Realizing Minimum Spanning Trees from Random Embeddings

Ion Mandoiu [*], Saad Quader [†], and Alexander Russell [‡]

Department of Computer Science & Engineering, University of Connecticut, USA.

March 3, 2018

Abstract

Let $T = (V, E)$ be an undirected tree with n vertices. For any arbitrary $x, y \in \mathbb{R}$, let $f : V \to \{x, y\}^d$ be a random embedding of the tree-vertices where each $f(v)$ is selected independently and uniformly at random. We study the event that there exist nonnegative weights $w_1, \ldots, w_d$ so that T is "realized" by this embedding as the unique minimum spanning tree of the points $f(V)$ under the scaled ℓ_2 metric $\|x\|^2 = \sum w_i x_i^2$. The realization occurs in the following sense: under this metric, the distance between two embedded vertices will be smaller than a threshold if and only if these vertices are neighbors in T. We wish to bound the dimensionality d for which it is possible to realize T with high probability.

We show that any tree can be realized with high probability when $d = \Omega(n \log n)$. The proof gives rise to a simple algorithm that needs only select $w_i \in \{0, 1\}$ and works for both ℓ_2 and ℓ_1 metrics. We additionally study the case for general undirected graphs. We show two sufficient conditions in this case: we show that $d = \Omega(na^2 \log n)$ is sufficient to realize any graph with high probability where a is the arboricity of that graph, and that $d = \Omega(nr^{-2} \log n)$ is also sufficient where r is the smallest effective resistance of the edges in the graph. The former bound becomes $d = \Omega(n|E| \log n)$ in the worst case. We also show that $d = \Omega(n^2)$ and $d = \Omega(n)$ are necessary to realize an Erdős-Rényi random graph and a random n-vertex tree, respectively. We develop a probabilistic analog of Radon's theorem on convex sets, which may be of independent interest.

Variants of this natural "realizability problem" play a basic role in statistical inference of gene expression data, where the existence of such a scaled metric is taken as evidence for the relevance of the expression data to the biological dynamics modeled by the tree.

[*] Email: ion@engr.uconn.edu

[†] Corresponding author. Email: saad.quader@uconn.edu

[‡] Email: acr@cse.uconn.edu