The Relational Shape of Structural Complexity and Neighborhood Density

J. R. Landers

Abstract

A mathematical object becomes more richly specified as metadata is added to its description. A natural question is whether this additional structure also changes the object's *relational position*: does a more detailed object tend to lie farther from the other objects in a population, and therefore occupy a sparser neighborhood? For a broad class of additive distances the answer is exact rather than heuristic. Additional distinguishing coordinates can only increase pairwise separation, while fixed-radius neighborhoods form a nested, progressively thinning family. This yields a natural rate of relational isolation, and in a simple random metadata model the expected local density decays exponentially with descriptive complexity. The effect also admits a direct information-theoretic interpretation.

Objects, descriptions, and distance

Let

$$\mathcal{O} = \{O_1, \dots, O_K\}$$

be a finite collection of K objects. At descriptive complexity N, object O_i is represented by metadata

$$O_i^{(N)} = (o_{i1}, \dots, o_{iN}),$$

where o_{in} denotes the value of the nth metadata coordinate. Assume separation is measured by an additive nonnegative distance

$$\delta_N(O_i, O_j) = \sum_{n=1}^{N} d_n(o_{in}, o_{jn}), \qquad d_n(x, y) \geq 0,$$

with each d_n measuring dissimilarity along coordinate n. The next level of structural complexity appends one further coordinate, so that

$$\delta_{N+1}(O_i, O_j) = \delta_N(O_i, O_j) + d_{N+1}(o_{i,N+1}, o_{j,N+1}).$$

Thus added structure acts geometrically: it contributes another nonnegative direction along which two objects may differ. For a fixed object O_i, define its mean relational separation

$$D_N(i) = \frac{1}{K-1} \sum_{j \neq i} \delta_N(O_i, O_j).$$

Also define its normalized radius-r neighborhood density by

$$\rho_N(i; r) = \frac{1}{K-1} \sum_{j \neq i} \mathbf{1}\{\delta_N(O_i, O_j) \leq r\},$$

where $\mathbf{1}\{\cdot\}$ is the indicator function. The quantity $\rho_N(i; r)$ is simply the fraction of the remaining objects lying within distance r of O_i.

Proposition — Structural separation and neighborhood thinning. For every i, every $j \neq i$, and every radius $r \geq 0$,

$$\delta_{N+1}(O_i, O_j) \geq \delta_N(O_i, O_j),$$

and consequently

$$D_{N+1}(i) \geq D_N(i), \qquad \rho_{N+1}(i; r) \leq \rho_N(i; r).$$

If the appended metadata distinguishes O_i from at least one other object, then $D_{N+1}(i) > D_N(i)$ whenever the corresponding coordinate distance is strictly positive.

Proof. Because $d_{N+1} \geq 0$,

$$\delta_{N+1}(O_i, O_j) - \delta_N(O_i, O_j) = d_{N+1}(o_{i,N+1}, o_{j,N+1}) \geq 0.$$

Averaging this inequality over $j \neq i$ gives $D_{N+1}(i) \geq D_N(i)$. Since no pairwise distance decreases, an object that lies outside a radius-r neighborhood at level N cannot move into that neighborhood merely by appending a nonnegative distance contribution. Hence every indicator in $\rho_N(i; r)$ is nonincreasing, and so is their average. Strict growth of $D_N(i)$ follows if at least one added contribution is positive. □

The proposition gives a simple relational picture: richer descriptions do not merely enlarge the internal representation of an object. When the new structure is distinguishing, they push that object outward relative to the population, thinning its fixed-radius neighborhood.

Neighborhoods as a filtration

The density statement can be sharpened from a scalar inequality to a set-theoretic one. For radius $r \geq 0$, define the radius-r neighborhood of O_i at complexity N by

$$\mathcal{N}_r^{(N)}(O_i) = \{O_j \in \mathcal{O} : j \neq i,\ \delta_N(O_i, O_j) \leq r\}.$$

Its normalized cardinality is exactly the density introduced above,

$$\rho_N(i; r) = \frac{|\mathcal{N}_r^{(N)}(O_i)|}{K-1}.$$

Corollary — Nested relational neighborhoods. For every fixed object O_i and radius $r \geq 0$,

$$\mathcal{N}_r^{(N+1)}(O_i) \subseteq \mathcal{N}_r^{(N)}(O_i).$$

Hence increasing descriptive complexity generates the descending filtration

$$\mathcal{N}_r^{(1)}(O_i) \supseteq \mathcal{N}_r^{(2)}(O_i) \supseteq \cdots \supseteq \mathcal{N}_r^{(N)}(O_i) \supseteq \cdots.$$

Proof. If $O_j \in \mathcal{N}_r^{(N+1)}(O_i)$, then $\delta_{N+1}(O_i, O_j) \leq r$. Since $\delta_N(O_i, O_j) \leq \delta_{N+1}(O_i, O_j)$, it follows that $\delta_N(O_i, O_j) \leq r$, so $O_j \in \mathcal{N}_r^{(N)}(O_i)$. □

This filtration makes the relational effect of complexity visible without averaging: once an object leaves a fixed-radius neighborhood, it cannot re-enter under further additive nonnegative refinement. Complexity therefore removes neighbors monotonically, although it may remove none at a particular step.

A complementary scalar quantity records how rapidly an object becomes isolated. Define the *relational isolation velocity*

$$v_i(N) = D_{N+1}(i) - D_N(i).$$

Using the additive form of δ_N gives the identity

$$v_i(N) = \frac{1}{K-1} \sum_{j \neq i} d_{N+1}(o_{i,N+1}, o_{j,N+1}) \geq 0.$$

Thus $v_i(N)$ is not merely an analogy to motion: it is exactly the average distinguishing contribution of the newly appended coordinate. A coordinate shared by the entire population has zero velocity; a coordinate that strongly separates O_i from its peers produces a larger outward step.

The information carried by a new coordinate

For categorical metadata, the connection to information is especially transparent. Suppose coordinate n takes values in a finite alphabet $\mathcal{A}_n$ with probabilities $p_n(x)$, and use Hamming distance,

$$d_n(x, y) = \mathbf{1}\{x \neq y\}.$$

For two independently sampled objects,

$$\Pr(o_{in} \neq o_{jn}) = 1 - \sum_{x \in \mathcal{A}_n} p_n(x)^2.$$

Therefore

$$\mathbb{E}\,\delta_N(O_i, O_j) = \sum_{n=1}^{N} \left(1 - \sum_{x \in \mathcal{A}_n} p_n(x)^2\right).$$

Each coordinate contributes exactly its probability of distinguishing two sampled objects. A constant coordinate contributes zero; a highly diverse coordinate contributes more. The term $\sum_x p_n(x)^2$ is the collision probability and is directly related to Rényi entropy of order 2,

$$H_2(p_n) = -\log\left(\sum_x p_n(x)^2\right).$$

Thus the growth of expected distance is governed not by metadata count alone, but by the *distinguishing information* carried by the appended structure.

In the uniform q-ary case, $p_n(x) = 1/q$, so

$$\mathbb{E}\,\delta_N = N\left(1 - \frac{1}{q}\right).$$

Expected separation grows linearly with description length. For real-valued independent coordinates of variance σ^2, squared Euclidean distance yields the analogous identity

$$\mathbb{E}\,\delta_N^2 = 2N\sigma^2,$$

so the typical Euclidean scale grows on the order of $\sqrt{N}$.

A stochastic thinning law

The filtration above is deterministic. Under a simple random metadata model, one can also quantify how rapidly a fixed neighborhood disappears. Suppose every metadata coordinate is drawn independently and uniformly from a q-symbol alphabet, and let δ_N be Hamming distance. For any pair of independently sampled objects, the number of mismatched coordinates is binomial:

$$\delta_N(O_i, O_j) \sim \mathrm{Bin}(N, p), \qquad p = 1 - \frac{1}{q}.$$

Therefore, for a fixed radius $r \geq 0$ and $R = \lfloor r \rfloor$,

$$\mathbb{E}\,\rho_N(i; r) = \Pr\{\delta_N(O_i, O_j) \leq R\} = \sum_{m=0}^{R} \binom{N}{m} \left(1 - \frac{1}{q}\right)^m \left(\frac{1}{q}\right)^{N-m}.$$

Equivalently,

$$\mathbb{E}\,\rho_N(i; r) = q^{-N} \sum_{m=0}^{R} \binom{N}{m} (q-1)^m.$$

For fixed r and fixed $q \geq 2$, the finite sum is polynomial in N of degree R, while q^{-N} is exponential. Hence

$$\mathbb{E}\,\rho_N(i; r) = \Theta\big(N^R q^{-N}\big) \qquad (N \to \infty).$$

So in this model local density does not merely decline: every fixed-radius neighborhood becomes exponentially sparse, up to a polynomial factor. The result separates two scales cleanly. Pairwise distance grows linearly in expectation, while the probability of remaining genuinely nearby collapses exponentially.

Presumptions of the principle

The essential distinction is between *more metadata* and *more distinguishing structure*. Appending the same value to every object increases description length but leaves all distances unchanged. Redundant coordinates may likewise add little genuinely new separation. Moreover, dimension-normalized distances can behave differently because the normalization itself changes with N.

Under a distance that accumulates nonnegative coordinate-wise differences, however, the relational consequence is sharp:

distinguishing structural complexity increases separation and thins local neighborhoods.

This suggests viewing structural complexity not only as an intrinsic property of an object, but as a geometric property of its position among alternatives. As an object becomes more specifically described, its identity is expressed partly through the progressive disappearance of nearby peers.