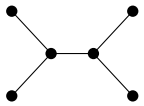
Why Four-Point Tests Cannot Recognize Cactus Metrics

1. From distances to trees and cacti

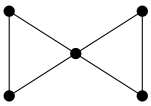
Suppose we know the distances between a finite collection of objects, but not the network connecting them. Can we reconstruct a network, or at least decide what kind of network could produce those distances? For example, the objects might represent observed species, while unobserved branching points represent common ancestors. Tree reconstruction from dissimilarity data is a classical motivation for this question [1].

A *metric* on a set X is a function $d : X \times X \to [0, \infty)$ such that $d(x, y) = d(y, x)$, $d(x, y) = 0$ exactly when $x = y$, and $d(x, z) \leq d(x, y) + d(y, z)$. A finite connected graph with positive edge lengths defines a metric: the distance between two points is the length of a shortest path joining them. Throughout, each edge includes all the points of an interval of the assigned length. The resulting space is the graph's *geometric realization*. Thus points may lie inside edges, and a shortest path has a midpoint.

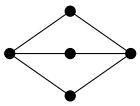
A *tree* is a connected graph with no cycles. A *cactus* is a connected graph in which any two distinct simple cycles have at most one vertex in common. A simple path has no repeated vertices; a simple cycle repeats only its starting vertex at the end. Trees are cacti, as are single cycles and collections of cycles attached to one another at single vertices. A weighted cactus with its path distance is a *metric cactus*; its distinct simple cycles meet in at most one point, even when edge interiors are included.



a tree



a cactus



not a cactus

We say that (X, d) has a *tree realization* or a *cactus realization* if its labelled points can be placed in such a weighted graph with all their distances preserved. Additional, unlabelled points and vertices are allowed. This is a condition on the distances, not on a particular drawing of a graph.

Write $d_{ij} = d(i, j)$ for labelled points. Every metric on at most three points has a tree realization. For three labels, use a three-armed star with arm length $(d_{ij} + d_{ik} - d_{jk})/2$ at label i. The triangle inequalities make these lengths nonnegative; zero-length arms are contracted.

The classical *four-point theorem* says that a finite metric has a tree realization if and only if, for every four distinct labels, the largest of

$$S_1 = d_{12} + d_{34}, \qquad S_2 = d_{13} + d_{24}, \qquad S_3 = d_{14} + d_{23}$$

is attained at least twice. For the four vertices of a unit square, in cyclic order, these sums are $2, 4, 2$. Its distances therefore cannot be realized by a tree, although the square itself is a cactus.

We use the classical theorem as background. The question of this note is whether some other test on these same six distances can recognize cactus realizations. We will prove that no such universal four-point test exists, and exhibit a five-point obstruction.

2. Every four-point metric has a cactus realization

Proposition 1. *Every metric on four labels has a cactus realization.*

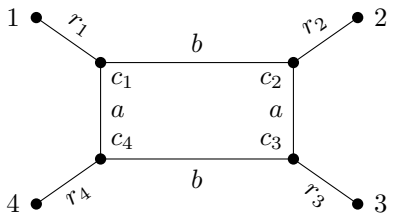
Proof. Relabel the points so that $S_2 \geq S_1, S_3$, and set

$$a = \frac{S_2 - S_1}{2}, \qquad b = \frac{S_2 - S_3}{2}.$$

Make a cycle with corners c_1, c_2, c_3, c_4 in this order and consecutive edge lengths b, a, b, a. At corner c_i, attach a segment of length r_i and place label i at its other end, where

$$r_1 = \frac{d_{12} + d_{14} - d_{24}}{2}, \qquad r_2 = \frac{d_{12} + d_{23} - d_{13}}{2},$$
$$r_3 = \frac{d_{23} + d_{34} - d_{24}}{2}, \qquad r_4 = \frac{d_{14} + d_{34} - d_{13}}{2}.$$

The numbers a, b are nonnegative by the choice of S_2, and all r_i are nonnegative by the triangle inequalities.



Only the edges of the rectangle are present; its interior is not part of the graph. Between adjacent corners the direct edge is a shortest route, and between opposite corners both routes have length $a + b$. Substitution gives all six required distances:

$$r_1 + r_2 + b = d_{12}, \qquad r_2 + r_3 + a = d_{23},$$
$$r_3 + r_4 + b = d_{34}, \qquad r_4 + r_1 + a = d_{14},$$
$$r_1 + r_3 + a + b = d_{13}, \qquad r_2 + r_4 + a + b = d_{24}.$$

Contract all zero-length segments. If exactly one of a, b is zero, keep only one of the resulting parallel edges; if both are zero, the cycle becomes a single point. These operations preserve distances and leave a tree or a cactus. Distinct labels remain distinct because their prescribed distances are positive. □

This construction gives a geometric reading of the four-point condition. Since S_2 is maximal, that condition holds exactly when $a = 0$ or $b = 0$: the cycle collapses to a tree. Allowing a genuine cycle therefore admits *every* four-point metric. To find a restriction on cactus metrics, we must look beyond an isolated set of four labels.

3. A global restriction: two separating points

A set $P \subset Z \setminus \{x, y\}$ *separates* x and y if deleting P leaves no path between them. Separation concerns all paths, not only shortest paths. Any path in a finite graph contains a simple path with the same endpoints, obtained by removing loops. On a tree, one interior point of the unique x–y path suffices. On a circle, one point never suffices, but two suitably chosen points do.

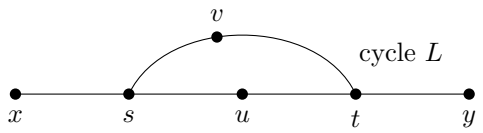
Theorem 2. *Let Z be the geometric realization of a finite connected positively weighted graph. Then Z is a cactus if and only if each pair $x \neq y$ can be separated by at most two points. In a cactus, the separating set P can moreover be chosen so that*

$$1 \leq |P| \leq 2, \qquad d(x,p) \geq \frac{d(x,y)}{2}, \qquad d(y,p) \geq \frac{d(x,y)}{2} \quad \textit{for every } p \in P.$$

Proof. First suppose the graph is not a cactus. Two distinct cycles sharing more than one vertex contain a *theta configuration*: three paths between some two points s, t, with pairwise disjoint interiors. To see this, take a portion of one cycle outside the other, together with the two routes in the other cycle between its endpoints. Two points outside $\{s, t\}$ cannot meet all three paths. Thus they cannot separate s, t.

Now suppose Z is a cactus. Fix $x \neq y$, put $r = d(x, y)$, and choose a shortest path g from x to y. Let u be its midpoint. If $\{u\}$ separates the endpoints, we are done.

Otherwise choose a simple x–y path β avoiding u. Starting at u, follow g in each direction until first meeting β, at points s, t, with s before u and t after u. The interior of this portion of g is disjoint from β. Together with the s–t portion of β, it forms a simple cycle L.



Any other simple path β' avoiding u produces a cycle L' in the same way. Both L and L' contain a small portion of g around u, so they share more than one point. The cactus property forces $L' = L$. Consequently every such β' contains $L \setminus g$.

We can choose $v \in L \setminus g$ with $d(x, v) = r/2$. Indeed, along the s–t portion of β, the continuous function $z \mapsto d(x, z)$ starts below $r/2$ and ends above $r/2$. A point where it equals $r/2$ cannot lie on g, since the only such point on g is u. The triangle inequality then gives $d(y, v) \geq r/2$.

Every simple x–y path meets u or v, hence $P = \{u, v\}$ separates the endpoints. Since u is a midpoint, both points satisfy the required bounds. □

The constant $1/2$ is best possible: any separating set meets g, and the distances from a point of g to its two endpoints sum to r. Theorem 2 is the finite-graph case of a more general separation theorem [2, Proposition 5.7].

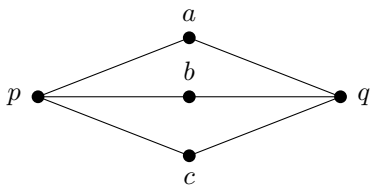
4. A five-point obstruction

The distance bounds in Theorem 2 have a useful consequence. A *midpoint* of x, y is a point m satisfying $d(x, m) = d(m, y) = d(x, y)/2$. Different shortest paths could, in principle, have different midpoints.

Corollary 3. *In a metric cactus, any two distinct points have at most two midpoints.*

Proof. Write $r = d(x, y)$ and choose the separating set P from Theorem 2. For any midpoint m, concatenate shortest paths from x to m and from m to y. Their total length is r, so this is a shortest x–y path, and it must meet P. On this path the only point at distance at least $r/2$ from both endpoints is m. Thus $m \in P$. All midpoints belong to the same set P, so there are at most two. □

Now consider the complete bipartite graph $K_{2,3}$: its two vertex classes are $\{p, q\}$ and $\{a, b, c\}$, with an edge between every pair in different classes and no other edges. Give every edge length 1.



Its vertex metric satisfies

$$d(p, q) = d(a, b) = d(a, c) = d(b, c) = 2, \qquad d(p, z) = d(q, z) = 1 \quad (z \in \{a, b, c\}).$$

If these five points had any cactus realization, the images of a, b, c would be three distinct midpoints of the images of p, q, contradicting the corollary. This rules out *every* cactus realization, not just the graph drawn above.

Nevertheless, each four-point restriction has a cactus realization: deleting a, b, or c leaves the distances of a unit four-cycle; deleting p or q leaves the distances of a three-leaf star with unit edges. Thus separately realizable four-point data need not fit into one common cactus.

Corollary 4. *There is no test depending only on the six distances of four labels such that a finite metric has a cactus realization exactly when every four-point restriction passes the test.*

Proof. Any necessary test must accept every four-point metric, by Proposition 1. It therefore accepts every four-point restriction of the $K_{2,3}$ metric, although that metric has no cactus realization. □

Five is consequently the smallest number of points at which a metric can fail to have a cactus realization. This does not establish a sufficient five-point test. The separation theorem supplies a different kind of characterization: it uses auxiliary points and paths in the whole space, rather than only the distances within each small sample.

References

[1] P. Buneman, *The recovery of trees from measures of dissimilarity*, in Mathematics in the Archaeological and Historical Sciences, Edinburgh University Press, 1971, pp. 387–395.

[2] R. Khanfir, *Convergences of looptrees coded by excursions*, Ann. Inst. H. Poincaré Probab. Statist. **62** (2026), 363–408. doi:10.1214/24-AIHP1514.