The Geometry of Interaction:
Proximity Graphs in Modern Network Biology

Expert Insight in Computational Graph Theory

January 19, 2026

Abstract

While biological networks are often visualized as topological "hairballs," the underlying logic is increasingly geometric. This blog explores the transition from abstract Protein-Protein Interaction (PPI) networks to Proximity Graphs. We define the formal frameworks of k-NN, RNG, and NSG, revealing how biological systems like contact maps and spatiotemporal expression are, at their core, proximity-based manifolds.

1 Introduction: The Geometric Turn

In traditional network biology, edges are often treated as binary experimental observations. However, as we move into the era of high-throughput single-cell sequencing and structural proteomics, we are no longer just observing connections; we are calculating them.

The move from "Interaction Networks" to "Proximity Graphs" represents a shift from observing a handshake to measuring the distance between two people in a room. In graph theory, a **Proximity Graph** $G = (V, E)$ is a graph where an edge (u, v) exists if and only if the distance $\delta(u, v)$ satisfies a specific geometric constraint relative to the rest of the set S.

2 Formal Definitions and Taxonomy

To understand how these graphs function in biology, we must first define the fundamental structures.

> **Definition 1: The Euclidean Distance (l_2 norm)**
> Given two biological entities p and q represented as vectors in d-dimensional space E^d (where d might be 978 landmark genes in LINCS), the distance is defined as:
>
> $$\delta(p, q) = \|p - q\|_2 = \sqrt{\sum_{i=1}^{d} (p_i - q_i)^2}$$

2.1 1. The k-Nearest Neighbor Graph (k-NNG)

The k-NNG is the most ubiquitous proximity graph in bioinformatics, used heavily in UMAP and t-SNE visualizations.

- **Rule:** An edge exists from u to v if v is among the k closest points to u.
- **Biological Use:** Single-cell RNA-seq clustering. We assume cells with the most similar transcriptomic profiles belong to the same lineage.

2.2 2. The Relative Neighborhood Graph (RNG)

RNG is a sparser, more elegant structure that removes redundant edges.

- **Rule:** An edge (u, v) exists if there is no third point w such that:

$$\delta(u, v) > \max(\delta(u, w), \delta(v, w))$$

- **Biological Use:** Protein folding contact maps. RNG identifies "essential" contacts where two residues are close, and no other residue is positioned between them to block the interaction.

3 Hidden Proximity Graphs in Biology

We often use these structures without formally labeling them. Here are the "hidden" proximity graphs currently driving discovery:

3.1 Contact Maps as Geometric Constraints

In structural biology, a Protein Contact Map is effectively a proximity graph. By applying a threshold τ (usually 8 Angstroms), we create a **Unit Disk Graph**:

$$E = \{(u, v) \mid \delta(u, v) < \tau\}$$

This graph dictates the physical stability of the protein fold.

3.2 Spatiotemporal Proximity

In spatial transcriptomics, we are building proximity graphs where the dimensions E^d are literal X, Y, Z coordinates in a tissue slice combined with time T.

- **The Logic:** Two cells are connected if they are physically adjacent and their expression profiles are co-varying.
- **NSG Application:** The Navigating Spreading-out Graph (NSG) is used here to ensure that "long-range" signaling (hormones/distal communication) can be searched as efficiently as "local" paracrine signaling.

4 Why the NSG is the Future of Network Biology

The primary challenge in modern PPI networks is the "Hairball" problem. High-dimensional data ($d > 100$) suffers from the curse of dimensionality.

> **The NSG Diversification Rule**
> To prevent redundant edges that slow down search (Algorithm 1), the NSG ensures that for any node p and its neighbors n_1, n_2, the following condition is checked:
>
> $$\delta(n_1, n_2) > \delta(p, n_2)$$
>
> If this is false, the edge to n_2 is pruned because n_1 already provides a better path to that "region" of the biological space.

By using NSG, we can perform **Complex Detection** by finding "cliques" in the proximity graph that represent stable molecular machines. This is significantly more robust than searching raw PPI data because the geometric constraints of the NSG filter out the false-positive "noise" typical of yeast two-hybrid or Mass Spec experiments.

5 Conclusion

The transition from topology to geometry is not just a mathematical preference; it is a biological necessity. Whether we are mapping the 1.3 million profiles of LINCS or the atomic distances in a complex, we are navigating a proximity graph. By understanding the rules of these graphs, we can build more efficient, accurate, and scalable models of life.