Life and Motion of Collective Phenomena

Antony Galton

School of Engineering, Mathematics and Physical Science
University of Exeter, UK

Workshop on Collectives in Space and Time

Rostock, Germany, June 2011
TIME . . .

. . . SPACE . . .

. . . COLLECTIVES

The Logic of Aspect: An Axiomatic Approach (Oxford University Press, 1984)


. . . translated into German by Bertram Kienzle as:

‘Die Logik des Vorkommens’ (in Zustand und Ereignis, Suhrkamp Verlag, 1994)

The RCC system of qualitative relations between spatial regions.

RCC = Regional Connection Calculus = Randell, Cui, and Cohn


Geographical-scale phenomena which, depending on the point of view from which they are being described, might be modelled as belonging to ontologically distinct categories.

A protest march: is it
- an *event* (as seen by an onlooker)?
- a *process* (as seen by a participant)?
- an *object* (as seen by a police surveillance helicopter)?

Other examples: storms, floods, wildfires, traffic jams.

A key desideratum for a spatio-temporal geo-ontology is to provide the technical means to represent these phenomena in such a way that their different aspects can be accessed equally easily.
“In many cases these [multi-aspect] phenomena involve large numbers of similar units acting together in a more or less coordinated way; but the phenomenon does not consist of those units since the units can have lives separate from the phenomenon and the phenomenon may outlast the participation of any individual unit. Thus the unity of the phenomenon as a whole goes beyond the separate unities of its constituent parts.”

This is the point at which I first began to think seriously about collective phenomena . . .
Collectives may be *dynamic* in two senses

- they exhibit movement
- they undergo change of membership

This paper presented a formal analysis within which dynamic collectives could be defined in a way that did justice to their “dual-aspect” nature.

**WARNING:** Technical stuff ahead! (But not for long . . .)
Definition 1

The **lifeline** of a continuant entity $c$ is the set of spatio-temporal positions occupied by $c$ in the course of its existence:

\[
\text{lifeline}(c) = \{ \langle s, t \rangle \in S \times T \mid s \in \text{pos}(c, t) \}\]

Definition 2

An **episode** is a connected subset of the lifeline of a continuant:

\[
\text{epi}(c, t_1, t_2) = \{ \langle s, t \rangle \in \text{lifeline}(c) \mid t_1 \leq t \leq t_2 \}\]
**Definition 3**

A **collective dynamic** is a collection $C$ of episodes from two or more individual lifelines, closed under the sub-episode relation.

**NOTE:** A collective dynamic represents the event-like aspect of a dual-aspect collective phenomenon.

**Definition 4**

A **dynamic collective** is that “notional” continuant $C^*$ whose lifeline is the aggregation of the episodes in the collective dynamic $C$, i.e., such that

$$\text{lifeline}(C^*) = \bigcup C.$$
Properties of Dynamic Collectives

From the definition of a dynamic collective one can go on to define formally such notions as

- the **participants** in a dynamic collective;
- the **members** of a dynamic collective at a time;
- the **lifetime** of a dynamic collective;
- the **participation** of an individual member in a dynamic collective.

(Details in the COSIT paper!)

But the question which interested me most was: *How can we specify the position of a collective?* — in other words, what value should be assigned to

\[ \text{pos}(C^*, t)? \]
“To represent the spatial region occupied by the points, we could simply take the set of points themselves; but we may want something less detailed, some simply-specified region sufficient to indicate the area over which the points are distributed, or their broad-brush configuration.”

The obvious choice, mathematically, is the convex hull of the points. But this is often unsatisfactory, assigning the same position to very different-looking collectives . . .
We noted that

- There is no such thing as *the* region occupied by a set of points;
- There are many different conditions that a user may or may not require a region for a set of points to satisfy, e.g.
  - Must the region include all of the points?
  - If so, must they all lie in the region’s interior, or can some (or all) of them lie on its boundary?
  - Should the region consist of a single connected component?
  - Must it be topologically regular?
  - &c, &c, ...

We described various algorithms for generating “footprints” for a set of points, including three of our own, and compared them with respect to these conditions.
“A typical paper in this area will propose an algorithm for generating a shape from a pattern of dots, explore its mathematical and/or computational characteristics . . . , and examine its behaviour when applied to various dot patterns. The evaluation of this behaviour is typically very informal, often amounting to little more than observing that the shape produced by the algorithm is a ‘good approximation’ to the perceived shape of the dots. While lip-service is generally paid to the fact that there is no objective definition of such a ‘perceived shape’, little is said about how to verify this, or indeed, exactly what it means.”
Which of the innumerable possible shapes ("footprints") associated with a given dot pattern might be considered to be in some sense the "best" ones?

To bring the problem down to a manageable size, we confine our attention to a restricted class of footprint:

**Definition**

A **polygonal hull** for a collection of dots in the plane is a polygon such that

1. Every vertex of the polygon is one of the dots;
2. Every dot which is not a vertex of the polygon lies in the interior of the polygon;
3. The boundary of the polygon forms a Jordan curve.
A DOT PATTERN
THE CONVEX HULL
ANOTHER POLYGONAL HULL
A THIRD POLYGONAL HULL
Which is the best polygonal hull for this dot pattern?

- Hull 1 includes too much empty area — it doesn’t correspond well to the perceived shape of the dot pattern.
- Hull 2 is too “spiky” — its perimeter goes in and out in a way that doesn’t correspond to anything we see in the dot pattern.
- Hull 3 seems to fit better with our intuitive perception. It achieves a compromise between two conflicting goals: reducing the area and reducing the perimeter.
A conjecture

The polygonal hulls which human observers would regard as representing intuitively “good” footprints would be found to be optimal with regard to reconciling the two conflicting objectives of reducing the area and reducing the perimeter.
Multi-objective optimisation

A polygonal hull $H_1$ **dominates** another polygonal hull $H_2$ so long as the area and perimeter of $H_1$ are no greater than those of $H_2$, and at least one of them is smaller. In this case, according to the conjecture, $H_1$ should be preferred to $H_2$.

Those polygonal hulls which are not dominated by any other hulls are said to be **Pareto optimal**. If plotted on a graph of perimeter against area, they lie along a line called the **Pareto front**.
Dominance: $H_1$ dominates $H_2$, but neither of $H_3$ and $H_4$ dominates the other.

All the hulls for a given dot pattern. Hulls along the Pareto front are not dominated by any others.
Restated conjecture:

The points in area-perimeter space corresponding to polygonal hulls which best capture a perceived shape of a dot pattern lie on or close to the Pareto front.

To test the conjecture, 13 subjects were each presented with eight dot patterns and asked to draw polygonal hulls which, in their opinion, best captured the shape formed by each pattern.

Out of the 104 ($= 8 \times 13$) responses, 57 were on the Pareto front for their dot pattern, and all the rest were very close to it.

Statistical analysis of the results led to the conclusion that “the chance that Pareto-optimality has no influence on the subjects’ choices is effectively zero.”
A limitation of the this study is that it applied only to footprints in the form of polygonal hulls.

But for a region to be in some way representative of the location of the dots, it does not have to be a polygonal hull:
Intrinsic footprint criteria

C  Single connected component.
R  Topologically regular.
P  Boundary made up only of straight lines.
JC Boundary of each component is a Jordan curve.
SCC  Each component is simply connected.

Relational footprint criteria

CED  All curvature extrema of boundary coincide with dots.
ADB  All dots lie on the boundary.
NDB  No dots lie on the boundary.
FC  All dots in the topological closure.
Example classifications of footprint types

Polygonal hulls:

\[ [+C, +R, +P, +JC, +SCC, +CED, \pm ADB, -NDB, +FC] \]

Minimum bounding rectangle:

\[ [+C, +R, +P, +JC, +SCC, -CED, \pm ADB, -NDB, +FC] \]

Circle centred on centroid, with standard deviation as radius:

\[ [+C, +R, -P, +JC, +SCC, (+CED), -ADB, \pm NDB, -FC] \]

Union of covering disks:

\[ [\pm C, +R, -P, +JC, \pm SCC, -CED, -ADB, +NDB, +FC] \]
Application of footprint type classification

<table>
<thead>
<tr>
<th>FOOTPRINT ALGORITHM $ALG$</th>
<th>APPLICATION $APP$</th>
</tr>
</thead>
<tbody>
<tr>
<td>generates footprints of type $[+, +, -, -, +, \pm, +, -, -]$</td>
<td>requires footprints of type $[\pm, +, -, +, \pm, +, +, -, -, -]$</td>
</tr>
</tbody>
</table>

\[
1 - \frac{1}{6}\Delta([+, +, -, -, +, \pm, +, -, -], [\pm, +, -, +, \pm, +, +, -, -, -]) = 1 - \frac{\sqrt{7}}{6} = 0.56
\]

This gives a measure (admittedly crude) of the suitability of algorithm $ALG$ for use in application $APP$, on a scale from 0 (useless) to 1 (perfect match).
Aspects of Dot Patterns

The research reported above concentrated on the spatial aspects of dot patterns and their footprints.

Other important aspects are

- **thematic** — what dot patterns represent (e.g., collectives)
- **temporal** — how dot patterns change
Five dimensions of variability within the class of collectives:

1. Membership
2. Location
3. Coherence
4. Differentiation of roles
5. Depth
Membership

- Is the membership constant or variable?
- If variable, is the cardinality constant or variable?
- If cardinality is variable, how low can it get without destroying the collective? (2? 1? 0?)
Location

- Can the collective as a whole be assigned a location?
- If so, is the location fixed or variable?
- Independently of this, is the location of the members fixed or variable?
- If both the collective and its members have variable location, is the motion of the former coordinated with the motions of the latter?
Does the collective owe its coherence to cause or purpose?

Is the source of collectivity internal or external to the collective?

If the source of collectivity is internal purpose, is this a shared collective purpose or independent individual purposes?
Differentiation of Roles

▶ Do all the members participate in the same way?
▶ If not, how are their roles differentiated?
   ▶ hierarchically
   ▶ oligarchically
   ▶ individualistically
   ▶ other?
Are some or all of the members of the collective collectives themselves?
If so, are the members of the members collectives?
Etc, etc!
When Dot Patterns Change

This is the temporal aspect.

But this aspect can be divided into two “sub-aspects:

1. **Spatial** sub-aspect: Analysis of dynamic dot patterns from a purely spatial point of view (e.g., how the footprint changes).

2. **Thematic** sub-aspect: What can the patterns of movement exhibited by a dynamic dot pattern tell us about the collective it represents?
We looked for an efficient way to track the footprint of an evolving dot pattern.

Recomputing the footprint at every time step can be computationally expensive.

Therefore we proposed a method by which “the position of the dots in relation to the most-recently computed footprint is continuously monitored, and the footprint is only recomputed when the mismatch between the dot positions and the current footprint exceeds some preassigned threshold of accuracy”
A change identifier is an easily computed characteristic of the dot patterns which can be inspected to determine how much the dot pattern has changed since the footprint was last recomputed.

We use one or more change identifiers to determine when the footprint needs to be recomputed next.

This reduces the computation time, but at the expense of accuracy. The time/accuracy trade-off makes this another multi-objective optimisation problem.
<table>
<thead>
<tr>
<th></th>
<th><strong>BASE</strong></th>
<th><strong>SURROGATE</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>POSITION</strong></td>
<td>Centroid, i.e., mean position of the dots</td>
<td>Centre of minimum bounding rectangle</td>
</tr>
<tr>
<td><strong>EXTENT</strong></td>
<td>Variance of the dot positions</td>
<td>Area of minimum bounding rectangle</td>
</tr>
<tr>
<td><strong>CARDINALITY</strong></td>
<td>Number of dots</td>
<td>Not applicable</td>
</tr>
<tr>
<td><strong>DENSITY</strong></td>
<td>Number of dots divided by variance of dot positions</td>
<td>Number of dots divided by area of minimum bounding rectangle</td>
</tr>
</tbody>
</table>
Max tested the change identifiers on synthetic data comprising streams of dot patterns generated using separation, cohesion, and alignment (as in the “Boids” system of Reynolds, 1987).

Each data stream was run several times:

1. Once recomputing the “true” footprint at every time step.
2. Once for each change identifier, using a pre-set threshold to determine when the footprint should be recomputed.
3. Three additional runs using combinations of change identifiers.

Accumulated error ("accuracy cost") computed using the area of the symmetric difference between the most recently computed footprint and the “true” footprint, expressed as a percentage of the area of the latter and integrated over the duration of the run.
The results: Accuracy Cost vs Computation Time for different change identifiers

The Three-Level Analysis

A full account of the motion of a collective should include components at three levels of spatial granularity:

1. **Coarse level:** The motion of the collective as a single entity, as given by the motion of a representative point such as its geometric centroid.
2. **Intermediate level:** The changes to the footprint (as e.g., in Max Dupenois’ work)
3. **Fine level:** The motions of the individual members, considered as points.
Temporal Granularity

Fundamental notion is a refinement of the notion of “episode” introduced in the COSIT 2005 paper:

An episode (in the refined sense) is a maximal “chunk” of process that looks homogeneous when viewed at a certain granularity.

Here homogeneity is assessed with respect to some set of qualitative motion descriptors.

The motion of an individual or collective over an extended period may be regarded as the concatenation of a sequence of episodes, punctuated by transitions at which one episode gives way to the next.
A set of qualitative motion descriptors for Level 1

**SPEED:**
- Zero
- Constant non-zero
- Increasing
- Decreasing

**DIRECTION:**
- Linear
- Curving left
- Curving right

A more refined set of descriptors might include, for speed, constant, increasing or decreasing acceleration; and for direction, circular, spiralling in, and spiralling out.
Decomposition of motion into qualitative episodes
The chief qualitative characters of a footprint are **size**, **shape**, and **orientation**.

**SIZE:**
- Constant size
- Expansion
- Contraction

**ORIENTATION:**
- Constant orientation
- Clockwise rotation
- Anticlockwise rotation

**SHAPE** — a minefield! There are innumerable dimensions of possible variation, but there has been a lot of work on readily computable and usefully discriminatory shape descriptors.
Qualitative descriptors for Level 3

Here the collective is considered at the granularity level at which the motions of the individual members is apparent. Qualitative descriptors include:

- Uncoordinated
- Convergent
- Divergent
- Parallel
- Lagged
- Parallel-lagged
Five types of coordinated collective motion

- Convergent
- Divergent
- Parallel
- Lagged
- Parallel-lagged
For a group of individuals to be recognised as a collective, there must be *something* which confers sufficient unity to justify their being considered as collectively constituting a single entity.

In many cases this unity manifests itself through the **spatio-temporal** properties of the group.

Zena proposed a set of three **spatial coherence criteria** which can be used to identify the presence of collectives within spatio-temporal data-sets.
Spatial Coherence Criteria

- COMMON LOCATION
  A set of individuals, if sufficiently large, is regarded as spatially coherent if they have similar locations at a sufficient number of distinct times.

- COMMON MOVEMENT
  A set of individuals, if sufficiently large, is regarded as spatially coherent if they have similar movement at a sufficient number of distinct times.

- COMMON FORMATION
  A set of individuals, if sufficiently large, is regarded as spatially coherent if they maintain similar relative positions at a sufficient number of distinct times.

Note the deliberately vague terms highlighted in blue!
Zena Wood has tested the criteria on both synthetic and real data-sets.

The synthetic data-sets were created using Max Dupenois’ program for generating streams of “boid”-like dot patterns.

The real-world data-sets comprised records of the postions of 480 ships with the Solent (off Hampshire, UK) over a 24-hour period.
Three “species” of boid:

▶ $A^+$: cohesion = 1, alignment = 1, separation = 0.3
▶ $A^-$: cohesion = 0.6, alignment = 0.6, separation = 0.3
▶ B: cohesion = 0.33, alignment = 0.33, separation = 0.8

The A species are designed to favour the formation of collectives; species B is designed to inhibit it.

Data-sets were of three types:

▶ SD1: 20 individuals each of species $A^+$ and B.
▶ SD2: 20 individuals each of species $A^-$ and B.
▶ SD3: 40 individuals of species B.
The Experiments

The spatial coherence criteria were implemented in a program which could be applied to spatio-temporal datasets to determine *prima facie* candidates for collectives.

The program was applied to 60 different data-sets of each type (SD1, SD2, SD3), using each of the spatial coherence criteria to identify collectives.

Any individual identified as belonging to a collective was assigned to species A; other individuals were assigned to species B.

The results were compared with the known species of each individual to determine the effectiveness of the spatial coherence criteria for identifying instances of collectivity.
How to evaluate the results

Results of a categorisation experiment are often presented by means of a **confusion matrix**, which in this case has the form

\[
\begin{array}{cc}
\text{Identified} & \text{A} & \text{B} \\
\text{A} & TP & FN \\
\text{B} & FP & TN \\
\end{array}
\]

where $TP = \text{true positives}$, $FN = \text{false negatives}$, etc. Then

\[
\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN}
\]

\[
\text{precision} = \frac{TP}{TP + FP}
\]

\[
\text{recall} = \frac{TP}{TP + FN}
\]
## The Results

<table>
<thead>
<tr>
<th></th>
<th>Common Location</th>
<th>Common Movement</th>
<th>Common Formation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SD1</td>
<td>SD2</td>
<td>SD1</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.8713</td>
<td>0.8796</td>
<td>0.7367</td>
</tr>
<tr>
<td>Precision</td>
<td>0.7988</td>
<td>0.8337</td>
<td>0.6862</td>
</tr>
<tr>
<td>Recall</td>
<td>0.9925</td>
<td>0.9483</td>
<td>0.9858</td>
</tr>
</tbody>
</table>

- Recall is better than precision, i.e., fewer errors of *omission* than errors of *commission*.
- Common Location is the most reliable spatial coherence criterion, Common Formation the least.
- Overall, a clear demonstration that collectivity can be detected using spatial coherence criteria.
The different spatial coherence criteria were found to identify different ways in which groups of ships might be considered to form collectives.

For example:

- Common Location enables the locations of ports to be identified, as fixed places where ships naturally congregate.
- Common Formation highlighted areas where the trajectories of ships were concentrated, mainly in the vicinity of ports and along major shipping lanes.
Collectives provide an example of **multi-aspect phenomena**. A collective’s location at a time can be represented by a **footprint**. There are many kinds of footprint, and many ways of generating and evaluating them. Likewise, there are many kinds of collective, with five major dimensions of variation along which to classify them, leading to a **taxonomy of collectives**. **Change identifiers** can provide an efficient means of tracking footprints in real time, although there is an inevitable trade-off between computation time and accuracy. Collective motion may be described at **three granularity levels** and decomposed into **episodes** homogeneous with respect to **qualitative motion descriptors**. Spatial collectives owe their unity to **spatial coherence criteria**, which may be used for automated identification of collectives in spatio-temporal data-sets.
THE END

Thank you for listening!

Any questions?