Reducing Overhead for Quantum Advantage in Topological Data Analysis

Dominic W. Berry[a], Yuan Su[b], Casper Gyurik[c], Joao Basso[b], Alexander D.T. Barba[b], Abhishek Rajput[d], Vedran Djunko[c], Nathan Wiebe[d], and Ryan Babbush[b]

[a]*School of Mathematical and Physical Sciences, Macquarie University, NSW 2109, Australia*

[b]*Google Quantum AI, Venice, CA 90291, United States*

[c]*Department of Computer Science, 2300 RA Leiden, The Netherlands*

[d]*Department of Computer Science, University of Toronto, ON M5S 2E4, Canada*

Suppose we are given a data set represented as an unordered set of points in some metric space, such as $\mathbb{R}^n$ with the Euclidean distance function. One of the goals in data analysis is to extract features of this data set and use them to cluster or classify the data. One approach for doing so is to convert the point cloud into a graph where the vertices are the given data points and the edges are determined by whether or not pairs of points lie within a chosen distance ε.

This approach can capture features such as connectivity but ignores potential higher dimensional features, especially if the data points are sampled from some underlying high dimensional manifold. Topological data analysis attempts to extract such higher dimensional global topological features of an underlying data set by applying techniques from the field of algebraic topology, in particular what is known as simplicial homology. An important feature to describe the data is the Betti number β_k, which is the number of k-dimensional holes of the complex. Classically, computing the Betti number is exponentially hard for large k, but quantum algorithms offer the promise of large speedups.

Lloyd *et al.* [1] were first to demonstrate the promise of quantum algorithms for computing Betti numbers. However, concrete estimates of the requirements for quantum advantage in this context have proved elusive. Here, we propose, analyse, and optimise a new quantum algorithm for topological data analysis with reduced scaling, including a method for preparing Dicke states, a more efficient amplitude estimation algorithm using Kaiser windows, and an optimal implementation of eigenvalue projectors based on Chebyshev polynomials. We compile our approach into a fault-tolerant gate set and obtain resource estimates for applying it to real data sets, clarifying conditions on the data that would enable quantum advantage. Finally, we discuss and analyse the viability of several specific applications in neuroscience and financial market modeling.

[1] Seth Lloyd, Silvano Garnerone, and Paolo Zanardi, "Quantum algorithms for topological and geometric analysis of data," Nature Communications **7**, 10138 (2016).