1163-68-446 **Manuel E Lladser*** (manuel.lladser@colorado.edu), Department of Applied Mathematics, University of Colorado, Boulder, CO 80309-0526. *Low-dimensional representation of genomic sequences.*

Numerous data analysis and data mining techniques require that data be embedded in a Euclidean space. When faced with symbolic datasets, particularly biological sequence data produced by high-throughput sequencing assays, conventional embedding approaches like binary and k-mer count vectors may be too high dimensional or coarse-grained to learn from the data effectively. Other representation techniques such as Multidimensional Scaling (MDS) and Node2Vec may be inadequate for large datasets as they require recomputing the full embedding from scratch when faced with new, unclassified data. To overcome these issues we amend the graph-theoretic notion of "metric dimension" to that of "multilateration." Much like trilateration can be used to represent points in the Euclidean plane by their distances to three non-colinear points, multilateration allows us to represent any node in a graph by its distances to a subset of nodes. Specializing to Hamming graphs, which are particularly well suited to representing biological sequences, we can readily generate low-dimensional embeddings to map sequences of arbitrary length to a real space. This work is in collaboration with Richard C. Tillquist, and has been partially funded by the NSF grant 1836914. (Received September 07, 2020)