Incompressibility Barriers to Neighborhood-Preserving Data Visualizations

A Preprint

Szymon Snoeck
Applied Mathematics Department
Columbia University
New York, NY 10027
sgs2179@columbia.edu

Noah Bergam
Computer Science Department
Columbia University
New York, NY 10027
njb2154@columbia.edu

Nakul Verma
Computer Science Department
Columbia University
New York, NY 10027
verma@cs.columbia.edu

February 26, 2025

Abstract

To what extent is it possible to visualize the neighborhood structure of high-dimensional point clouds in two- or three-dimensional space? We reframe this question in terms of embedding n-vertex graphs (representing the neighborhood structure of the input) into metric spaces of low doubling dimension d, in such a way that maintains the separation between neighbors and non-neighbors. We show preserving almost any n-vertex graph in this sense requires $d = \Theta(\log n)$. A similar bound holds for almost any k-regular graph for constant k. The landscape changes dramatically when we require that a single algorithm preserve *all* n-vertex graphs or *all* k-regular graphs in a normed space: the former requires $d = \Theta(n)$ (an exponential jump!) while the latter requires $d = \Omega(k \log n)$. Finally, we show that a point clouds sampled from a distribution over $\mathbb{R}^d$ with a continuous density function will with high probability require $\Theta(d)$ dimensions. Overall, these results challenge the aspiration that constant-dimensional visualizations can faithfully preserve neighborhood structure.