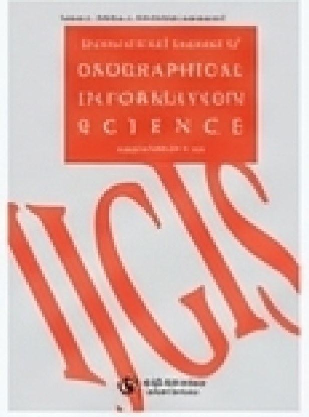
Beyond Coordinates: Measuring Distance in a Social World

An Introduction to the Social Distance Metric



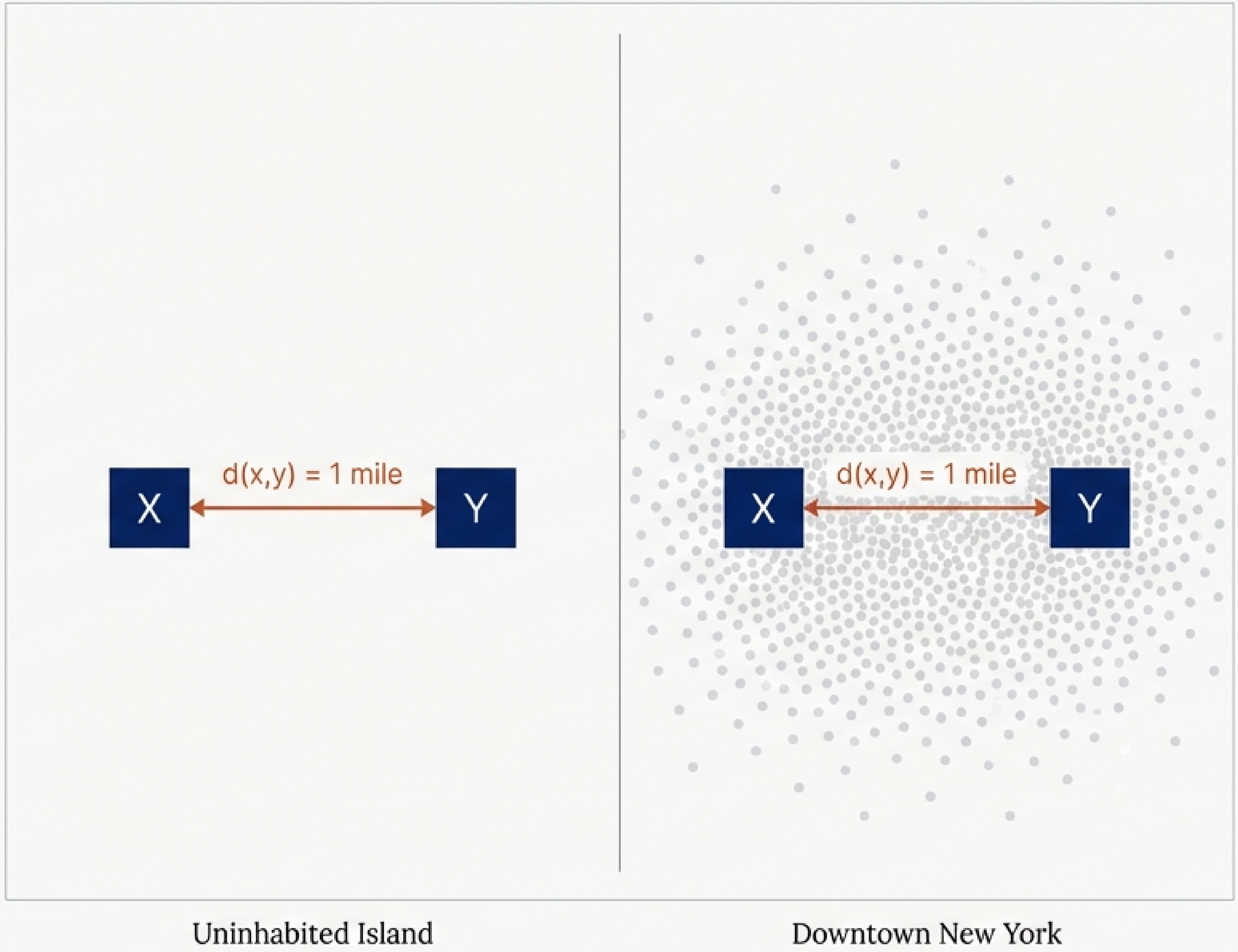
Based on “Social Distance metric: from coordinates to neighborhoods” by Vagan Terziyan, *International Journal of Geographical Information Science*, 2017.

Taylor & Francis
Taylor & Francis Group

Traditional Distance is Context-Blind

Standard distance metrics operate in a vacuum. They depend only on the coordinates of two points, ignoring the surrounding data landscape.

Consider this analogy: Two people, one mile apart on an uninhabited island (like Robinson Crusoe and Man-Friday), are socially much closer than two people one mile apart in downtown New York City, surrounded by millions. Traditional metrics see both distances as identical. They miss the context.



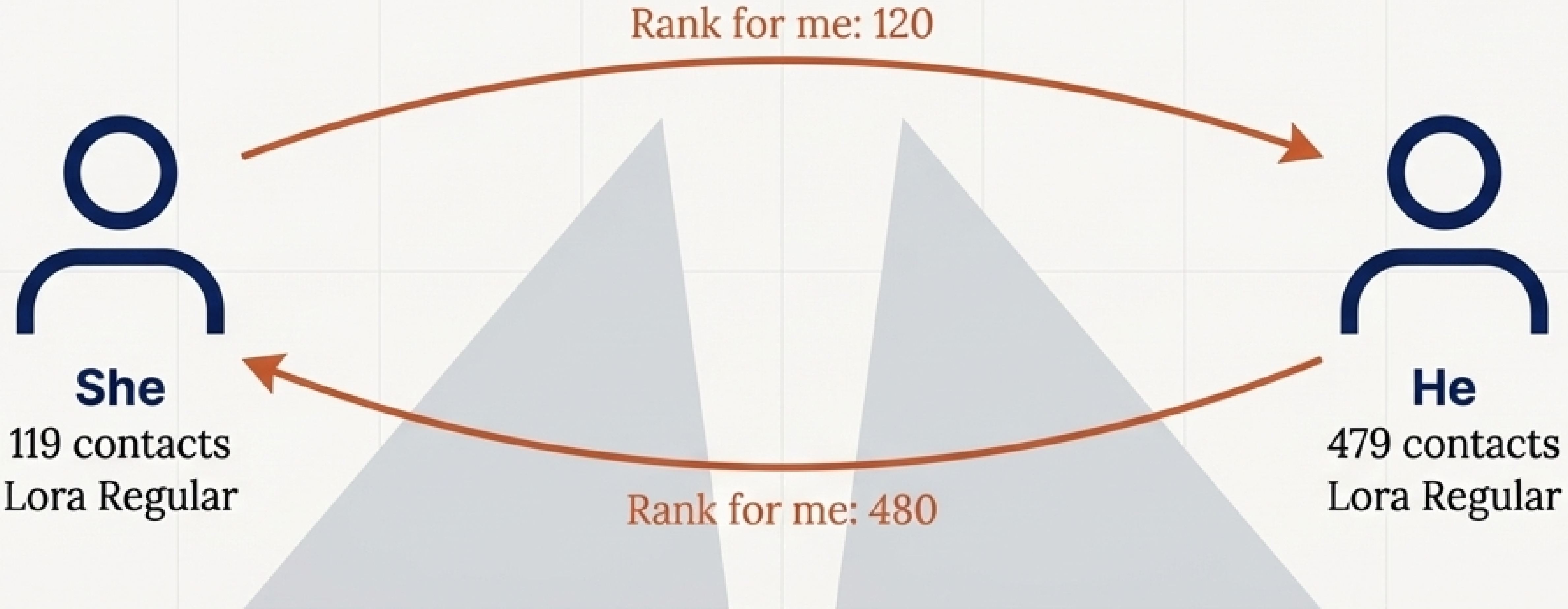
Are these distances truly the same?

A New Lens: Distance as a Social Relationship

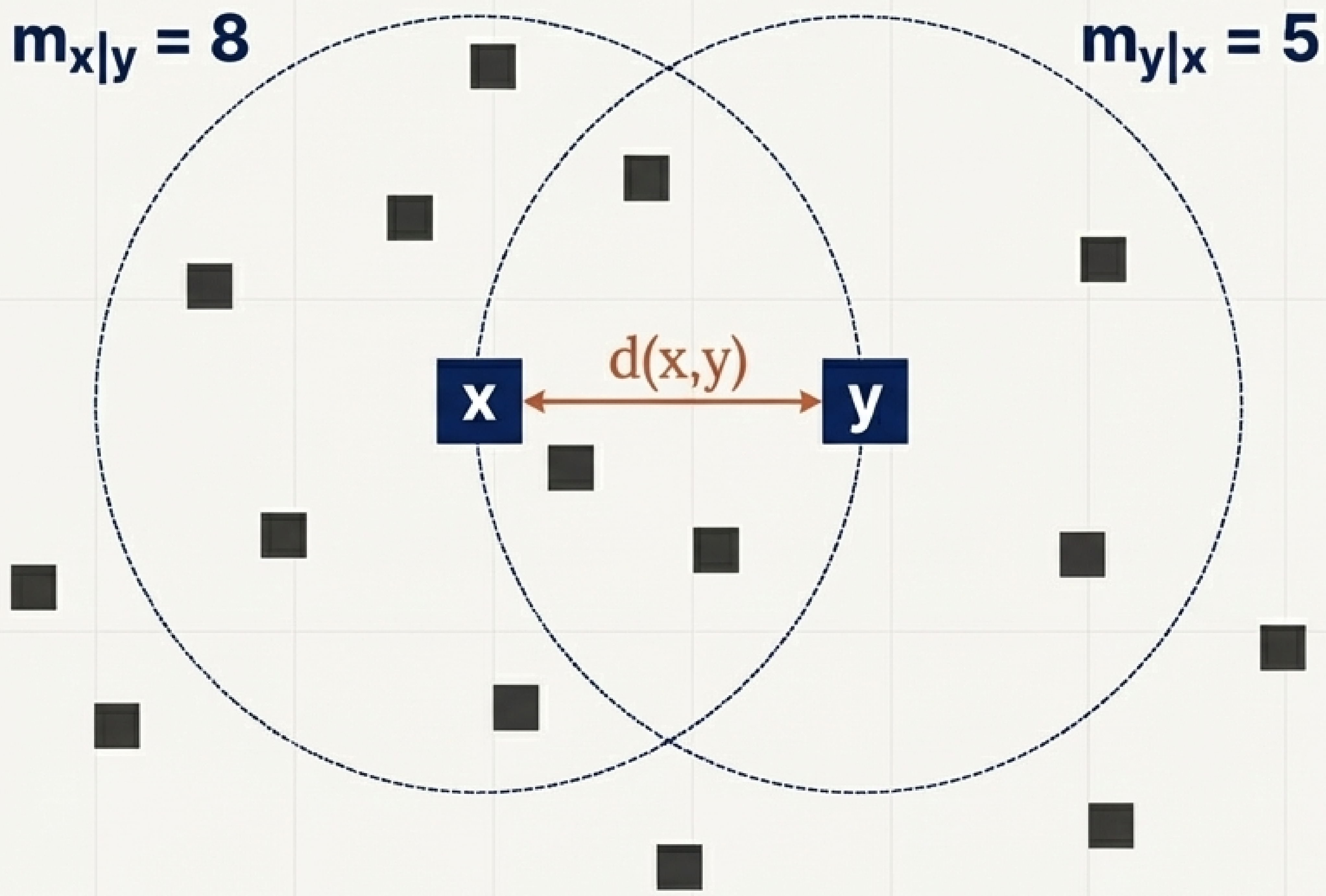
Instead of coordinates, what if we measured distance based on mutual importance, or ‘social rank’?

Let’s use a real-world example from LinkedIn:

- She has 119 contacts. He has 479 contacts.
- He invites her. If she accepts:
  - He becomes her 120th contact.
  - She becomes his 480th contact.
- There’s a clear asymmetry. He represents 1/480 (~0.2%) of her network’s attention, while she represents 1/120 (~0.83%) of his—four times more.
- This ‘social gap’ is a form of distance.



Quantifying the ‘Social Gap’ with Mutual Ranks



We can apply this concept to any dataset. For any two points, x and y, we calculate their mutual ranks based on a standard underlying metric (e.g., Euclidean).

$m_{x|y}$: The rank of y in the list of x’s nearest neighbors.
$m_{y|x}$: The rank of x in the list of y’s nearest neighbors.

Asymmetry is the norm: In this example, y is the 8th closest point to x. But x is only the 5th closest point to y.

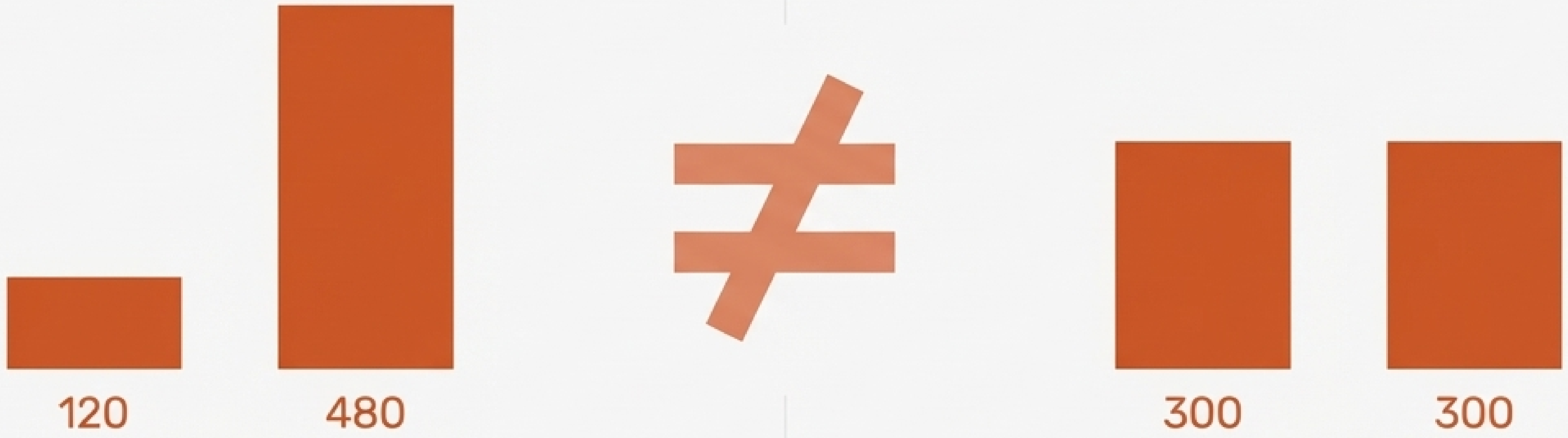
How do we combine $m_{x|y} = 8$ and $m_{y|x} = 5$ into a single, meaningful distance?

The Flaw of a Simple Average

An Arithmetic mean seems like an obvious way to combine the two ranks, but it fails to capture the essence of the "social gap." It sees no difference between a highly asymmetric relationship and a perfectly balanced one. We need a mean that penalizes asymmetry.

Case 1 (Asymmetric)

Case 2 (Symmetric)



Arithmetic Mean: $\frac{120 + 480}{2} = \mathbf{300}$

Arithmetic Mean: $\frac{300 + 300}{2} = \mathbf{300}$

The Lehmer Mean: A Tool for Penalizing Asymmetry

The Lehmer mean family provides a powerful way to average numbers while accounting for their difference.

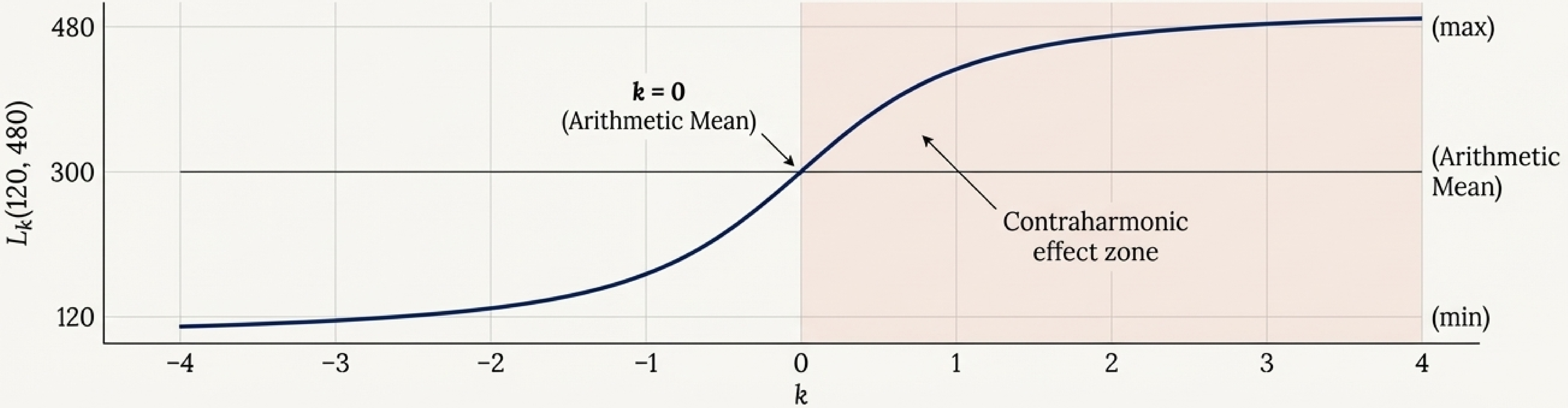
$$L_k(x_1, \ldots, x_n) = \frac{\sum(x_i^{k+1})}{\sum(x_i^k)}$$

- For $k > 0$, it produces a contraharmonic effect: the mean is greater than the Arithmetic mean, and the penalty increases as the difference between the numbers grows.
- $k = 1$ gives the Contraharmonic Mean: $(x_1^2 + x_2^2)/(x_1 + x_2)$
- $k = 0$ gives the Arithmetic Mean.
- $k < 0$ produces a harmonic effect.

Example (120 vs. 480):

Arithmetic Mean (k=0): 300

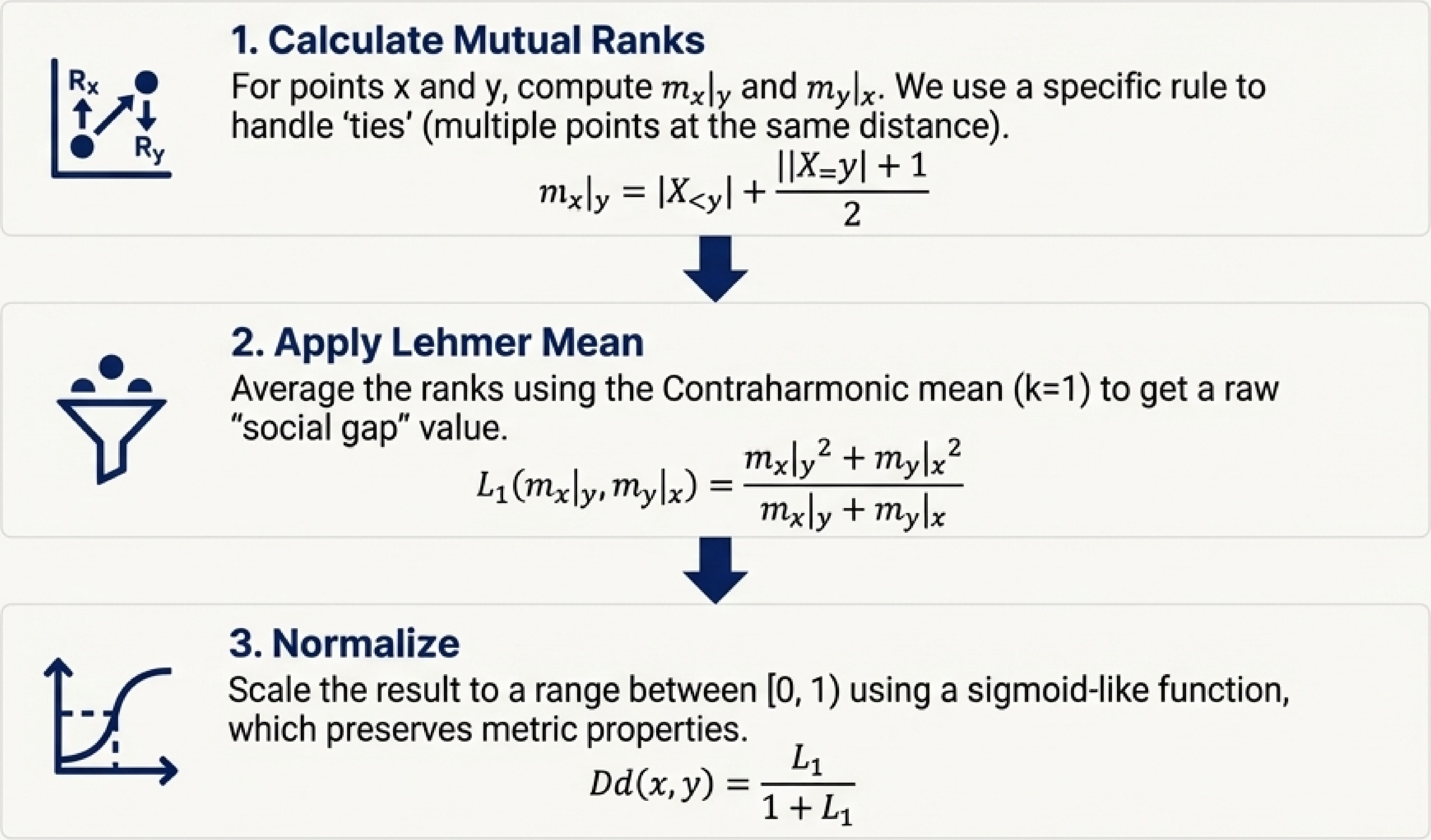
Contraharmonic Mean (k=1): **408** (A significant penalty)



NotebookLM

Defining the Social Distance Metric (Dd)

The Social Distance Dd(x,y) is built on an underlying metric d in three steps:



1. Calculate Mutual Ranks

For points x and y, compute $m_x|_y$ and $m_y|_x$. We use a specific rule to handle ‘ties’ (multiple points at the same distance).

$$m_x|_y = |X_{<y}| + \frac{||X_{=}y| + 1}{2}$$

2. Apply Lehmer Mean

Average the ranks using the Contraharmonic mean (k=1) to get a raw “social gap” value.

$$L_1(m_x|_y, m_y|_x) = \frac{m_x|_y{}^2 + m_y|_x{}^2}{m_x|_y + m_y|_x}$$

3. Normalize

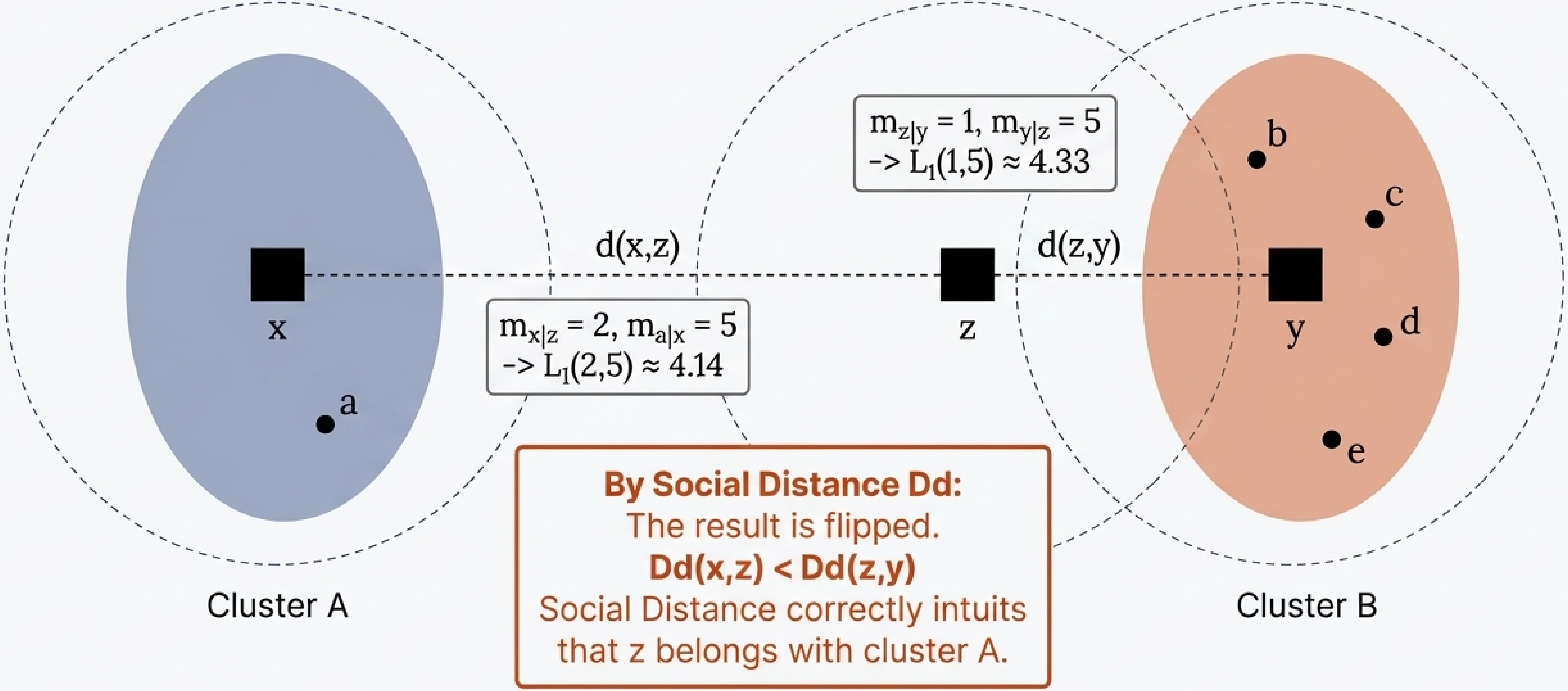
Scale the result to a range between [0, 1) using a sigmoid-like function, which preserves metric properties.

$$Dd(x, y) = \frac{L_1}{1 + L_1}$$

The function Dd is a true metric, satisfying non-negativity, identity, symmetry, and the triangle inequality.

The Contraharmonic Paradox: When Further is Closer

By Euclidean distance d:
z is closer to y in cluster B.
$d(x,z) > d(z,y)$



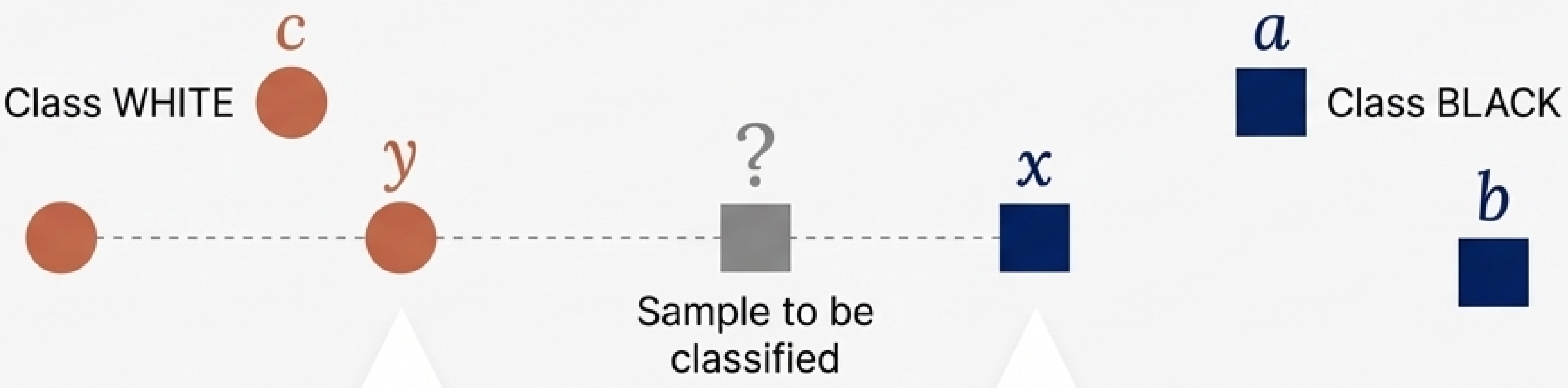
$m_{z|y} = 1$, $m_{y|z} = 5$
-> $L_1(1,5) \approx 4.33$

$m_{x|z} = 2$, $m_{a|x} = 5$
-> $L_1(2,5) \approx 4.14$

By Social Distance Dd:
The result is flipped.
$Dd(x,z) < Dd(z,y)$
Social Distance correctly intuits that z belongs with cluster A.

Application 1: More Intelligent k-NN Classification

The choice of metric is critical for k-NN classifiers. Social Distance can significantly alter classification outcomes by incorporating neighborhood context. The metric makes a classification decision based not just on proximity, but on the ‘social' fit within the local data structure.



Using standard distance d

The nearest neighbor is x (class BLACK).

$$d(?, x) < d(?, y)$$

Classification: **BLACK**

Using Social Distance D_d

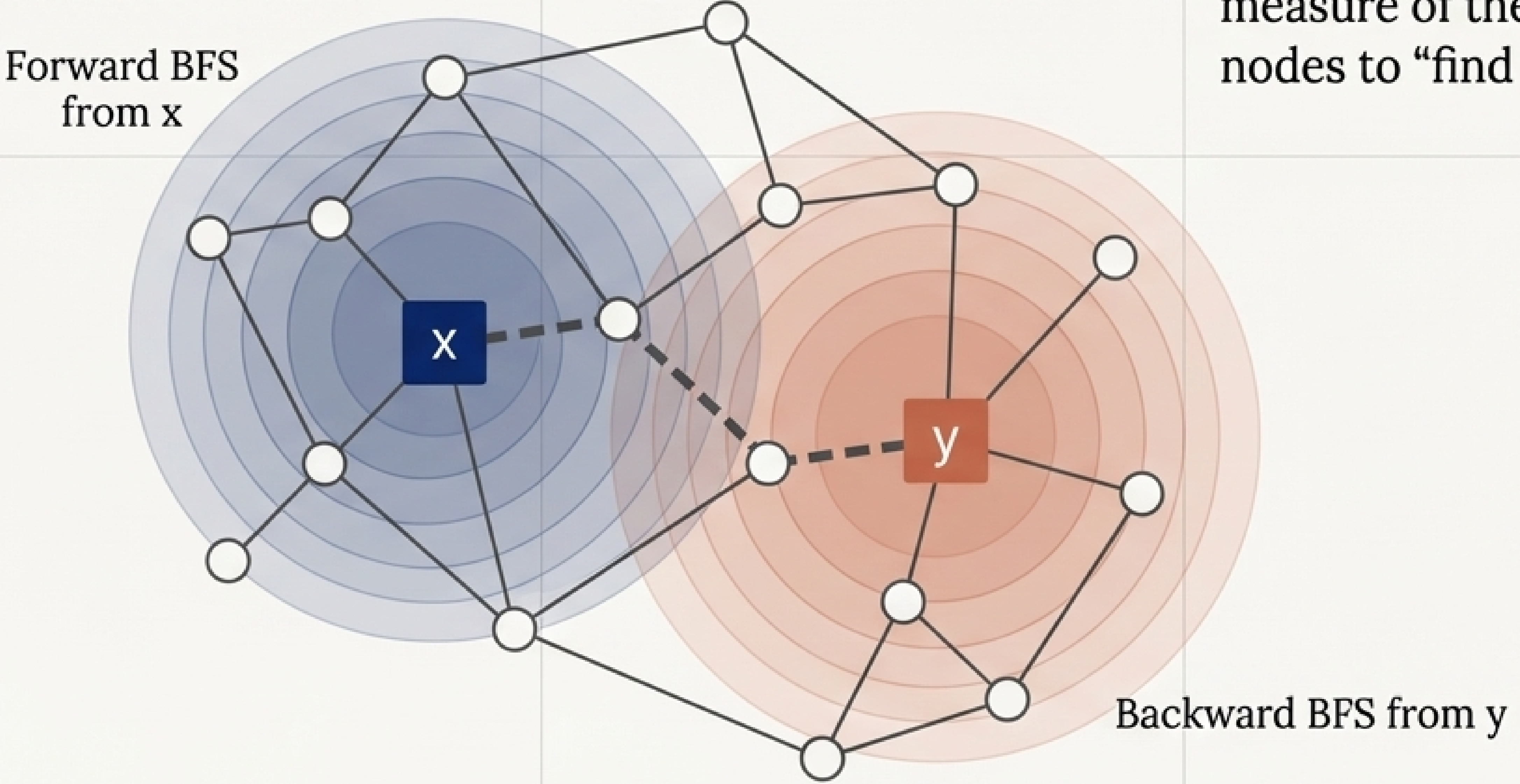
The metric accounts for local density. The nearest neighbor becomes y (class WHITE).

$$D_{1d}(?, x) \approx 0.714 > D_{1d}(?, y) = 0.6$$

Classification: **WHITE**

Application 2: Measuring Distance in Networks and Graphs

The Social Distance concept translates naturally to graphs, where distance is often more complex than shortest path. The underlying metric $d(x,y)$ is the shortest path length. The ranks $m_{x|y}$ and $m_{y|x}$ represent the number of nodes explored in a breadth-first search from each direction until the other other node is found. Social Distance beı an aggregate measure of the bidirectional effort required for two nodes to “find each other”.



Calculation for this graph:

$d(x,y) = 2$

$m_{x|y} = 8$ (Forward BFS from x)

$m_{y|x} = 7$ (Backward BFS from y)

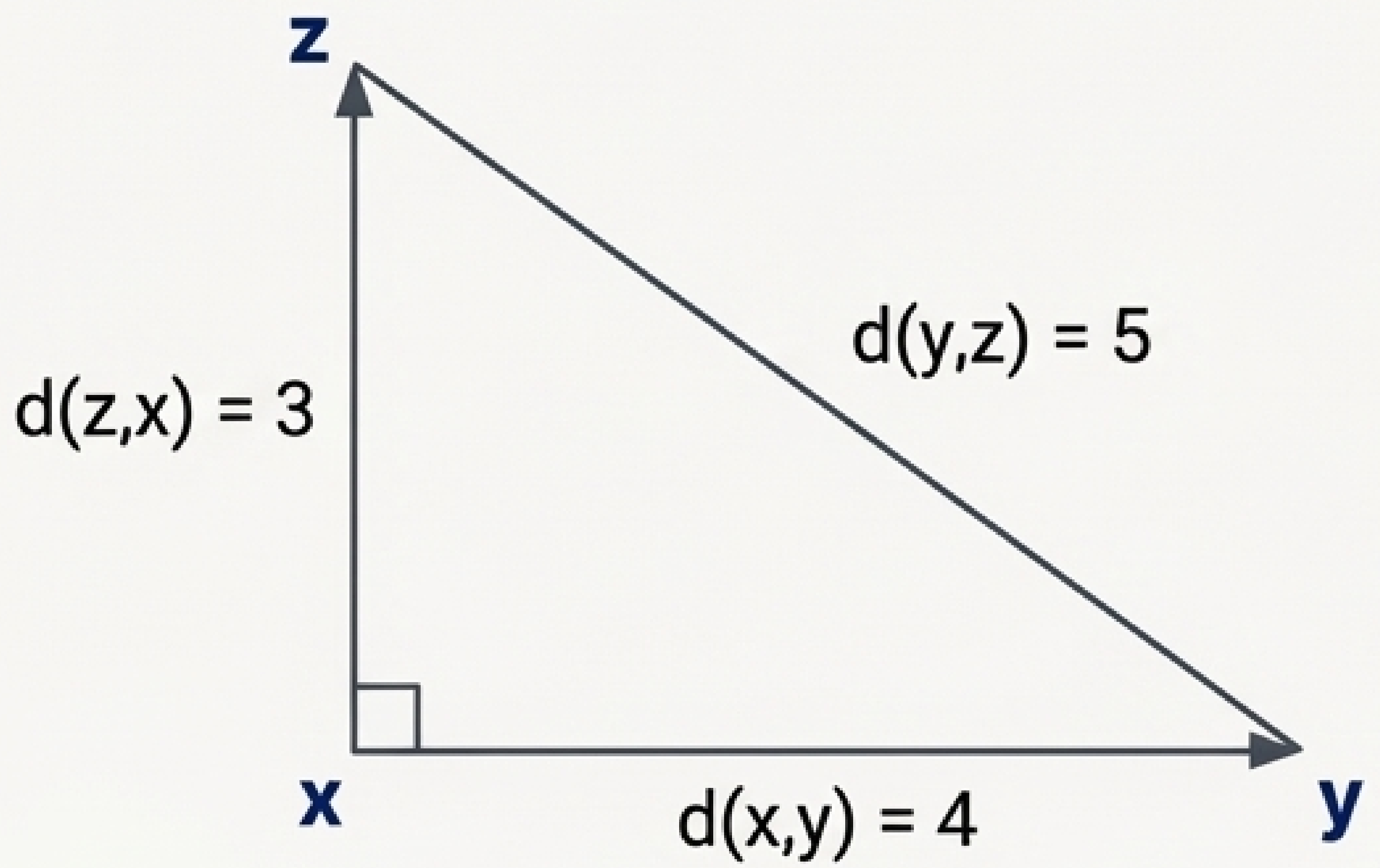
$L_1(8,7) = 7.53$

$D_1 d(x,y) \approx 0.88$

Variation 1: The ‘Social Traveler’ Distance

The standard Social Distance only uses the *order* of neighbors, not the actual distances to them. The **Social Traveler Distance** (~$\tilde{D}_d$) incorporates this information by using a weighted sum of ranks.

- $M_{X|y}$: Sum of distances $d(x, x_i)$ for all neighbors x_i closer than y.
- This variation is sensitive to the scale of the underlying space.



Social Distance (D_d)

$m_{y\|x}$	= 1
$m_{x\|y}$	= 2
D_{1d}(x,y)	= 0.625

If d(x,y) changes to 400, D_d remains **0.625**.

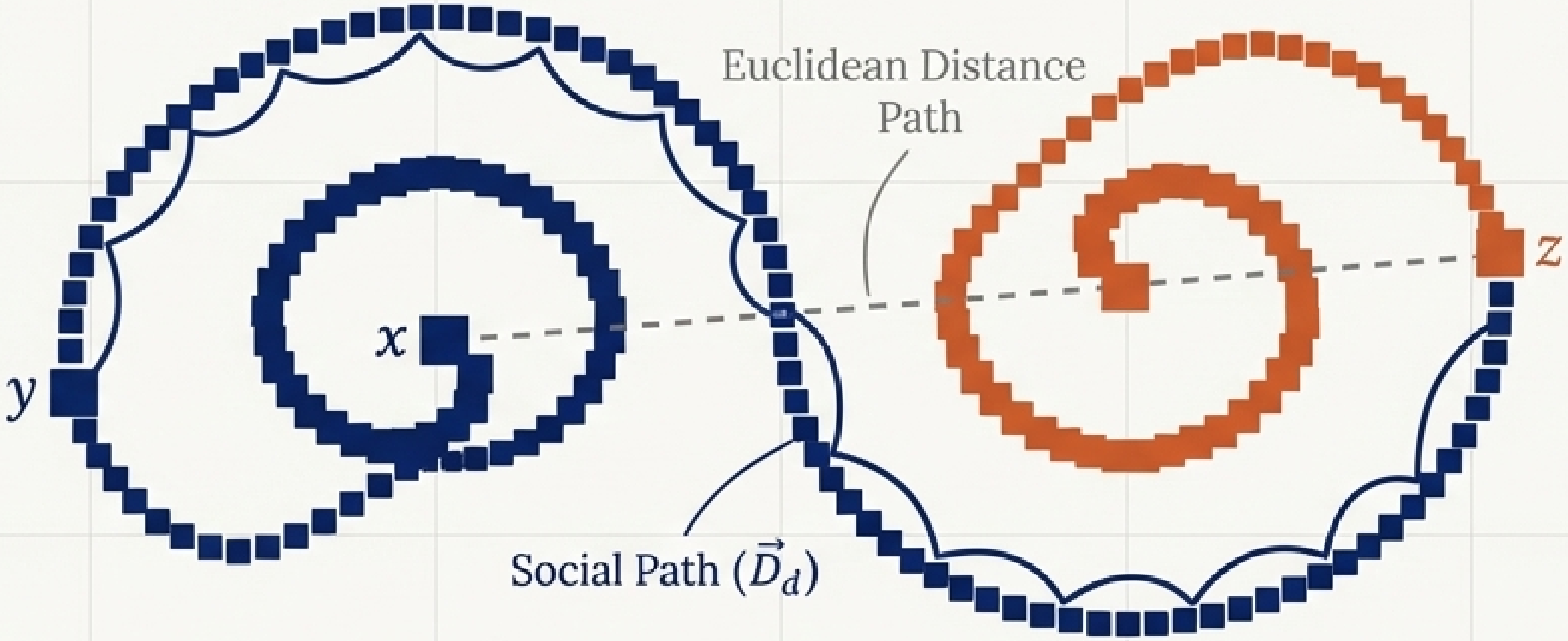
Social Traveler Distance (~D_d)

$M_{x\|y}$	d(x,z) + d(x,x) = 3 + 4 = 7
$M_{y\|x}$	d(y,y) = 4
~D_{1d}(x,y)	≈ 0.855

If d(x,y) changes to 400, ~D_d changes to **≈ 0.9975**.

Variation 2: The "Social Path" Metric for Complex Clustering

For discovering clusters with complex, non-convex shapes, we can combine Social Distance with a Minimax Path approach. The distance between two points (x, y) is the “cheapest” path between them, where the cost of a path is the maximum single “hop” length along it. By using D_d as the “hop” metric, we get the Social Path Metric $(\vec{D}_d)$. This allows us to find paths that stay within dense regions.



In the example, Euclidean distance would connect x to z. The Social Path Metric correctly finds that the “cheapest” social path connects x to y, keeping the path within the main cluster. $\vec{D}_d(x,y) < \vec{D}_d(x,z)$

NotebookLM

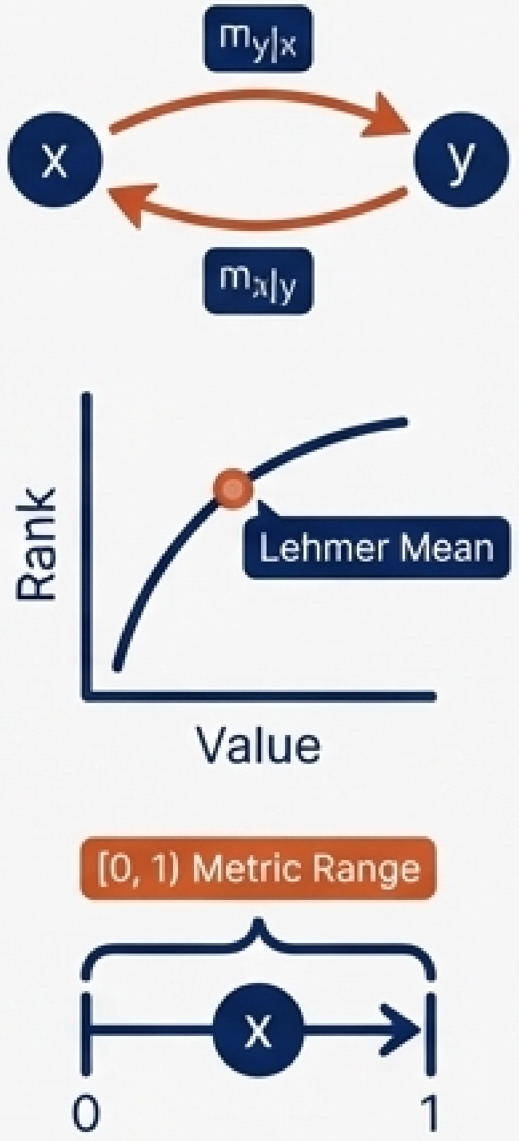
The Social Distance Metric: A Summary

Shift from absolute coordinates to relative, neighborhood-aware relationships.

Key Components

1. **Mutual Social Ranking**: Quantifies the relationship between two points from each one's perspective ($m_{x|y}$, $m_{y|x}$).
2. **Lehmer Mean Aggregation**: Averages mutual ranks with a contraharmonic effect that penalizes asymmetry.
3. **Normalization**: Ensures the final result is a true, well-behaved metric in the [0, 1) range.

Visual Metaphors



Key Benefits

- **Context-Aware**: Incorporates local data density and structure into the distance calculation.
- **Versatile**: Can be applied on top of any existing metric (d).
- **Powerful**: Improves performance in tasks like k-NN classification, graph analysis, and density-based clustering.

The true distance between two points is not just their separation in space, but the social landscape that lies between them.

The Social Distance Metric provides a formal language to measure this landscape, offering a more nuanced and powerful understanding of relationships within data.

Vagan Terziyan (2017) Social Distance metric: from coordinates to neighborhoods, International Journal of Geographical Information Science, 31:12, 2401-2426, DOI: 10.1080/13658816.2017.13677