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

April 17, 2025

ABSTRACT

To what extent is it possible to visualize high-dimensional point clouds in a two- or three-dimensional space? We reframe this question in terms of embedding n-vertex graphs (representing the neighborhood structure of the input points) into metric spaces of low doubling dimension d, in such a way that maintains the separation between neighbors and non-neighbors. We show that neighborhood preservation of almost any n-vertex graph requires, $d = \Theta(\log n)$. Even if one only considers k-regular graphs, $d = \Omega(\log n / \log\log n)$ is needed (for constant values of k). The landscape changes dramatically when embedding into normed spaces, which are more restrictive yet more desirable to practitioners. In particular n-vertex graphs require $d = \Theta(n)$ (an exponential jump!) while almost every k-regular graphs requires $d = \Omega(k \log n)$. Finally, we study the implications of these results for various types of practical data models featuring intrinsic cluster or manifold structure. Overall, these results challenge the aspiration that constant-dimensional visualizations can faithfully preserve neighborhood structure.