

CNP (Conditional Neighbor Probability) Formula

Complete Formula

$$CNP(i \rightarrow j) = \frac{\exp\left(-\frac{d(i,j)^2}{\sigma^2}\right)}{\sum_k \exp\left(-\frac{d(i,k)^2}{\sigma^2}\right)}$$

Where:

- i = Query provider (row index)
 - j = Candidate provider (column index)
 - $d(i, j)$ = Euclidean distance between provider i and j
 - σ = Bandwidth parameter (controls decay rate)
 - \sum_k = Sum over all providers k in the dataset
-

Step-by-Step Components

Step 1: Euclidean Distance

$$d(i, j) = \|embedding_i - embedding_j\|_2 = \sqrt{\sum_m (embedding_i[m] - embedding_j[m])^2}$$

Step 2: Gaussian Kernel

$$K(i, j) = \exp\left(-\frac{d(i, j)^2}{\sigma^2}\right)$$

Step 3: Row Normalization

$$CNP(i \rightarrow j) = \frac{K(i, j)}{\sum_k K(i, k)}$$

Matrix Form

$$D = \text{distance matrix } (N \times N)$$

$$K = \exp\left(-\frac{D^2}{\sigma^2}\right)$$

$$CNP[i, j] = \frac{K[i, j]}{\sum_k K[i, k]}$$

Property: Each row sums to 1: $\sum_j CNP[i, j] = 1$

Interpretation

$$CNP(i \rightarrow j) = P(\text{choosing } j \mid \text{starting from } i)$$

"If I'm at provider i and randomly walk to a neighbor weighted by similarity, what's the probability I land on provider j ?"

Key Property: Asymmetry

$$CNP(i \rightarrow j) \neq CNP(j \rightarrow i)$$

Example:

- Specialist A close to only 5 providers
 $\Rightarrow CNP(A \rightarrow B) = \frac{1}{5} = 0.20$ (if B is one of them)
- Generalist B close to 500 providers
 $\Rightarrow CNP(B \rightarrow A) = \frac{1}{500} = 0.002$ (A is just one of many)

This asymmetry captures "breadth" - generalists have broader coverage.