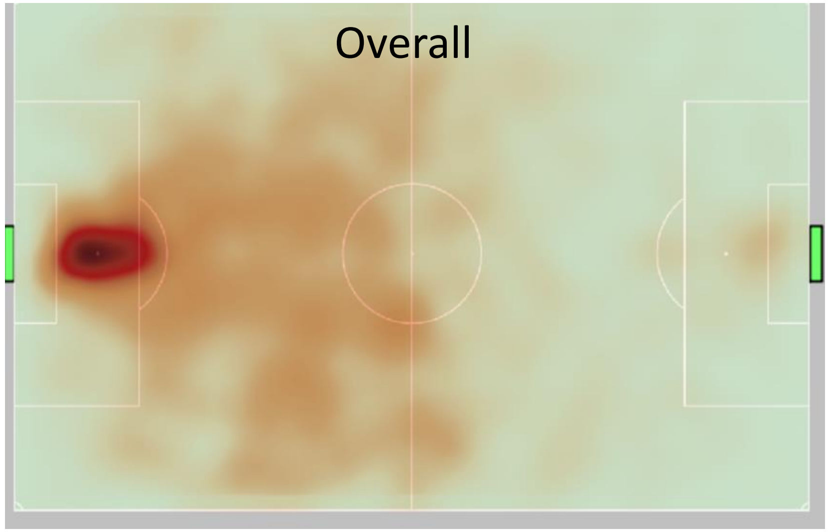
Red team’s passes

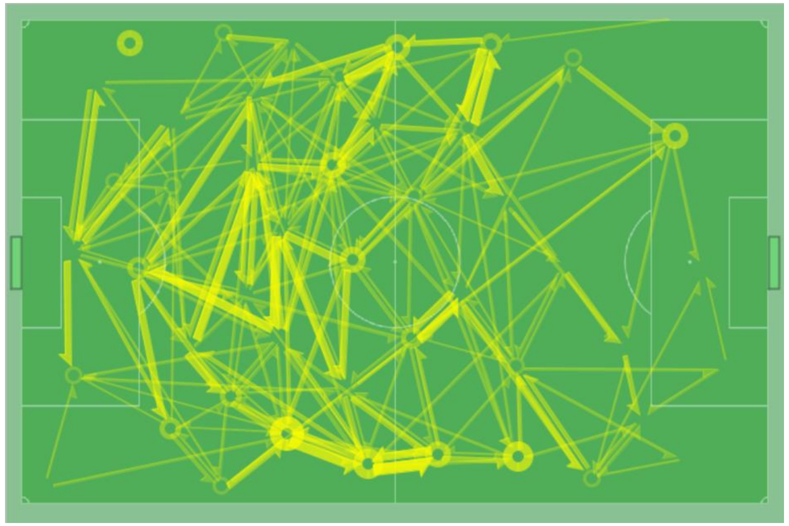
Spatial distribution of the red team's pressure





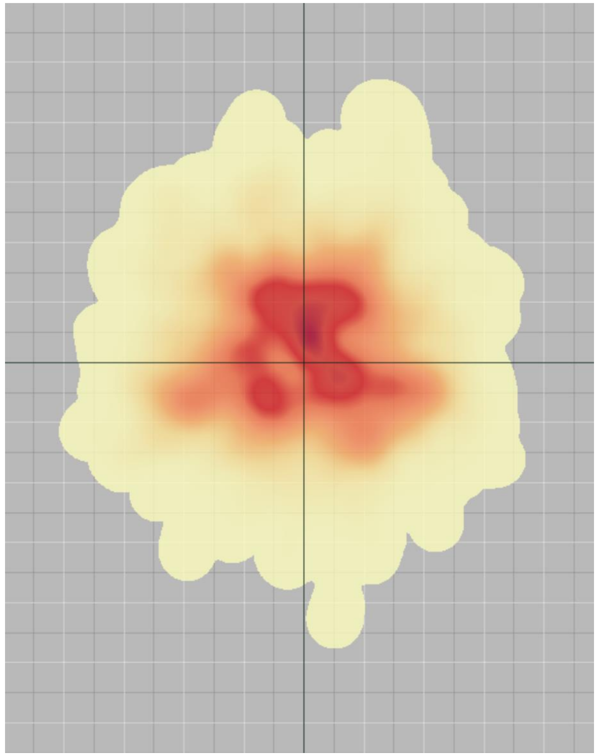




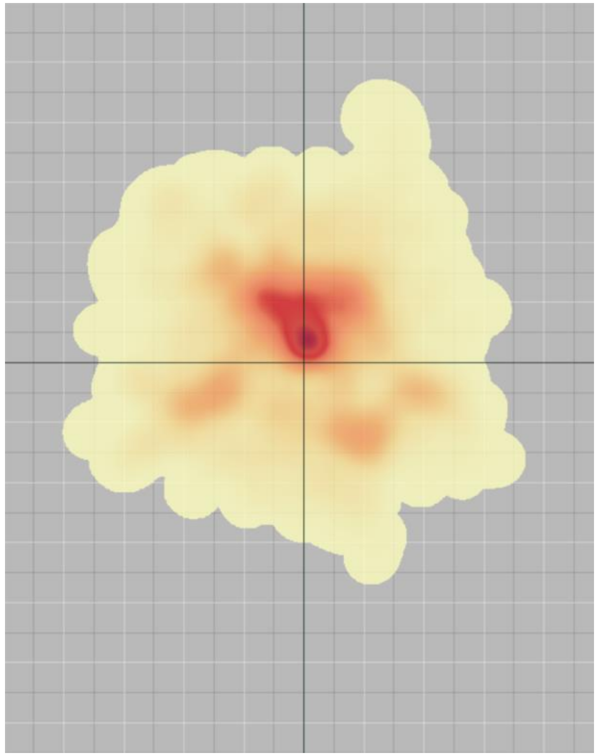


Yellow team's passes

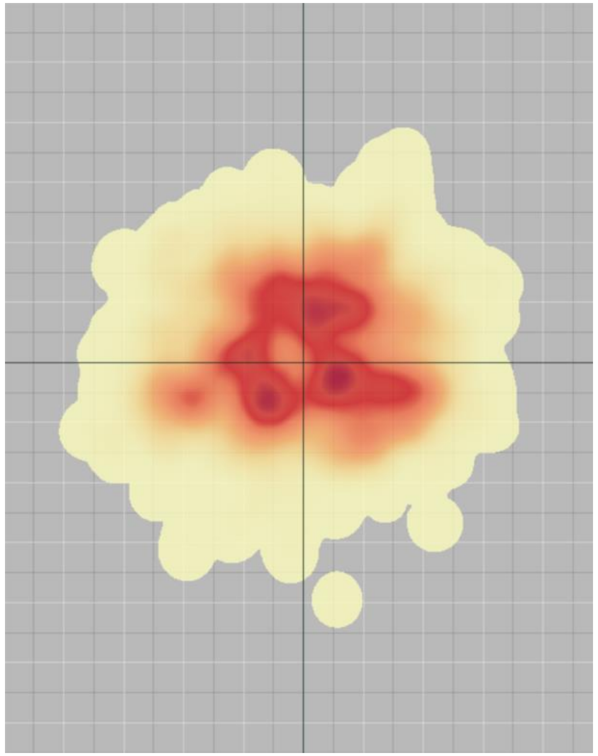
Distribution of the yellow team's pressure in the red team's space



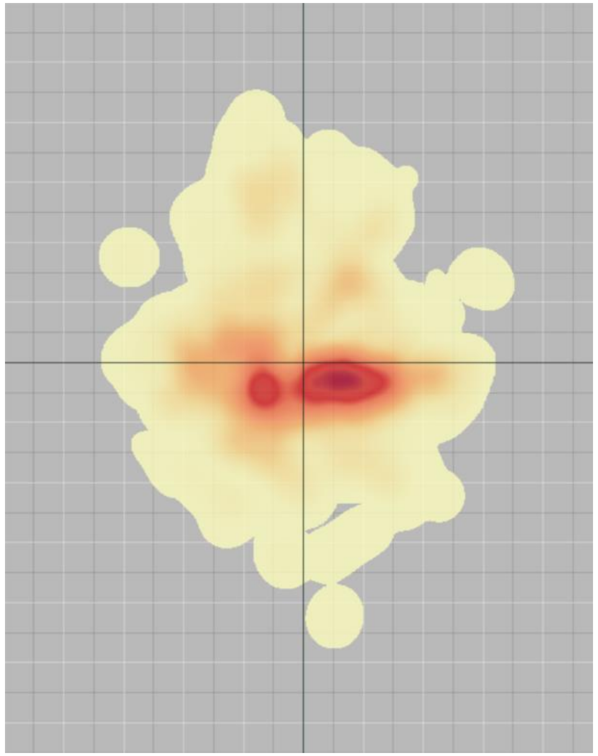
Overall



Ball in own (yellow) third

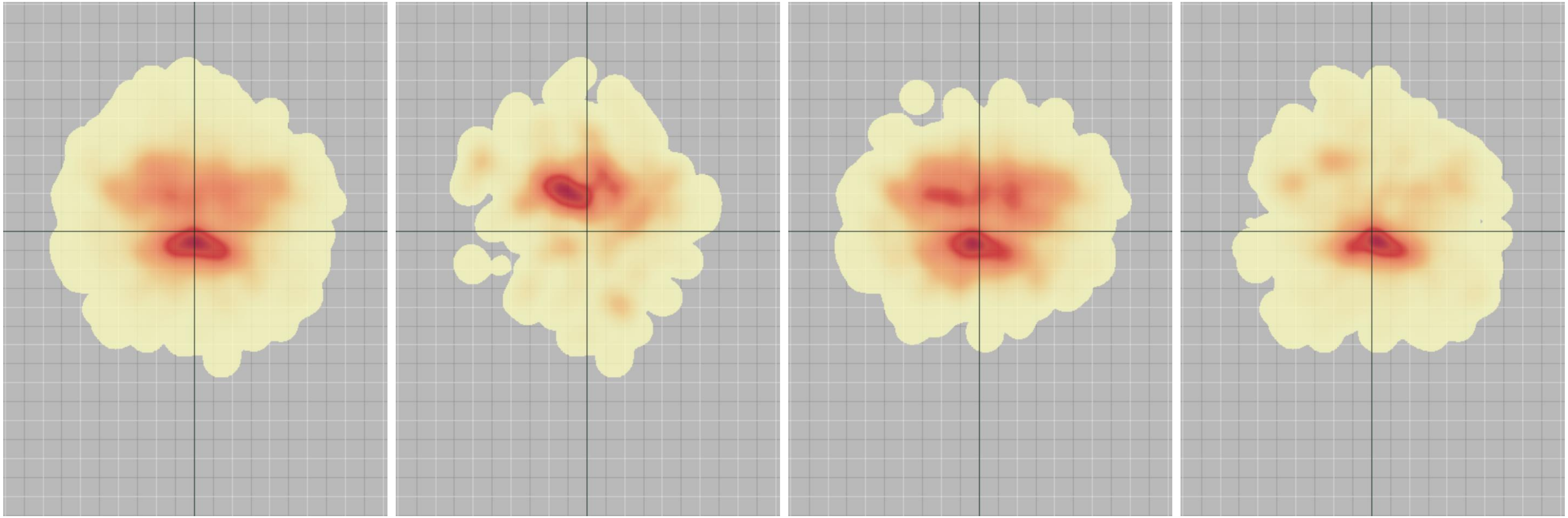


Ball in central third



Ball in opponents' (red) third

Distribution of the red team’s pressure in the yellow team’s space



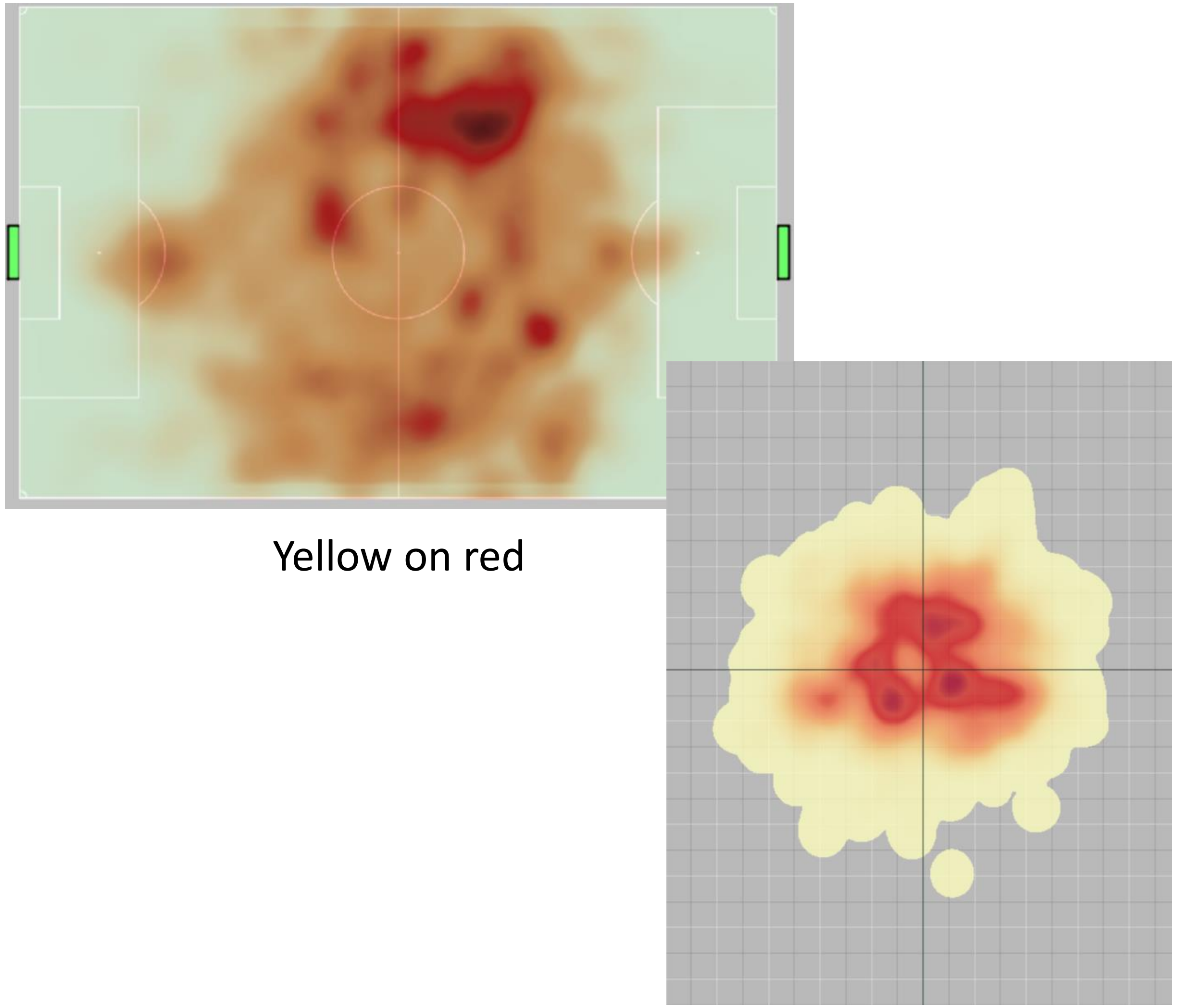
Overall

Ball in own (red) third

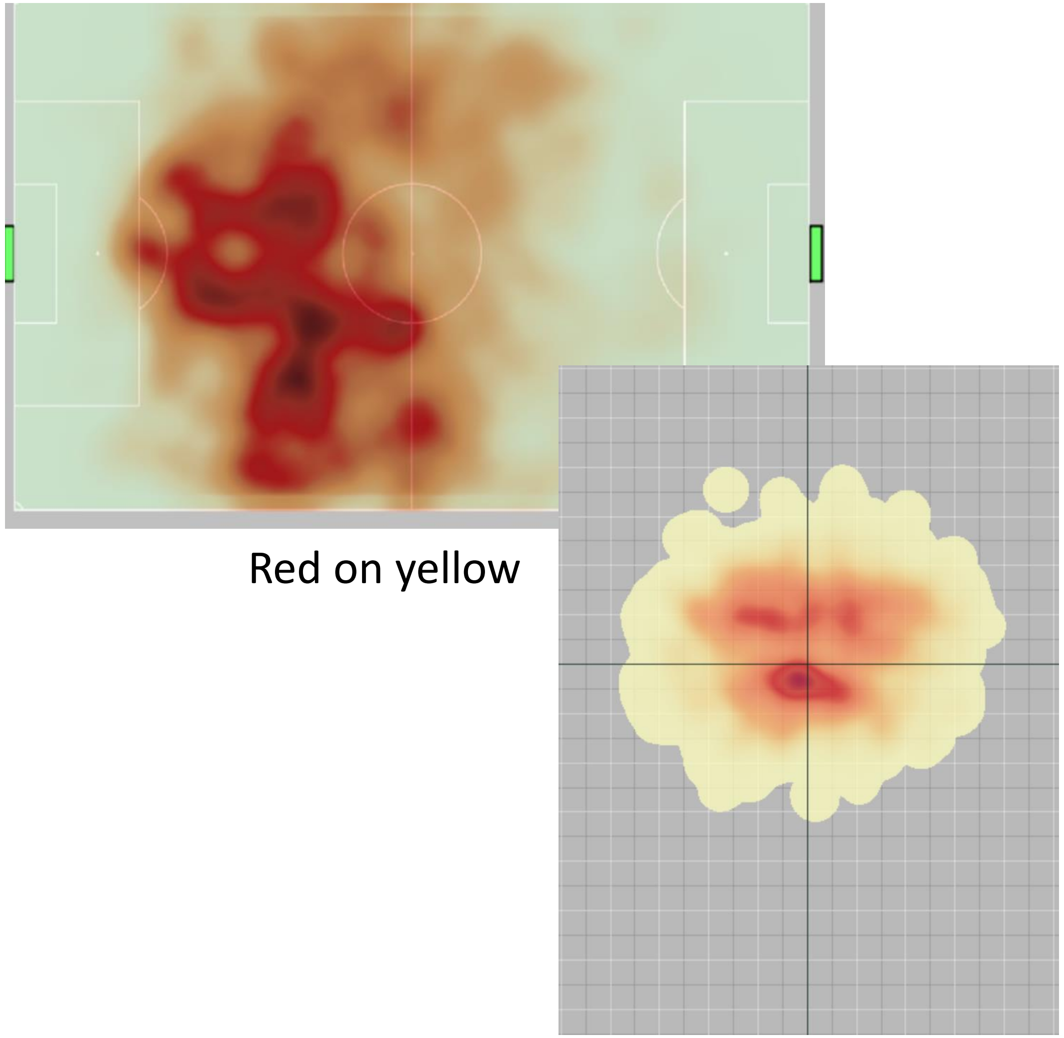
Ball in central third

Ball in opponents’ (yellow) third

Ball in central third: comparison of pressure distributions

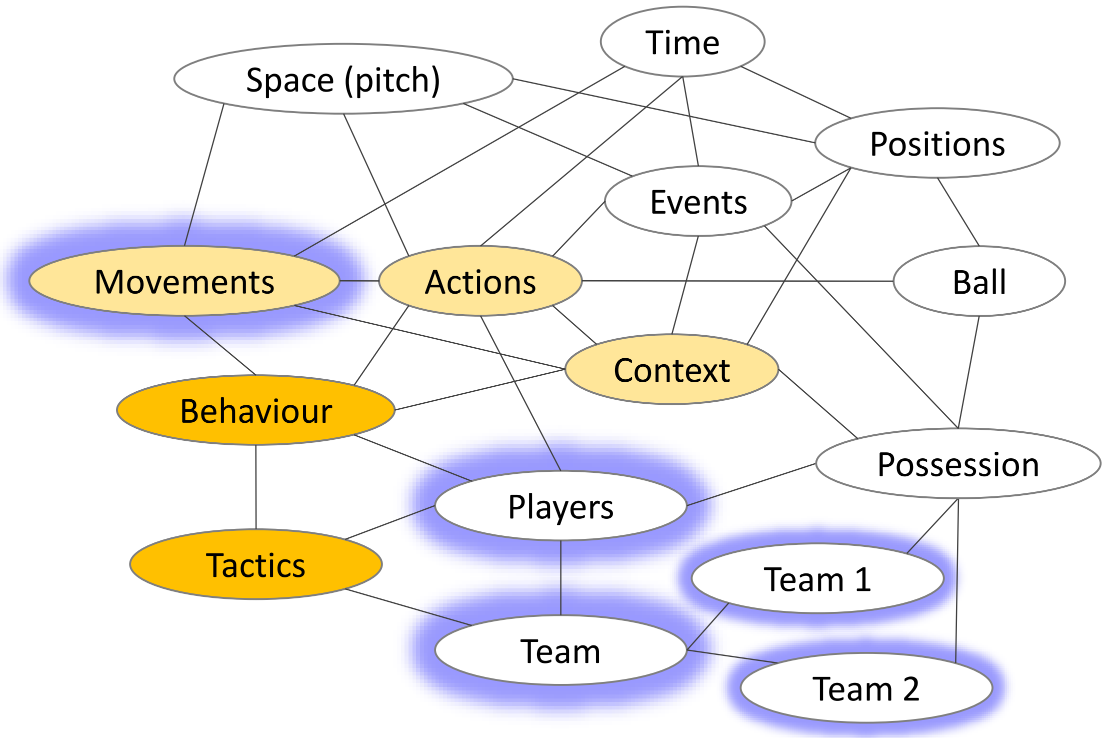


Yellow on red

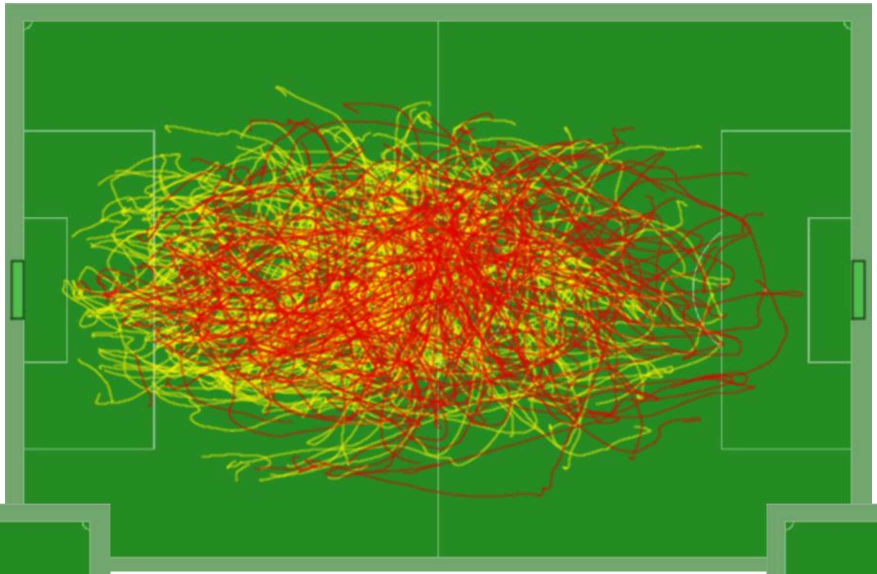


Red on yellow

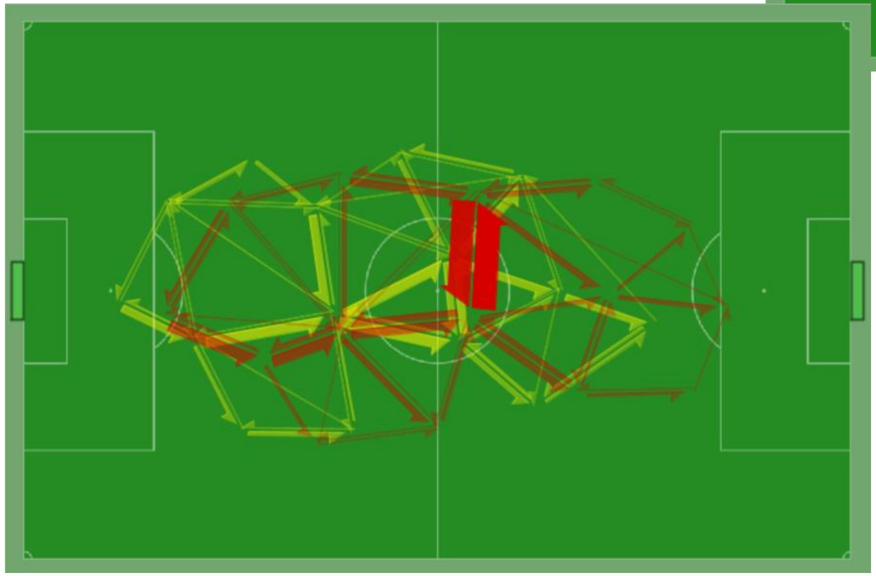
Movements of teams



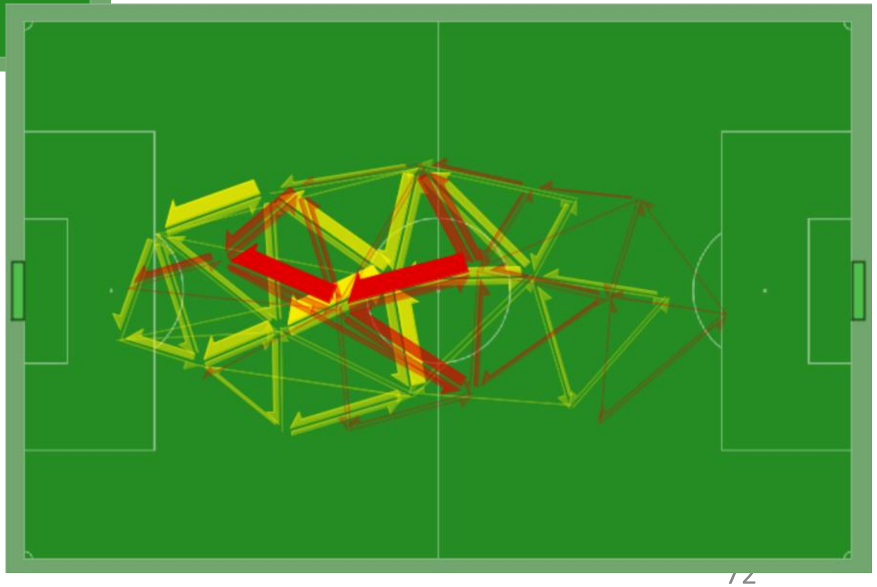
Movements of team centres



Under yellow team’s possession

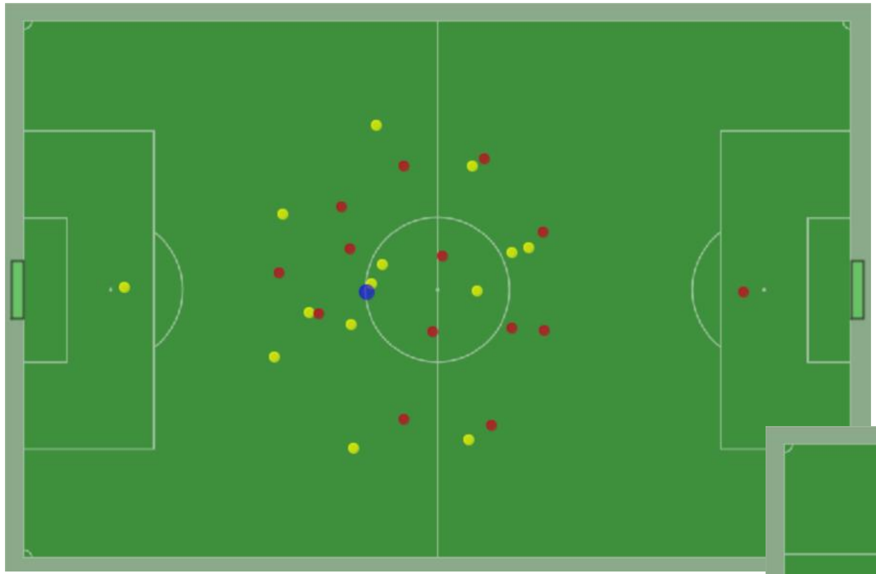


Under red team’s possession



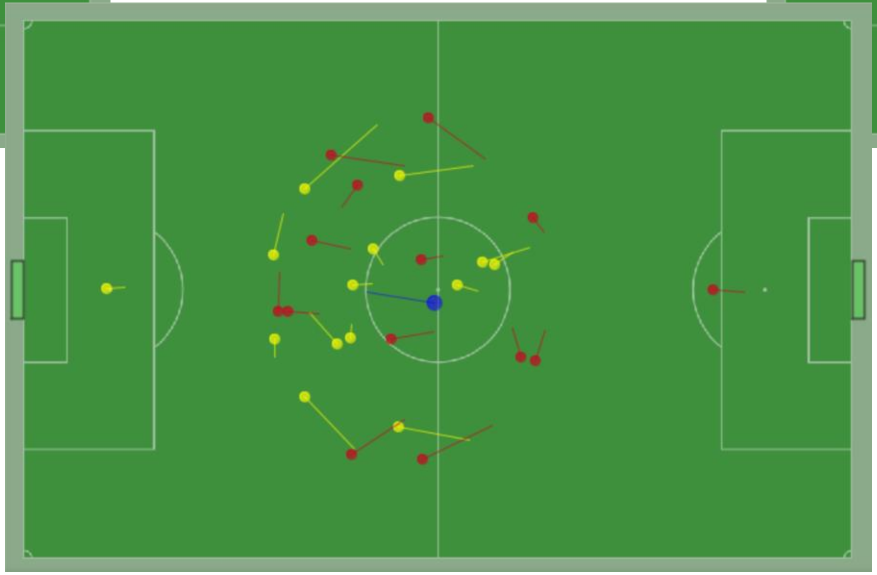
How to obtain a clear abstraction of collective movements?

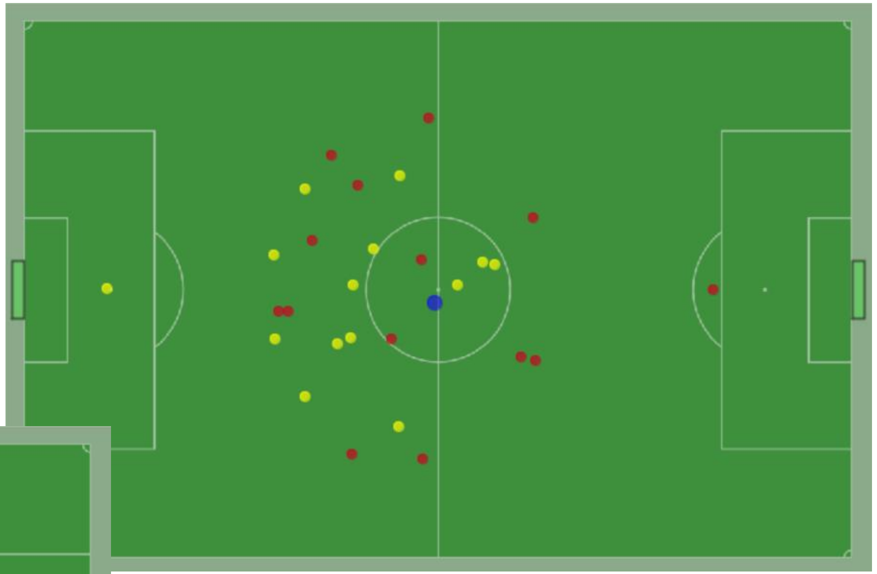
Let’s consider averaged positions of players under different conditions (contexts)



Ball possession by the yellow team

Changes

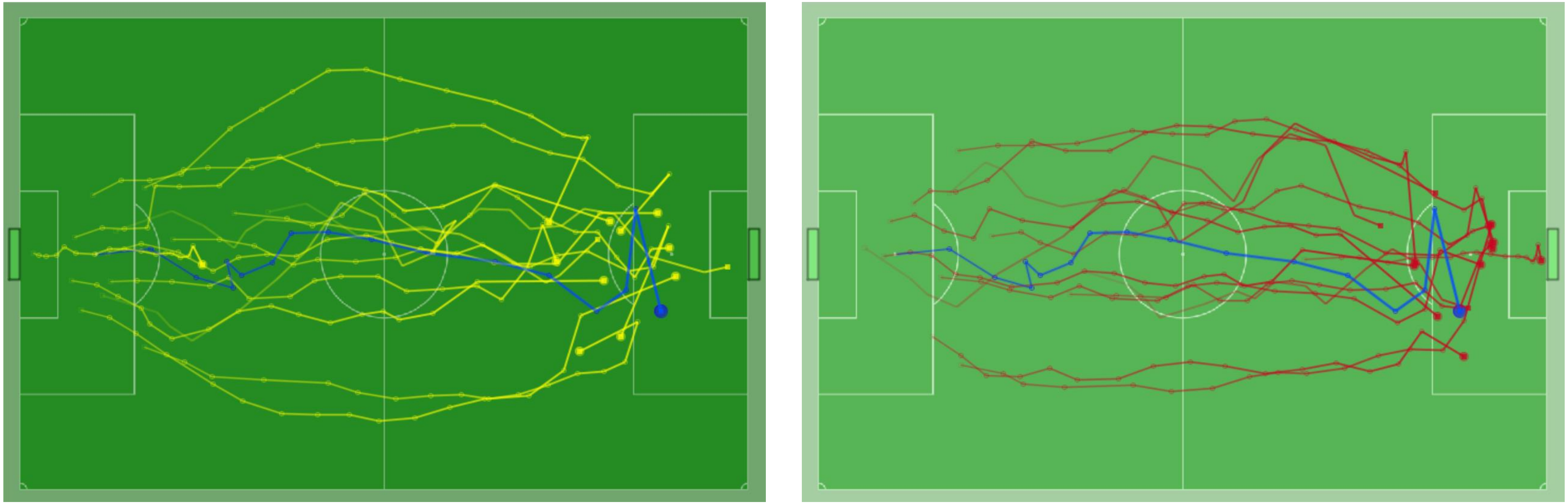




Ball possession by the red team

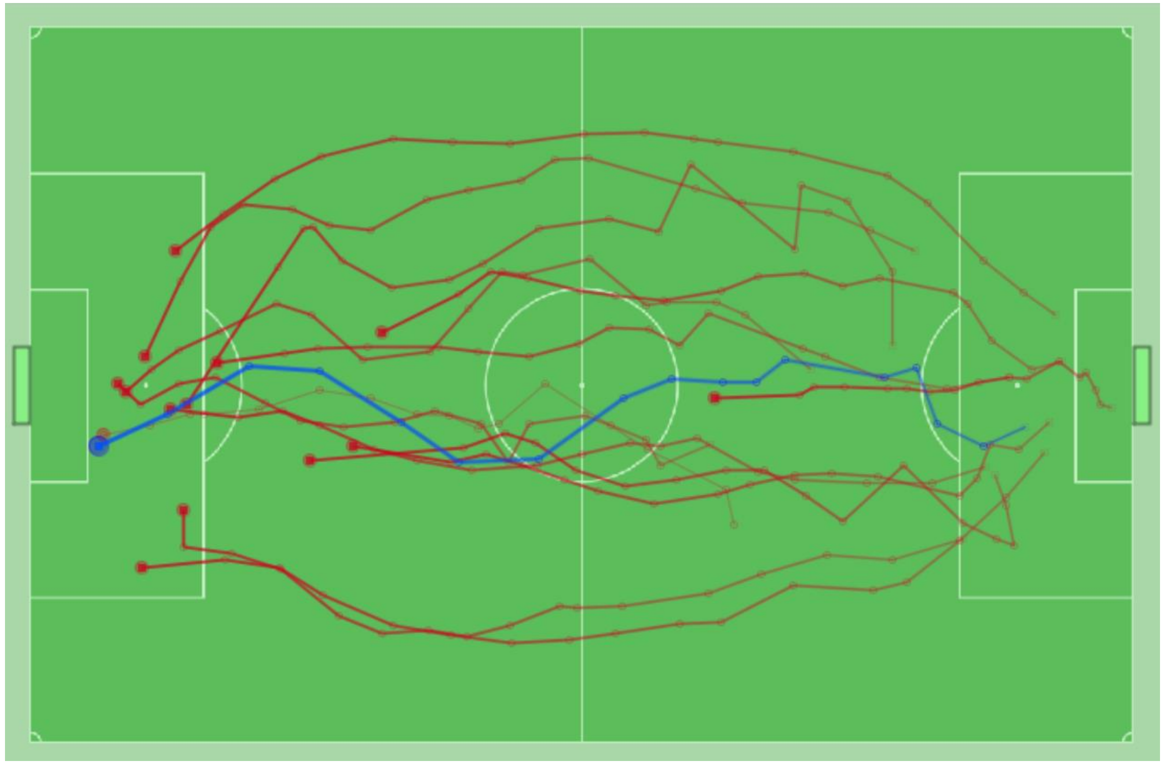
Constructing sequences of contexts to make abstractions

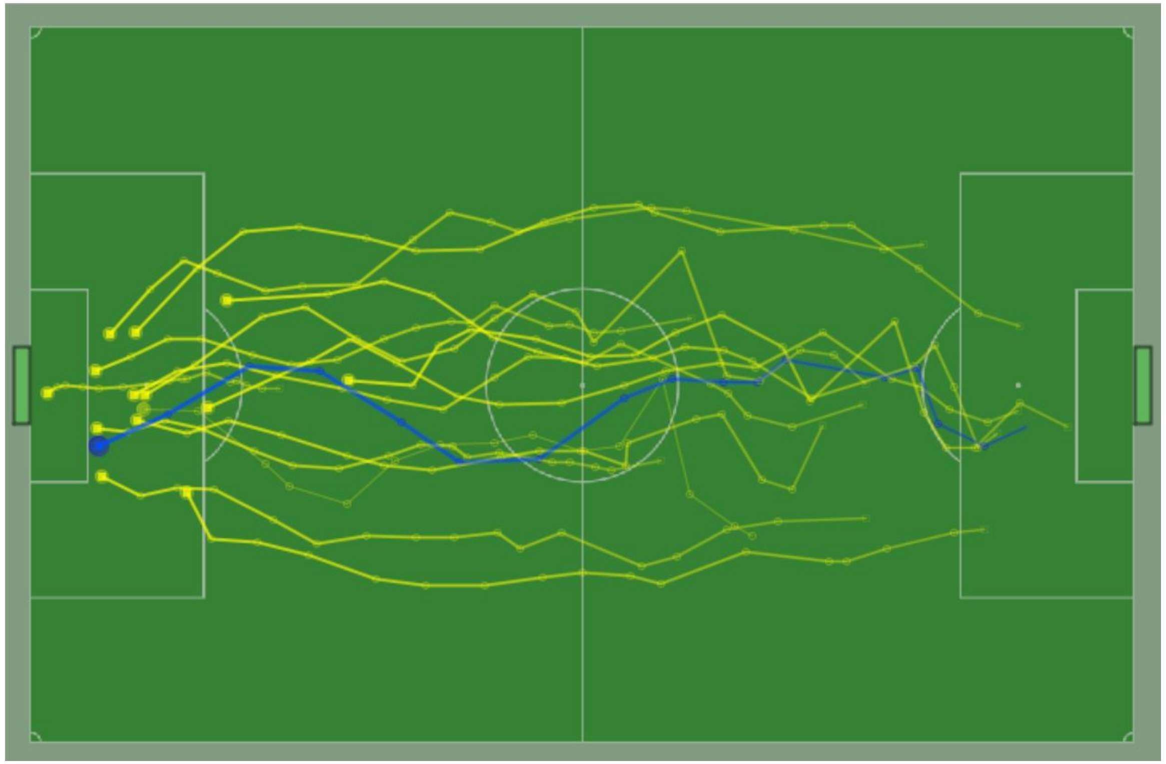
Changes of mean players' positions under the yellow team's possession depending on the position of the yellow team's centre, from left to right



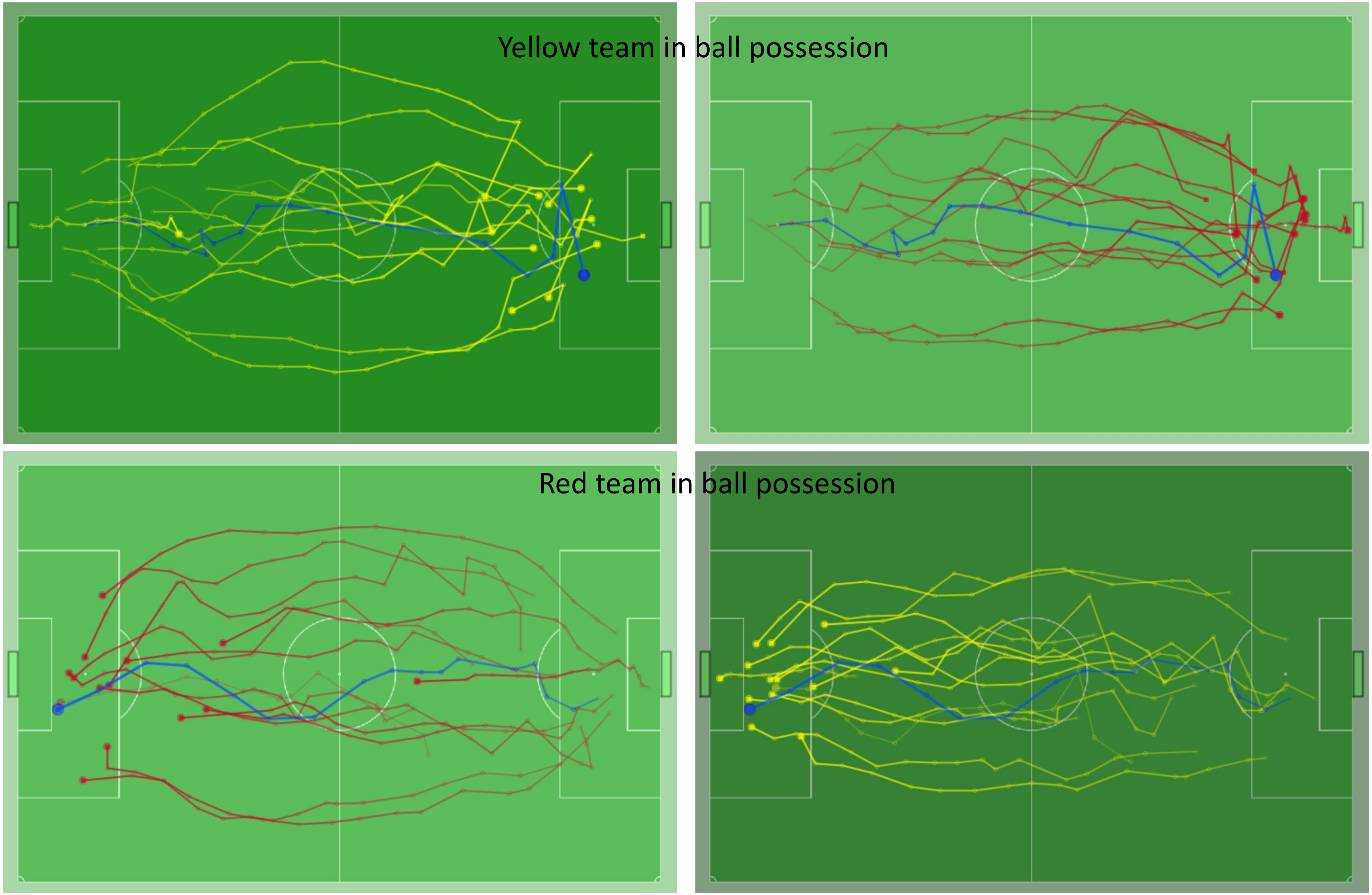
To facilitate abstraction, we create an artificial sequential arrangement of multiple contexts of interest. We exploit the ordering relationships between the contexts to link corresponding average positions into pseudo-trajectories, thus creating unified objects from multiple elements.

Changes of mean players' positions under the red team's possession depending on the position of the red team's centre, from right to left

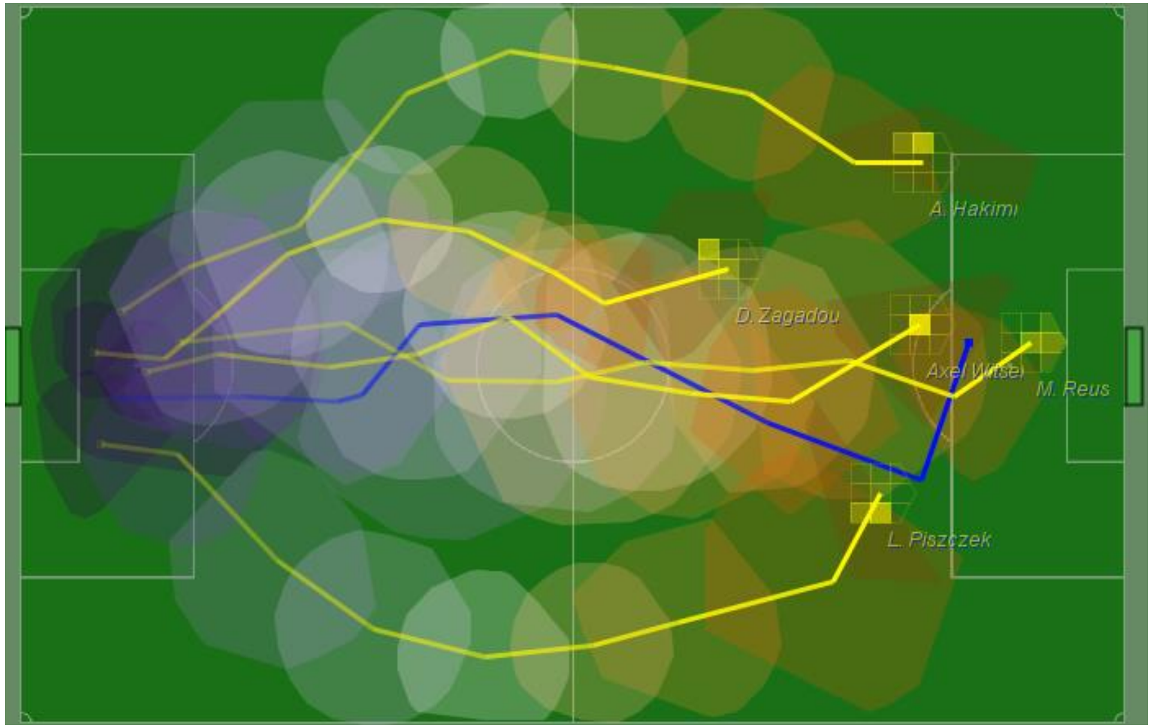


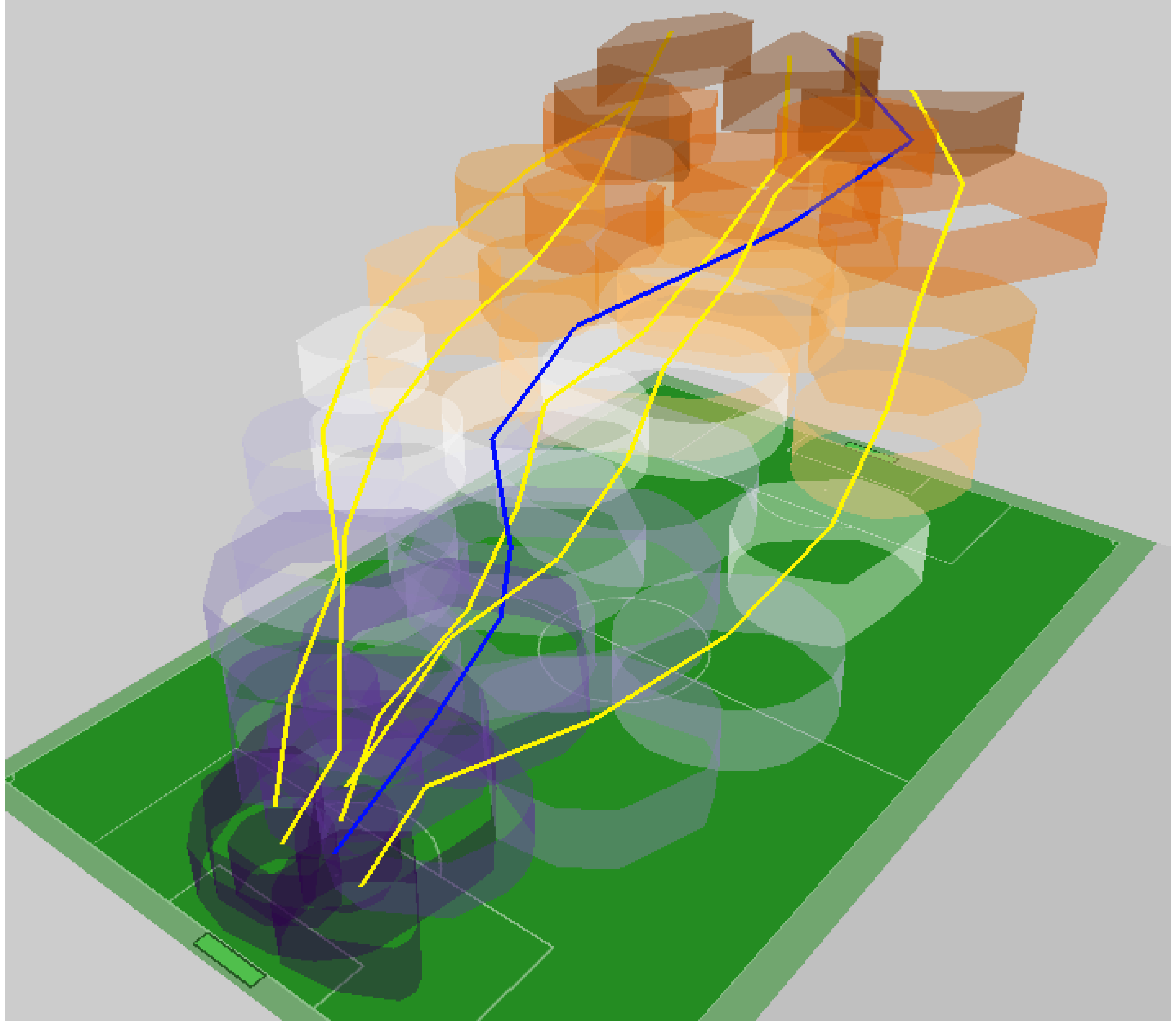


G. Andrienko, N. Andrienko, G. Anzer, P. Bauer, G. Budziak, G. Fuchs, D. Hecker, H. Weber, and S. Wrobel (2019)
Constructing Spaces and Times for Tactical Analysis in Football.
IEEE Transactions on Visualization and Computer Graphics, https://doi.org/10.1109/TVCG.2019.2952129.


Yellow team in ball possession
Red team in ball possession

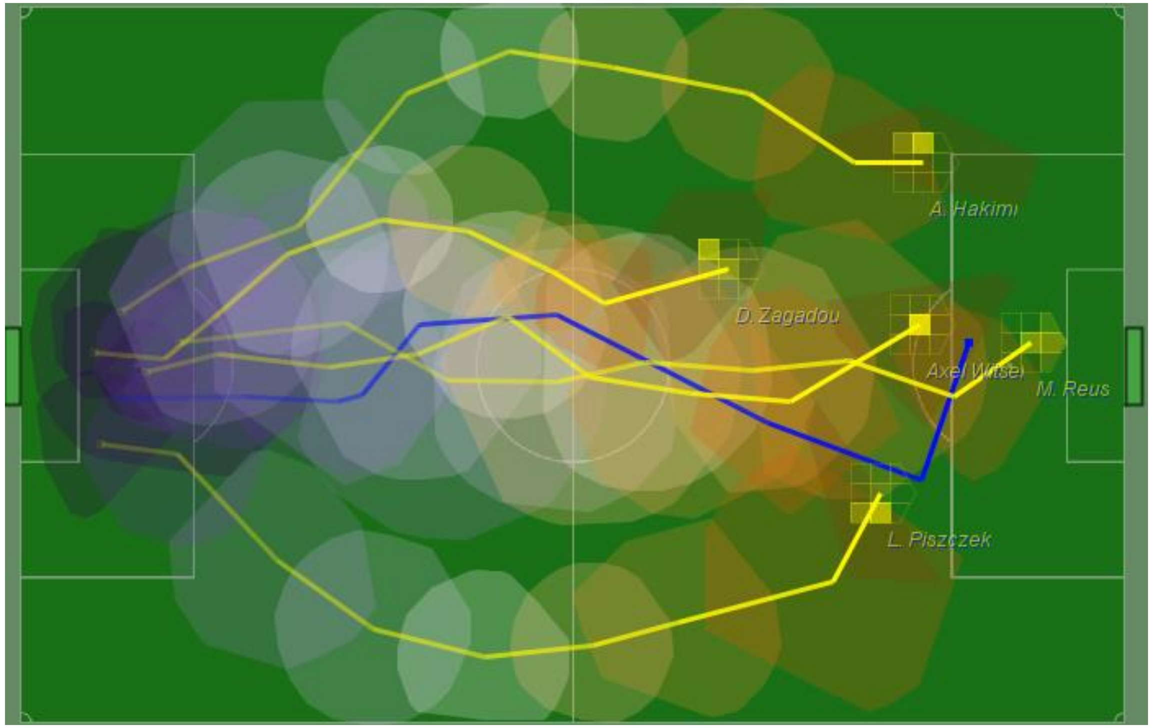
Assessing variation of positions: variability hulls

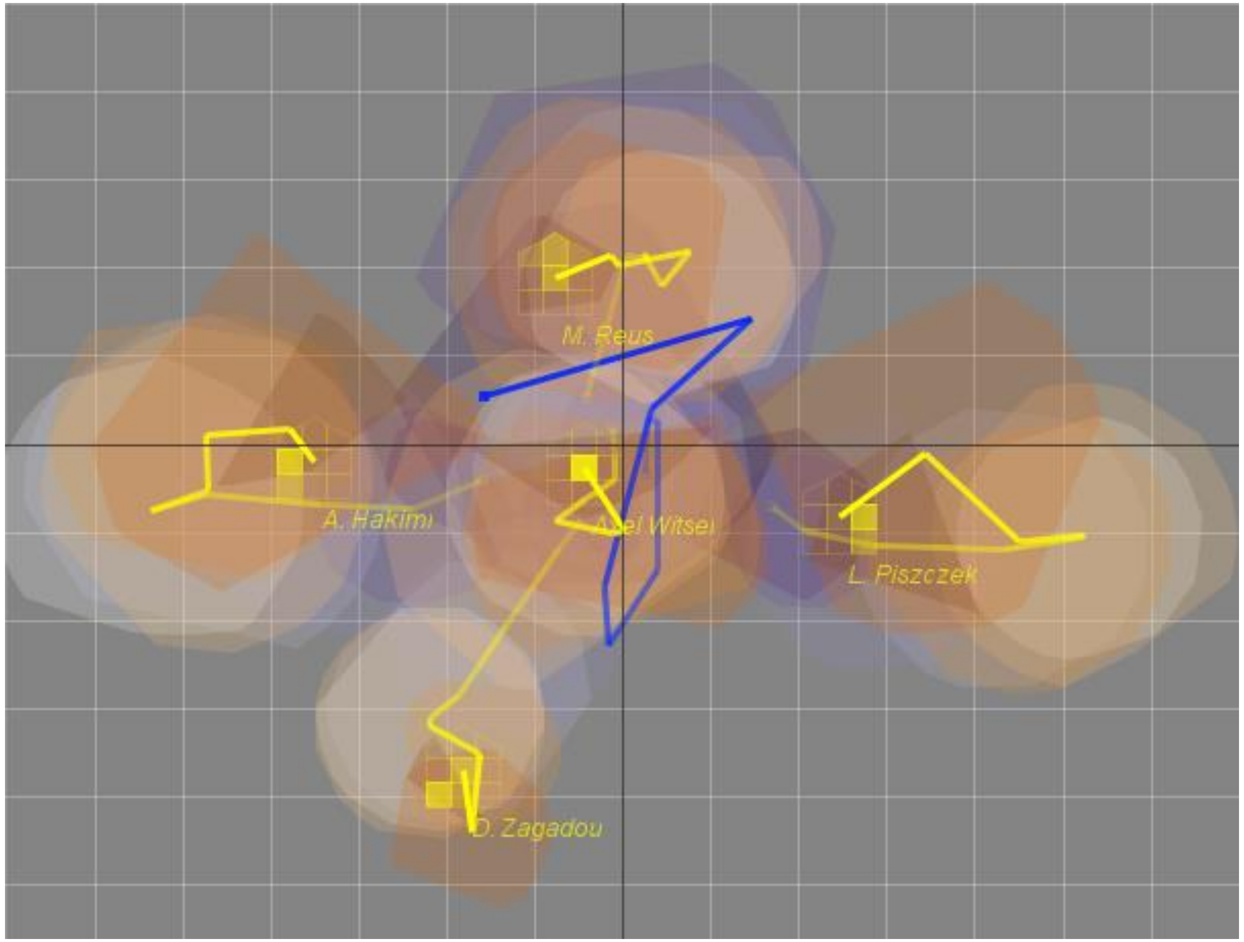




Convex hulls containing P% of players' positions around the average positions. Here P=50.

Assessing variation of positions: variability hulls

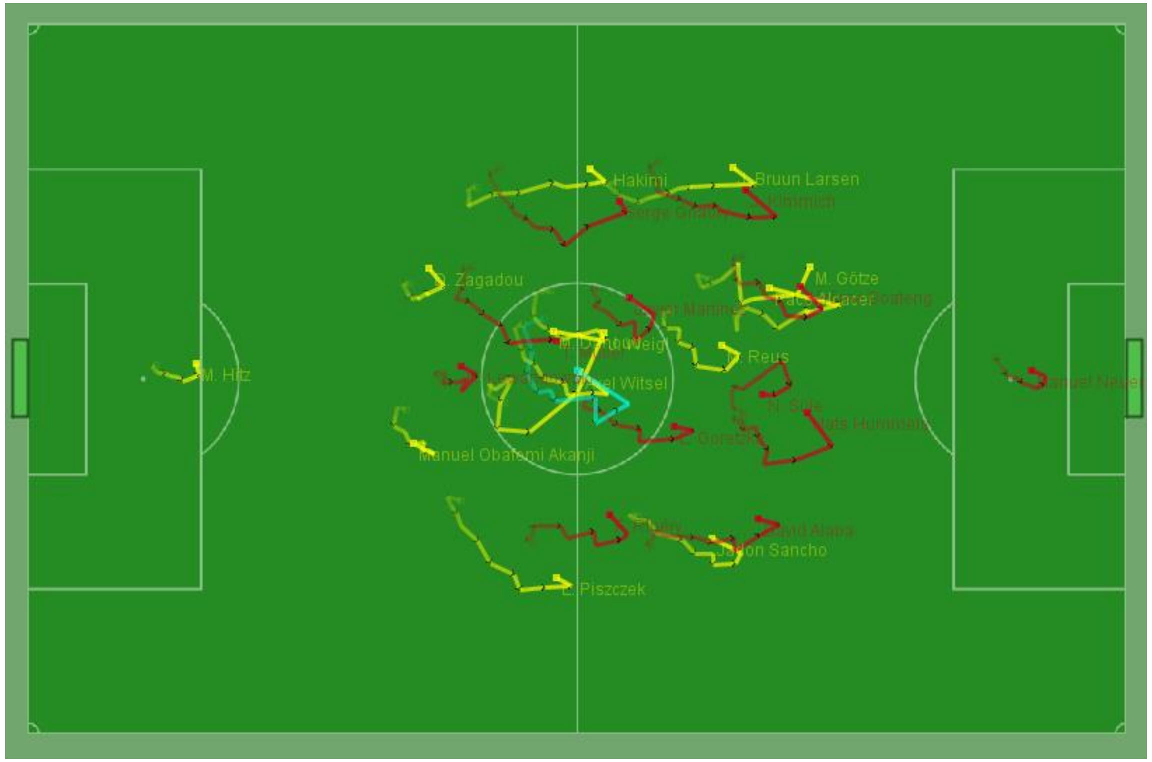




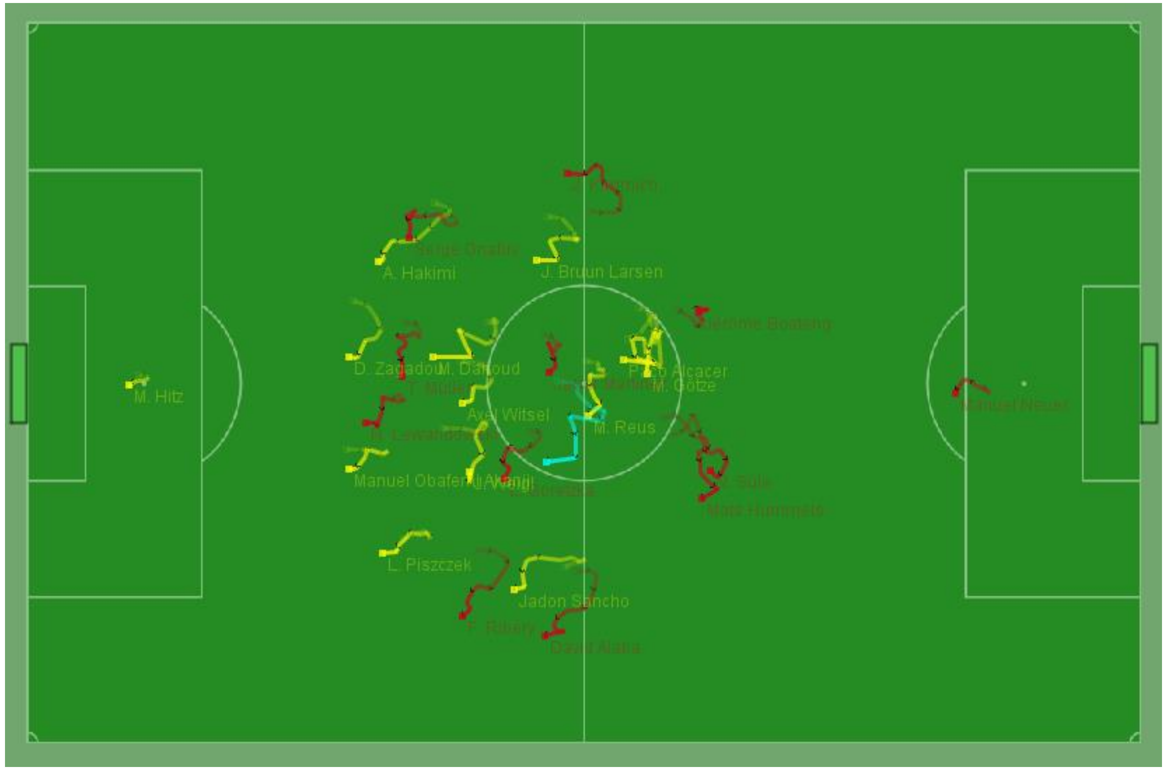
Pseudo-trajectories and variability hulls can be constructed both in the pitch space and in the team spaces.

Arrangement of contexts by relative time within episodes

Changes of the average players’ positions after changes of ball possession in the pitch centre.



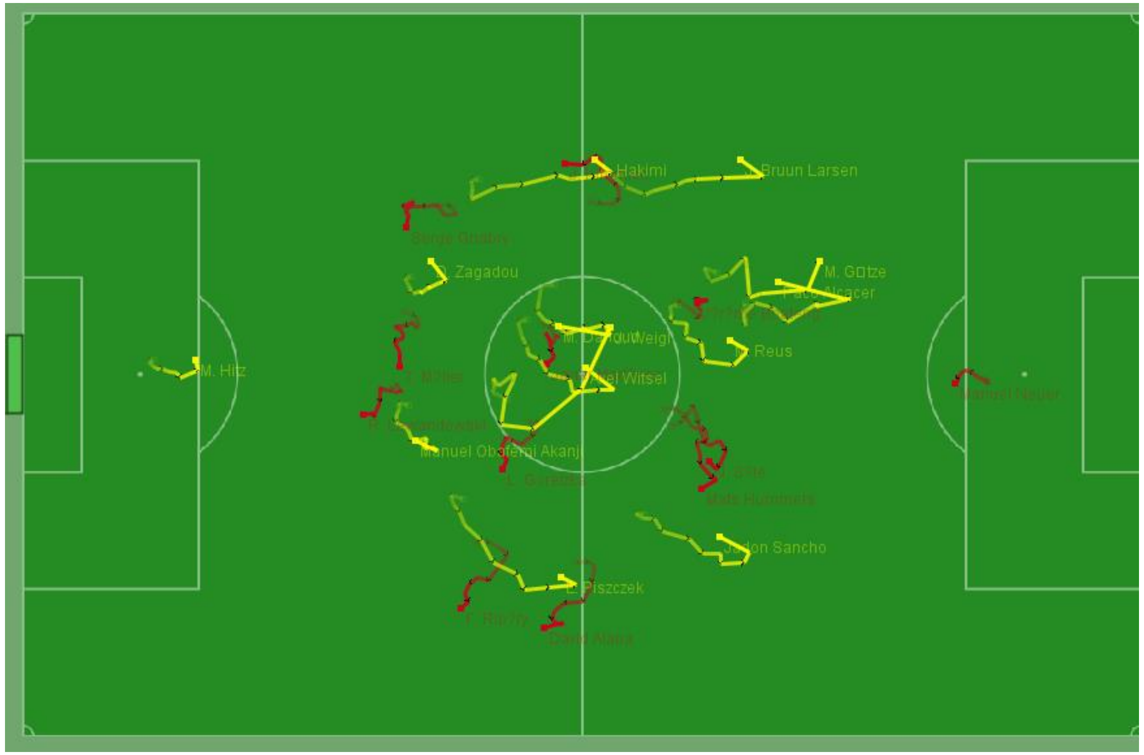
Ball gained by the yellow team



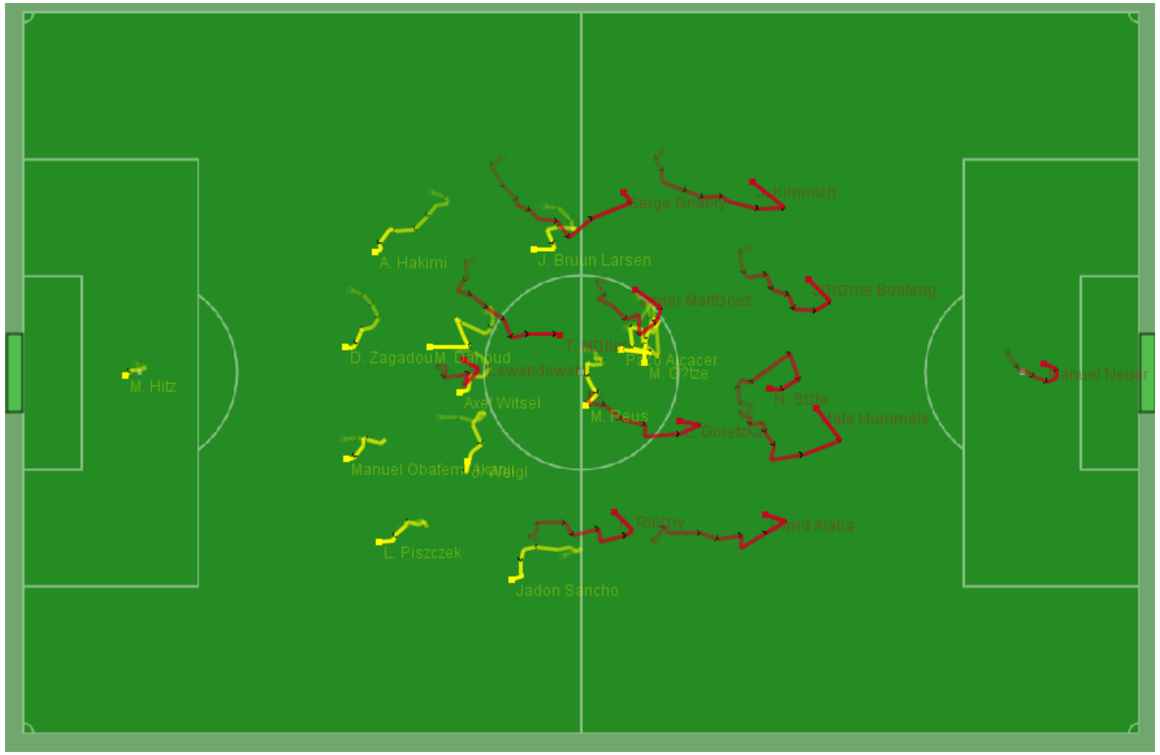
Ball gained by the red team

t_0: moment of possession change; +1sec ; +2sec; …; +12 sec

Comparison of teams’ behaviours



After ball gain

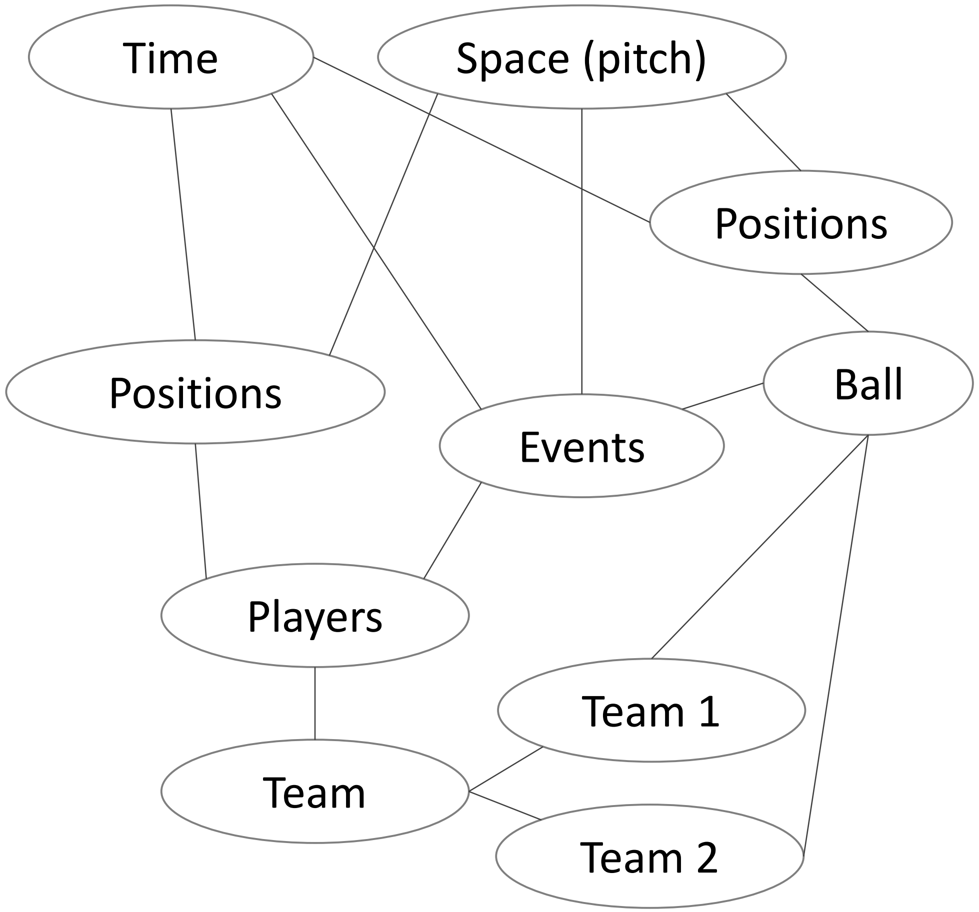


After ball loss

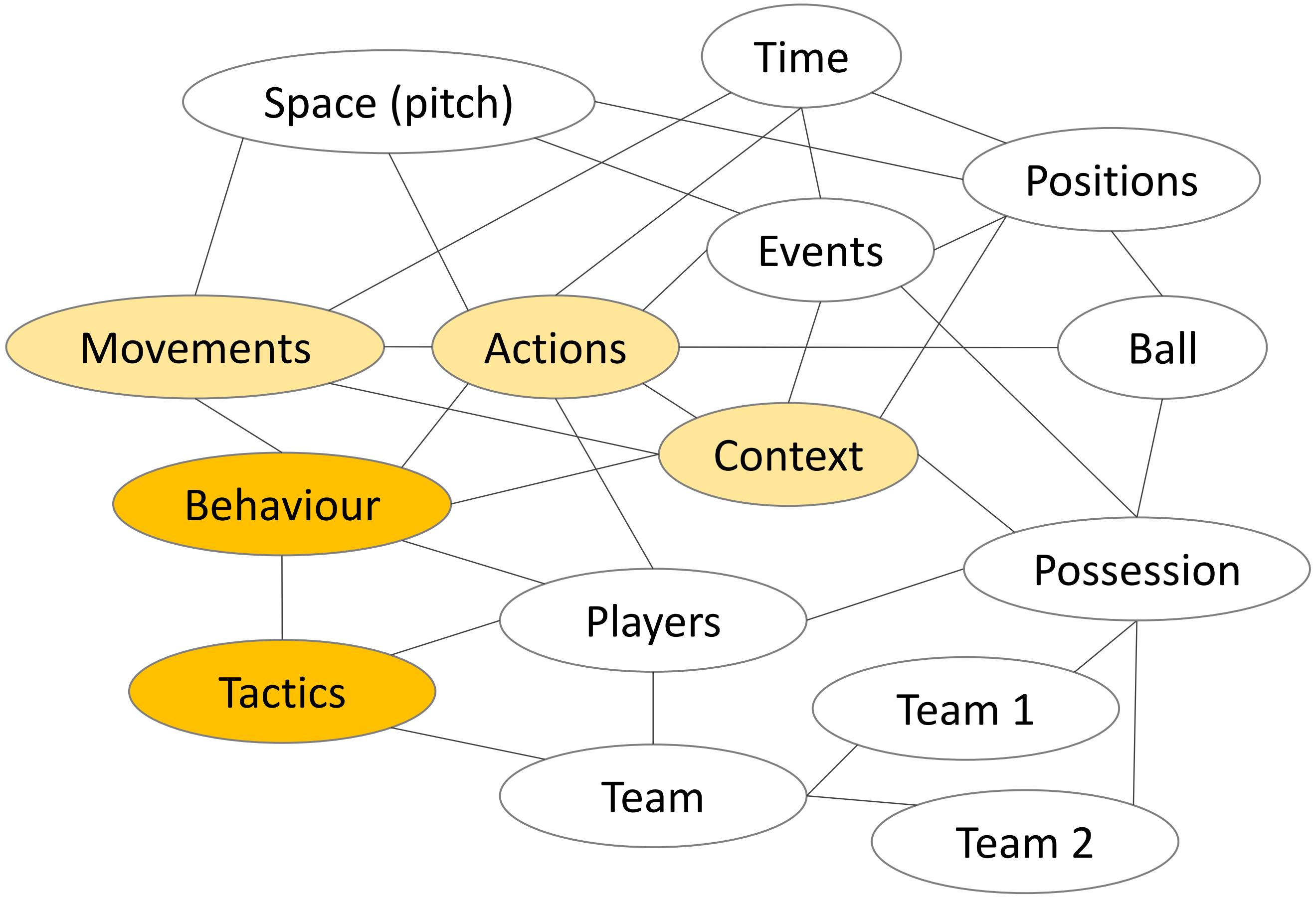
Pattern theory in football analysis

Consider relevant distributions, generate abstractions, observe patterns

Our analysis goals



Components of data



What we need to reconstruct and model*

* represent in a generalised manner

Our approach to football analysis*

- Determine which/what data components are relevant to our goals
  - Create missing data components from existing ones (e.g., define the set of relevant contexts, construct team spaces)
- Understand what distributions need to be considered
  - Relevant bases: pitch space, team spaces, team members, contexts, relative times of episodes (not absolute time as such).
  - Relevant overlays: individual movements, actions, team movements, team formations
- Understand what aspects of each distribution are relevant to analysis goals: composition, arrangement, or variation
  - Spatial arrangement of movements, spatial variation of movement amounts and properties, variation of movements, actions, and team formations over contexts, composition and variation of players' involvement in actions, …

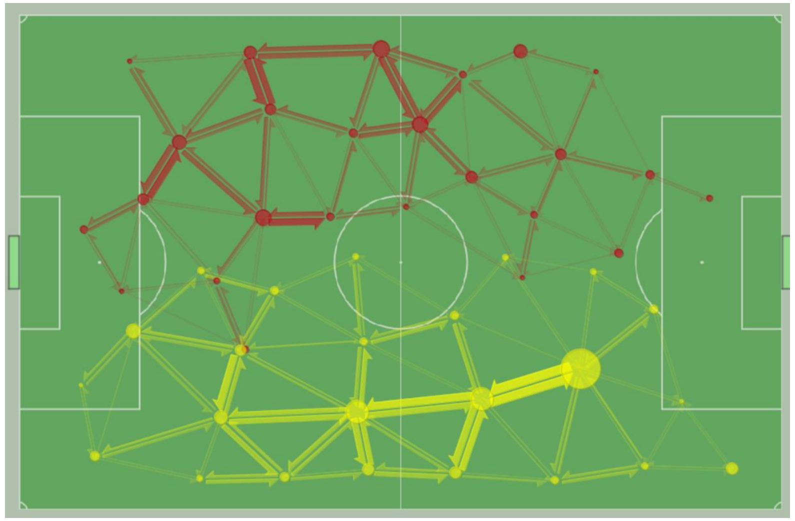
* All top-level list items apply to analysis of any data.

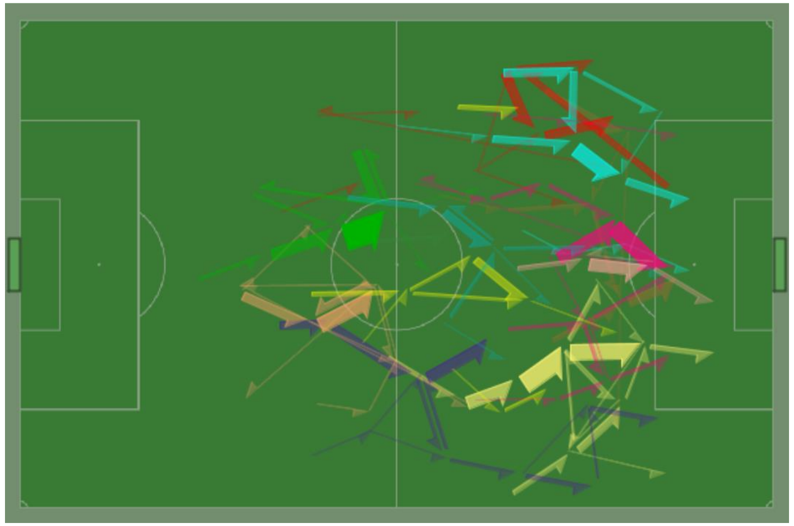
Our approach to football analysis* (continued)

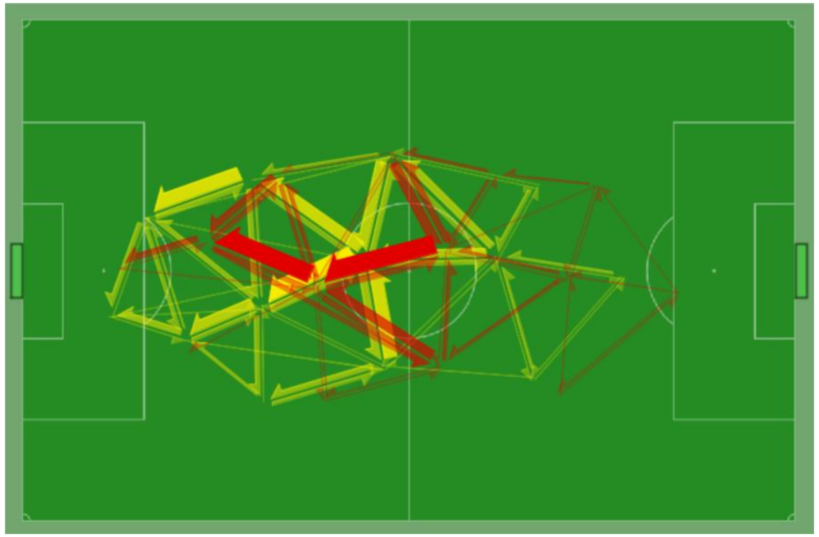
- Understand what relationships between base elements are exploitable for unification
  - Ordering: create sequences or networks; distances: create clusters; distances + directions: create shapes; distances + continuity: create fields; equivalence (same/similar or different): create groups
  - Define meaningful relationships when no suitable relationships exist (e.g., ordering relationships between contexts)
- Exploit the base relationships to arrange and aggregate data
  - Spatial aggregation: points → areas, fields; spatial aggregation + ordering: points → lines, areas → networks; equivalence: players → teams, contexts → context classes
  - Data aggregation is a technique supporting unification and abstraction
- Visualise the aggregates to observe and interpret patterns
- Compare patterns conveyed by different aggregates
- Discover relationships between patterns

* All top-level list items apply to analysis of any data.

Spatial distributions of movements

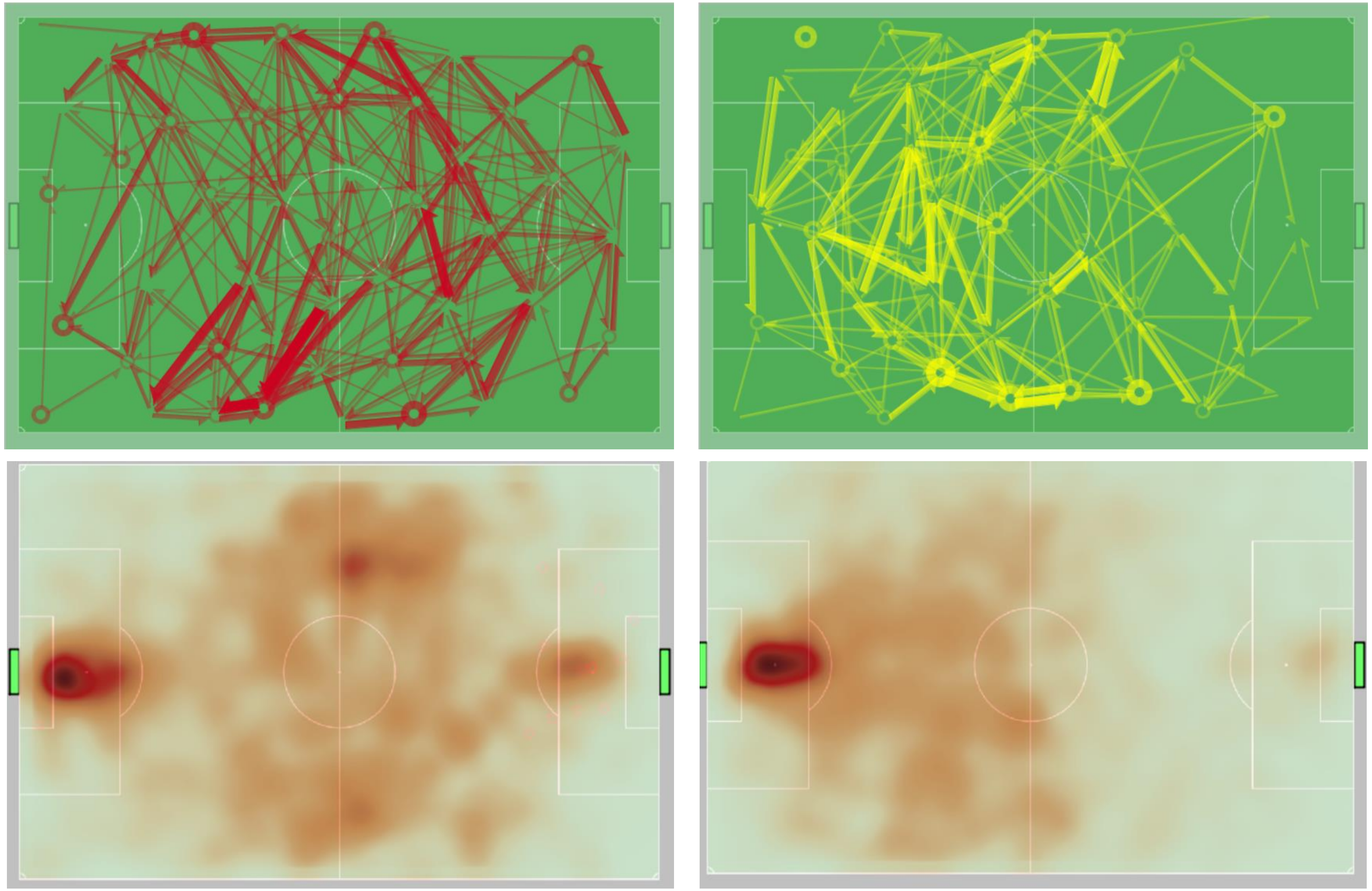






Temporal ordering relationships → arrangement of positions into sequences
Spatial distance relationships → arrangement of spatially close positions into spatial clusters, making areas
Sequential ordering relationships between positions → transitions between areas
----
Elementary data items (time stamped position records) → movement networks

Spatial distributions of actions



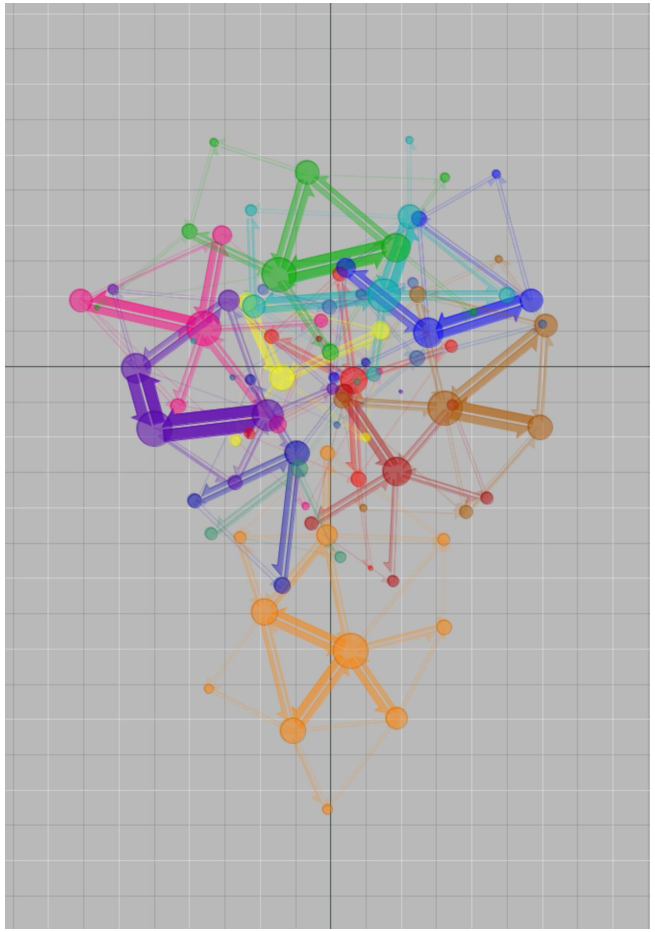
Passes (discrete actions): discrete spatial aggregation, as for movements

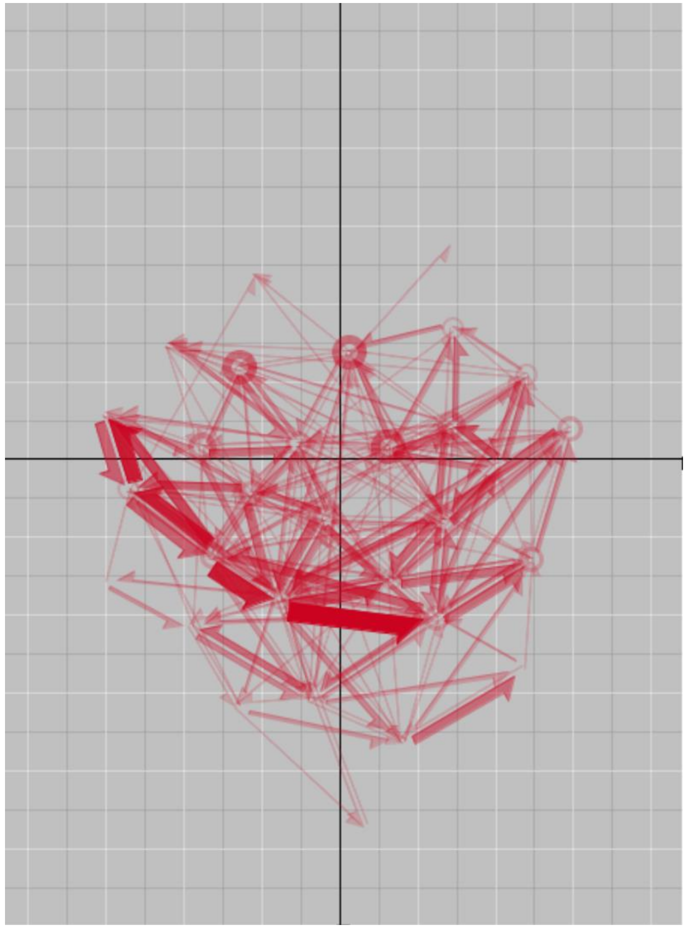
Pressure (continuous actions): continuous spatial aggregation and smoothing

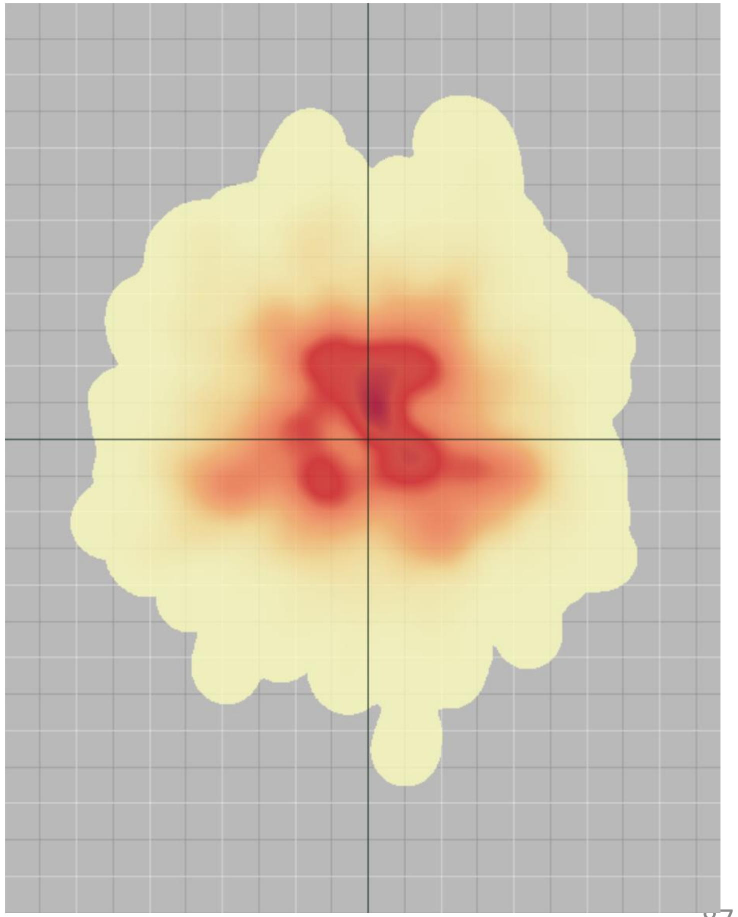
Exploited relationships: spatial distances and spatial continuity in the base, spatial autocorrelation in the overlay variation

Distributions in team spaces

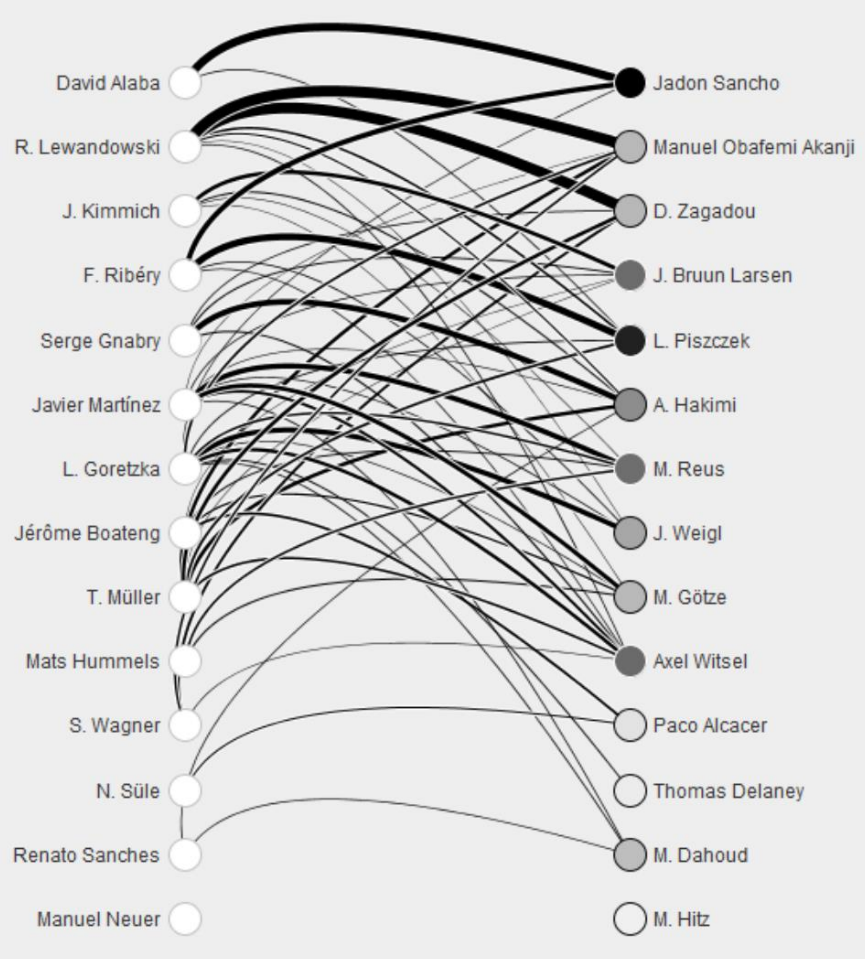
“Team space” is the relative spatial arrangement of team players. In considering team spaces we disregard absolute spatial positions on the pitch. Abstractions of movements and actions are constructed as in the pitch space.

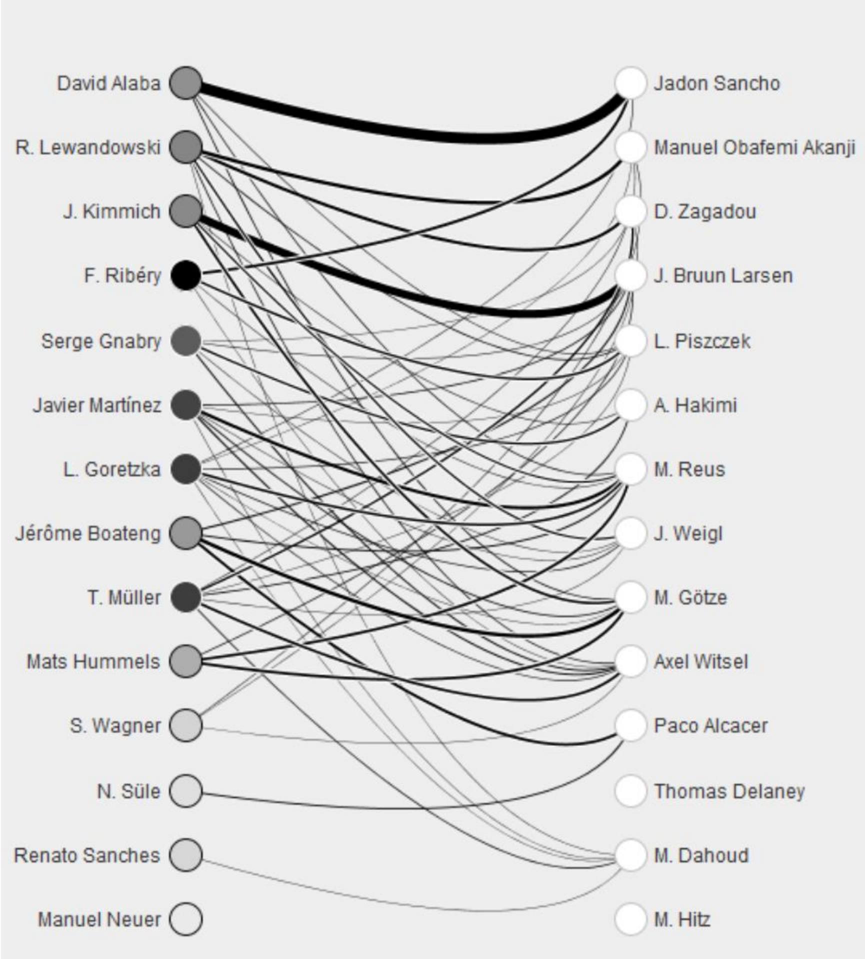






Distributions over players and pairs of players



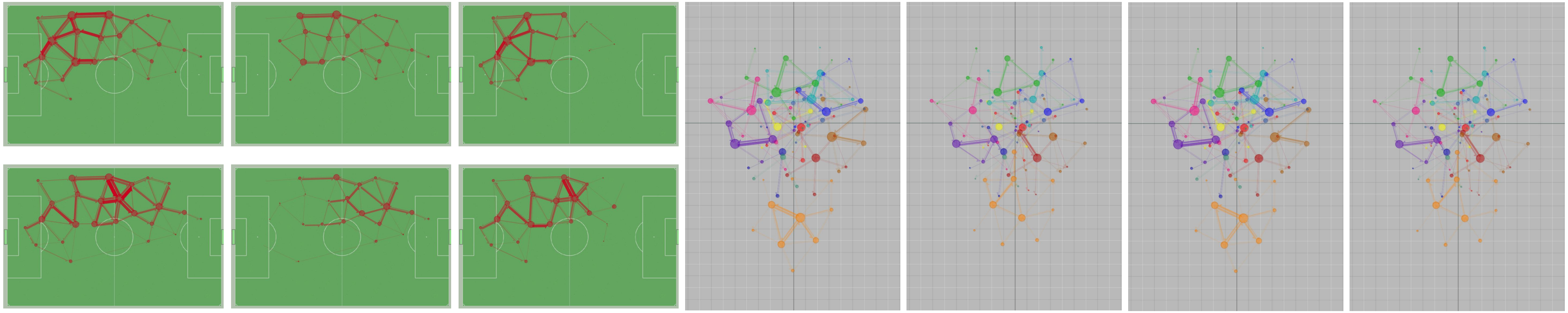


Composition of the players' involvement in the defensive pressure actions.

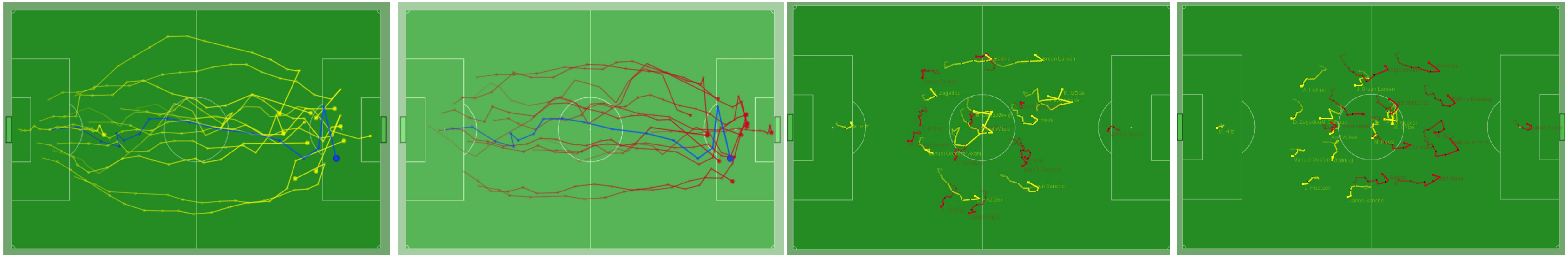
Variation of the amounts of the pressure over the pairs of players.

Distributions over contexts

Juxtaposition of data aggregates corresponding to different contexts



Exploitation of meaningfully defined ordering relationships between contexts

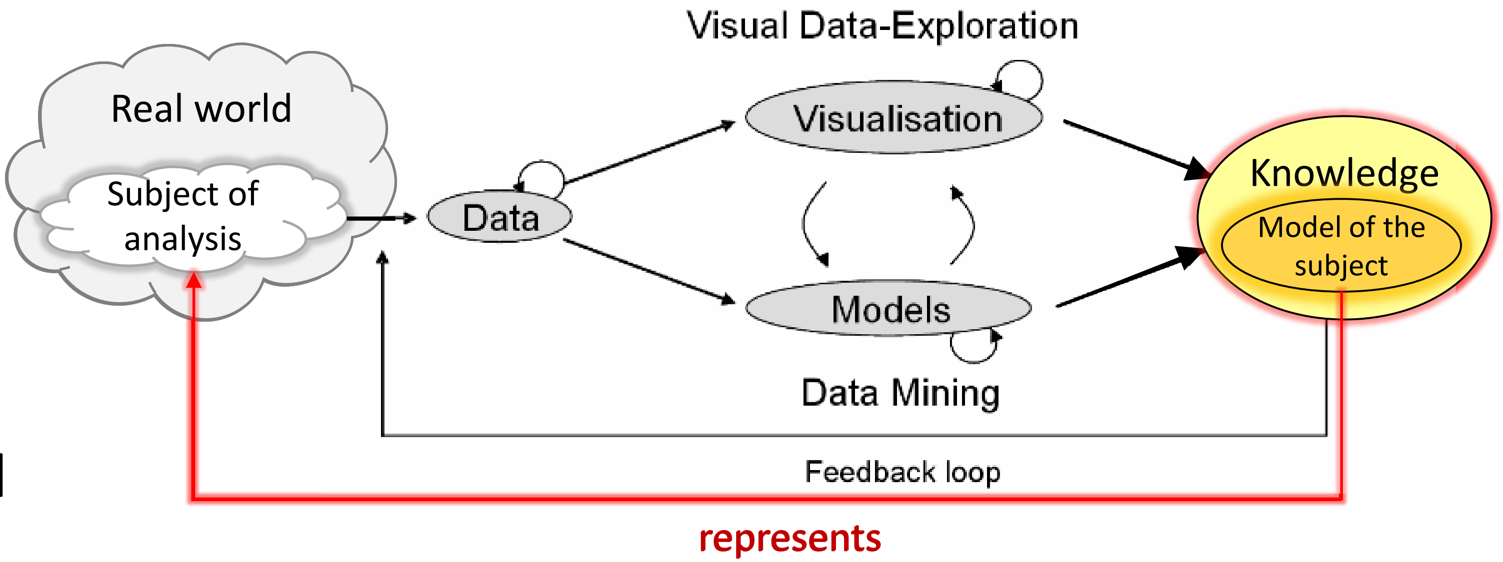


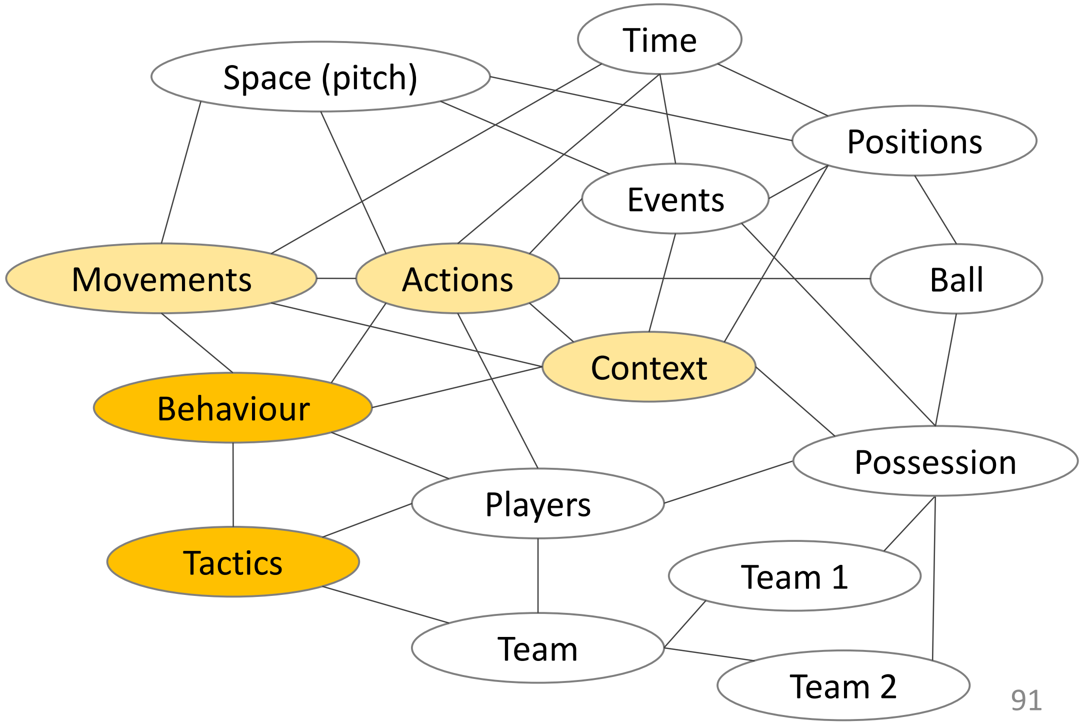
Types of patterns we discovered

- Major areas of presence and activities
- Major movement directions
- Major targets and sources of pressure
- Parallel, diverging, converging movements of players in teams
- Similarities and distinctions between data aggregates
- Impacts of context on movements and actions
- Relationships between movements and between actions of different teams

What is still missing

- Discovered patterns are not the final result of data analysis.
- A usual goal is to create an overall *model* of the analysis subject.
- Patterns provide material for the model that still needs to be built.
- Next operation after pattern discovery: model synthesis from abstracted data patterns.
  - Requires theoretical foundations.
  - Requires technical support (methods and tools).





Where to read more

N. Andrienko, G. Andrienko, S. Miksch, H. Schumann, and S. Wrobel (2020)
A theoretical model for pattern discovery in visual analytics.
Visual Informatics. https://doi.org/10.1016/j.visinf.2020.12.002

Andrienko, N., Lammarsch, T., Andrienko, G., Fuchs, G., Keim, D., Miksch, S., and Rind, A. (2018)
Viewing Visual Analytics as Model Building.
Computer Graphics Forum, 37: 275-299. https://doi.org/10.1111/cgf.13324

Andrienko, N., Andrienko, G., Fuchs, G., Slingsby, A., Turkay, C., and Wrobel, S. (2020)
Visual Analytics for Data Scientists.
Springer https://doi.org/10.1007/978-3-030-56146-8

Andrienko, G., Andrienko, N., Bak, P., Keim, D., and Wrobel, S. (2013)
Visual Analytics of Movement.
Springer https://doi.org/10.1007/978-3-642-37583-5

G. Andrienko, N. Andrienko, G. Budziak, J. Dykes, G. Fuchs, T. von Landesberger, and H. Weber (2017)
Visual Analysis of Pressure in Football.
Data Mining and Knowledge Discovery, **31**(6): 1793-1839, http://dx.doi.org/10.1007/s10618-017-0513-2

G. Andrienko, N. Andrienko, G. Anzer, P. Bauer, G. Budziak, G. Fuchs, D. Hecker, H. Weber, and S. Wrobel (2019)
Constructing Spaces and Times for Tactical Analysis in Football.
IEEE Transactions on Visualization and Computer Graphics, https://doi.org/10.1109/TVCG.2019.2952129.