Complexity-Theoretic Limitations on Quantum Algorithms for Topological Data Analysis

Alexander Schmidhuber[1,2,*] and Seth Lloyd[3,4]

[1]*Institute for Theoretical Physics, ETH Zurich, Zurich 8092, Switzerland*

[2]*Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*

[3]*Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*

[4]*Turing Inc., Cambridge, Massachusetts 02139, USA*

Quantum algorithms for topological data analysis (TDA) seem to provide an exponential advantage over the best classical approach while remaining immune to dequantization procedures and the data-loading problem. In this paper, we give complexity-theoretic evidence that the central task of TDA—estimating Betti numbers—is intractable even for quantum computers. Specifically, we prove that the problem of computing Betti numbers exactly is #P-hard, while the problem of approximating Betti numbers up to multiplicative error is NP-hard. Moreover, both problems retain their hardness if restricted to the regime where quantum algorithms for TDA perform best. Because quantum computers are not expected to solve #P-hard or NP-hard problems in subexponential time, our results imply that quantum algorithms for TDA offer only a polynomial advantage in the worst case. We support our claim by showing that the seminal quantum algorithm for TDA developed by Lloyd, Garnerone, and Zanardi achieves a quadratic speedup over the best-known classical approach in asymptotically almost all cases. Finally, we argue that an exponential quantum advantage can be recovered if the input data is given as a specification of simplices rather than as a list of vertices and edges.

I. INTRODUCTION

One of the major challenges in quantum information science is the development of quantum algorithms with an exponential advantage over the best classical counterpart. Some of the most promising candidates in that regard have been quantum machine-learning (QML) algorithms [1–4]. However, two issues have challenged the prospect of exponential speedups for many QML algorithms in recent years. On the one hand, a series of impressive dequantization results by Tang *et al.* [5,6] showed that several QML problems based on low-rank matrix analysis can be solved efficiently on a classical computer. On the other hand, most remaining QML algorithms require the encoding of a large input into a quantum state [7], which requires large quantum random access memory (QRAM).

These issues have resulted in increased efforts to find quantum algorithms that are immune to dequantization procedures and do not suffer from the data-loading problem. One prominent example is the quantum algorithm for topological data analysis (TDA) developed by Lloyd, Garnerone, and Zanardi (LGZ) [8], which has experienced a recent surge of attention from a theoretical [9–15] as well as experimental [16,17] perspective. This quantum algorithm, which we will refer to as the LGZ algorithm, does not require a large classical input and is immune to standard dequantization procedures. Nevertheless, it solves the task of estimating the Betti numbers of a simplicial complex—the essential task of TDA—and does so seemingly exponentially faster than the best classical algorithm, provided that certain assumptions regarding the input hold.

In this paper, we show that any quantum algorithm for TDA runs in exponential time in the worst case by proving that (under widely believed complexity-theoretic assumptions) the problem of topological data analysis is intractable even for quantum computers. Specifically, we prove that the problem of computing Betti numbers exactly is #P-hard and that the problem of approximating Betti numbers up to multiplicative error is NP-hard. Moreover, we show that both problems retain their hardness if the

_____

*alexsc@mit.edu

input is restricted to clique-dense complexes, which is the regime where the LGZ algorithm works best [8,12,13]. Building upon this result, we argue that the LGZ algorithm for TDA runs in exponential time not only in the worst case, but for asymptotically almost all inputs. We provide direct evidence for this claim by investigating the runtime of the LGZ algorithm on random Vietoris-Rips complexes, for which we show that it achieves a Grover-like speedup over the best-known classical algorithm in almost all cases. Our complexity-theoretic results and average-case analysis do not rule out an exponential quantum advantage over the best classical algorithm for some specific inputs. Indeed, there exists a known family of simplicial complexes for which the LGZ-algorithm achieves a superpolynomial speedup [14].

We further provide a complexity-theoretic analysis of each step of the LGZ algorithm, through which we show that the subroutine limiting quantum advantage in TDA is not actually the computation of Betti numbers given a description of the simplicial complex, but the construction of a quantum state representing the simplicial complex from a description of the underlying graph. We show that the latter problem is #P-hard by itself. It is thus the problem of sampling simplices given a list of vertices and edges that forms the bottleneck of topological data analysis. Provided with an oracle that enables random sampling from the k-simplices of a simplicial complex, a modification of the LGZ algorithm can estimate its Betti numbers in polynomial time, exponentially faster than the best-known classical approach with access to the same oracle. We study problems in data analysis and computational topology where such oracles appear naturally. In particular, when the data is given in the form of a list or specification of sets (simplices) and their members (vertices), an exponential quantum advantage could be recovered. This is the case, for example, when we are given the ability to sample from a list of Facebook groups together with the members of the sampled groups, and when a k-simplex is defined to be a group of $k + 1$ individuals all of whom are members of the same group.

The remaining article is structured as follows. In Sec. II, we introduce the framework of topological data analysis and describe the original LGZ Betti number quantum algorithm developed by Lloyd, Garnerone, and Zanardi, along with a brief summary of recent improvements. In Sec. III, we study the following formal problem of computing Betti numbers.

*Problem 1 (**Betti**).*—Given a clique complex S defined by its vertices and edges and an integer $k \geq 0$ as input, output the kth Betti number β_k of S.

The problem is defined exactly in the way that it would appear in practical applications of TDA, where the vertices represent a set of data points embedded in a metric space, and the edges describe whether two points are close to each other or not. The complexity of computing Betti

numbers was left as an open problem in the review article [18]. We show that the problem **Betti** is intractable for quantum computers even if restricted to the optimal regime of clique-dense complexes, by establishing its hardness via the following theorem.

Theorem 1.—The problem ***Betti*** is *(a)* #P-hard and *(b)* remains #P-hard when the input is restricted to clique-dense complexes.

However, the LGZ algorithm does not compute Betti numbers exactly but only approximately. Theorem 1 is therefore not enough to reason about the runtime of the LGZ algorithm. Indeed, there are examples of polynomial quantum algorithms that efficiently approximate #P-hard problems, such as the Jones polynomial [19] or the Potts model [20]. In Sec. IV, we show that even approximating Betti numbers up to any multiplicative error remains intractable for quantum computers (again, under the widely held complexity-theoretic assumption that quantum computers cannot access NP-hard problems). To do so, we will study the problem of deciding which Betti numbers are nonzero.

*Problem 2 (**Homology**).*—Given a clique complex S defined by its vertices and edges and an integer $k \geq 0$ as input, output *true* if $\beta_k > 0$ and *false* if $\beta_k = 0$.

Problem **Homology** is the natural decision version of the counting problem **Betti**. It was already shown by Adamaszek and Stacho [21] that this problem is NP-hard for complexes defined on co-chordal graphs. Here, we strengthen this result by showing that it remains NP-hard on clique-dense complexes.

Theorem 2.—The problem ***Homology*** is *(a)* NP-hard and *(b)* remains NP-hard when the input is restricted to clique-dense complexes.

Evidently, Theorem 2 directly implies that computing any multiplicative approximation of Betti numbers is NP-hard and therefore not accessible to quantum computers. Because Theorems 1 and 2 hold even if restricted to the regime where the currently known quantum algorithms for TDA perform best—and because the relative size of this regime compared to all possible simplicial complexes tends towards 0 as the number of vertices increases—our results imply that the LGZ algorithm runs in exponential time for almost all cases. In Sec. V, we verify this claim by investigating the runtime of the LGZ algorithm on random Vietoris-Rips complexes. We show that it achieves a Grover-like quadratic speedup over classical algorithms in asymptotically almost cases. However, up to superpolynomial speedups can be achieved for specifically selected inputs.

In Sec. VI, we subsequently discuss problems beyond TDA for which a modification of the LGZ algorithm achieves an exponential speedup. The main assumption enabling exponential quantum advantage is the access to random sampling from the simplices of a simplicial complex. We list examples of problems where such an oracle

might appear naturally. Finally, in Sec. VII, we comment on the computational hardness of computing normalized Betti numbers, which is a quantity that is not directly useful in practice but a more natural output of quantum algorithms for TDA.

Remark.—Recently, simultaneous independent work by Crichigno and Kohler [22] appeared on the arXiv. The authors prove that the problem **Homology** is QMA1-hard by investigating supersymmetric quantum systems. This result strengthens Theorem 2 stated here and yields an alternative proof of Theorem 1.

II. TOPOLOGICAL DATA ANALYSIS AND THE LGZ ALGORITHM

Topological data analysis is a recent approach to the analysis of large datasets that are high-dimensional, noisy or incomplete. The goal of TDA is to describe the shape of a dataset by extracting robust features—topological invariants—which are inherently insensitive to local noise. These topological invariants are the Betti numbers β_k, which for every k count the number of k-dimensional holes in the dataset. While the theory behind TDA is based on algebraic topology, it can be understood through simple linear-algebraic terms. This section gives a concise description of the TDA pipeline and its two main procedures: representing a dataset as a simplicical complex and subsequently computing its shape by estimating the Betti numbers of the complex.

A. Topological data analysis

The starting point of TDA is as follows: a dataset of interest is represented by a set of n points $\{x_i\}_{i=1}^n$ (a point cloud) embedded in a metric space. For simplicity, we will study the example where x_i are embedded in the real vector space $\mathbb{R}^d$ for some positive integer $d > 0$, but the formalism extends to arbitrary manifolds. A point cloud is not connected and consequently has a trivial topology. To study the shape of the data set, the point cloud is thus first

promoted to a topological object—a so-called simplicial complex—which can be interpreted as the generalization of a graph. This is done in two steps. First, a *grouping scale* $\epsilon > 0$ is chosen and every two points x_i, x_j for which

$$||x_i - x_j|| < \epsilon \qquad (1)$$

are connected. The point cloud and resulting graph G are depicted in Fig. 1. Because a graph is an object with at most one-dimensional building blocks (edges), it does not yet allow the extraction of higher-dimensional features. To remedy this, the graph G is promoted to a simplicial clique complex by identifying $(k + 1)$ cliques in G with k *simplices*. A k-simplex s_k of G is thus a fully connected k-dimensional subgraph of G, i.e., a set [23] of $k + 1$ points $\{v_0, \ldots, v_k\} \subset \{x_i\}_{i=1}^n$ that are pairwise connected to each other. By this definition, a 0-simplex is a point, a 1-simplex a line, a 2-simplex a triangle, a 3-simplex a tetrahedron and so on. Simplices are visualized in Fig. 2(a). The set of all possible simplices that can be constructed on G is called the *simplicial complex S* of the dataset. This particular type of simplicial complex—constructed by identifying cliques of G with simplices—is also referred to as *clique complex* of G, denoted Cl(G). Furthermore, the clique complex of a graph obtained by connecting pairwise close points embedded in a metric space is called the *Vietoris-Rips* complex.

Note that S is closed under inclusion: if a k-simplex $s_k = \{v_0, \ldots, v_k\}$ is in the simplicial complex, the $(k - 1)$-simplex $s_{k-1}(j)$ obtained by omitting the vertex v_j from s_k is also an element of the complex. These are the boundary simplices of s_k. Concretely, a line is bounded by two points, a triangle by three lines, and so on.

In anticipation of an quantum algorithm for TDA, and to avoid repetition at a later stage, we will write simplices as quantum states of a n-qubit system. Because each k-simplex $s_k = \{v_0, \ldots, v_k\}$ defines a subset of $\{x_i\}_{i=1}^n$, it can be naturally identified with a computational basis state $|s_k\rangle$ of Hamming weight $k + 1$ on n qubits, where the jth
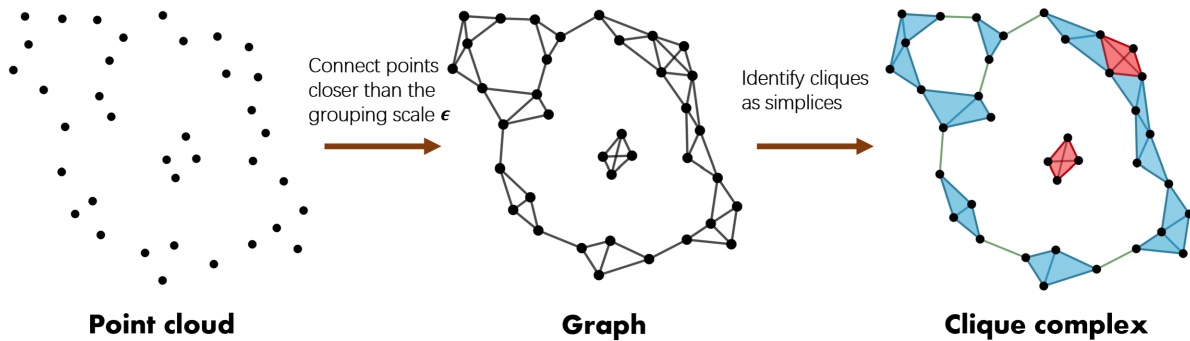

FIG. 1. The construction of a simplicial complex from a point cloud. First, a grouping scale ϵ is chosen and every two data points with pairwise distance smaller than ϵ are connected. The resulting graph is subsequently promoted to a (simplicial) clique complex by identifying $k + 1$ cliques with k-simplices.
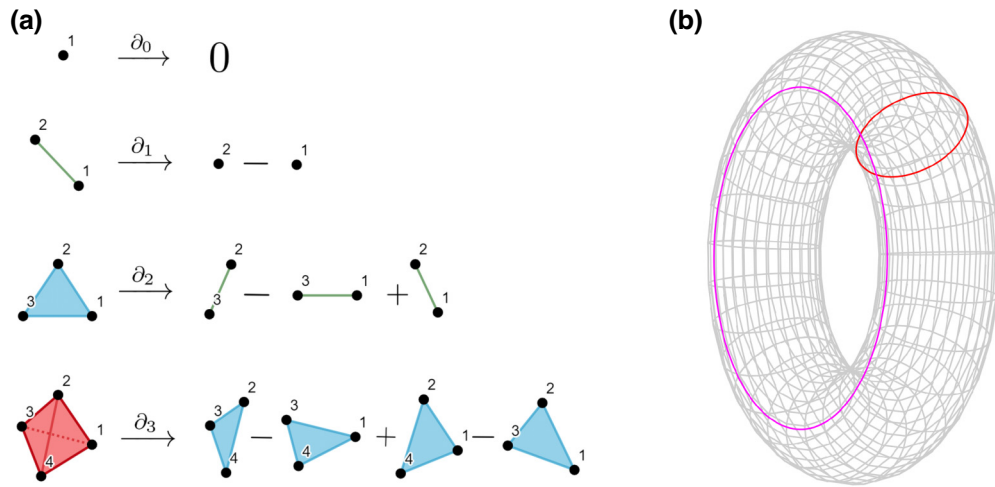
FIG. 2.   (a) A visualization of the simplicial boundary operator acting on a 0-simplex (point), 1-simplex (line), 2-simplex (triangle), and 3-simplex (tetrahedron). More generally, a k-simplex is mapped to the alternating sum of its $(k+1)$ boundary simplices. (b) A two-dimensional torus embedded in $\mathbb{R}^3$. The torus has one connected component, two distinct holes (red and purple circle) and a single void (the interior). Its first three Betti numbers are thus $(\beta_0, \beta_1, \beta_2) = (1, 2, 1)$ and all higher orders vanish.

bit of $|s_k\rangle$ is one if s_k contains the vertex x_j, and zero else. If we denote by $\mathcal{H}_k$ the Hilbert space spanned by the computational basis states of Hamming weight $k+1$, we have $|s_k\rangle \in \mathcal{H}_k$ for all k and $\oplus_{k=-1}^{n-1} \mathcal{H}_k = \mathbb{C}^{2^n}$. It will become convenient later to further define the set S_k of all k-simplices in the simplicial complex S, and the Hilbert space $\mathcal{H}_k(S) \subset \mathcal{H}_k$ spanned by $\{|s_k\rangle \mid s_k \in S_k\}$.

B. Simplicial homology

Having constructed the simplicial complex S as a representation of the dataset $\{x_i\}_{i=1}^n$, TDA aims to describe the shape of S. From a topological standpoint, this shape is described by the *Betti numbers* $\beta_0, \ldots \beta_n$ of S, which intuitively count the number of k-dimensional holes in the simplicial complex. We will first state their technical definition: Betti numbers are defined as the ranks of the homology groups induced by the boundary map $\partial_k : \mathcal{H}_k(S) \rightarrow \mathcal{H}_{k-1}(S)$, which is a combinatorial operator

$$\partial_k |s_k\rangle = \sum_{i=0}^{k} (-1)^i |s_{k-1}(i)\rangle , \qquad (2)$$

mapping each k-simplex to the alternating sum of its boundary simplices. The action of ∂ is visualized in Fig. 2(a). As before, $|s_{k-1}(i)\rangle$ is the $(k-1)$-simplex obtained by setting the ith entry of the bitstring $|s_k\rangle$ to zero. Note that the boundary of a boundary is zero,

$$\partial_{k-1}\partial_k = 0. \qquad (3)$$

Thus ∂_k induces a (finite) chain complex on S,

$$0 \xrightarrow{\partial_n} \mathcal{H}_{n-1}(S) \xrightarrow{\partial_{n-1}} \ldots \xrightarrow{\partial_1} \mathcal{H}_0(S) \xrightarrow{\partial_0} 0, \qquad (4)$$

with the corresponding kth *homology group* $\mathbf{H_k}$ defined as the quotient group of the kernel of ∂_k over the image of ∂_{k+1}:

$$\mathbf{H_k} = \text{Ker}\partial_k \big/ \text{Im}\,\partial_{k+1} . \qquad (5)$$

The kth Betti number $\beta_k = \dim \boldsymbol{H_k}$ equals the dimension of the kth homology group [24]. Intuitively, the kth Betti number counts the number of k-simplices in the complex that are boundaryless but not themselves a boundary, which can be interpreted as the number of k-dimensional holes. Thus β_0 counts the number of connected components, β_1 the number of holes, β_2 the number of voids, and so on. The Betti numbers of a torus are visualized in Fig. 2(b). As for any chain complex that satisfies Eq. (3), Hodge theory provides a relationship between the homology group H_k and the *Hodge Laplacian* $\Delta_k := \partial_k^*\partial_k + \partial_{k+1}\partial_{k+1}^*$:

$$\boldsymbol{H_k} \cong \text{Ker}\,\Delta_k = \text{Ker}\,(\partial_k^*\partial_k + \partial_{k+1}\partial_{k+1}^*). \qquad (6)$$

In the special case presented here, this Hodge relation can be understood intuitively as follows: both $\partial_k^*\partial_k$ and $\partial_{k+1}\partial_{k+1}^*$ are positive Hermitian operators, thus any state $|\psi\rangle$ annihilated by the Hodge Laplacian lies in the intersection of the kernels of ∂_k and ∂_{k+1}^*. The first operator enforces that $|\psi\rangle$ has no boundary, and the second operator guarantees that only one representative per equivalence class is counted. Two cycles are considered equivalent if they can be continuously deformed to each other within the simplicial complex, i.e., if they differ by a boundary. Equation (6) reduces the computation of β_k to the linear-algebra problem of constructing and estimating the rank of a $\binom{n}{k+1} \times \binom{n}{k+1}$ matrix. In the next section, we describe

a quantum algorithm for precisely this task, developed by Lloyd, Garnerone, and Zanardi.

C. The LGZ algorithm

The LGZ quantum algorithm [8] estimates the Betti numbers of a simplicial complex through repeated quantum phase estimation. Before providing a detailed description, we summarize its three main steps: First, create the mixed density state

$$\rho_k = \frac{1}{|S_k|} \sum_{s_k \in S_k} |s_k\rangle \langle s_k| \qquad (7)$$

of all k-simplices in the simplicial complex S. Here $|S_k|$ is the number of k-simplices of S. Secondly, exponentiate the Hodge Laplacian Δ,

$$\Delta = \bigoplus_{k=0}^{n-1} \Delta_k = \begin{pmatrix} \Delta_0 & 0 & \cdots & 0 \\ 0 & \Delta_1 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & \Delta_{n-1} \end{pmatrix} \qquad (8)$$

and perform quantum phase estimation of $e^{i\Delta}$ with the eigenvector register starting in ρ_k. This yields random sampling from the eigenvalue distribution of Δ_k. In a third step, steps one and two are repeated multiple times. The relative frequency of the zero-eigenvalue gives an estimate of the Kernel dimension of Δ_k, and thus of the kth Betti number:

$$\beta_k = \dim \boldsymbol{H_k} = \dim \operatorname{Ker} \Delta_k. \qquad (9)$$

In the following, each step is explained in detail.

(1) Projection onto k-simplices. A variety of approaches can be used to arrive at the state ρ_k. LGZ construct it via Grover search applied to the membership function f_k, with

$$f_k(s) = \begin{cases} 1 & s \in S_k \\ 0 & \text{else} \end{cases}. \qquad (10)$$

Given access to the pairwise distances between points x_i, f_k can be implemented in $\mathcal{O}(k^2)$ steps. The multisolution version of Grover's algorithm then allows us to construct

$$|S_k\rangle = \frac{1}{|S_k|^{1/2}} \sum_{s_k \in S_k} |s_k\rangle, \qquad (11)$$

while simultaneously revealing $|S_k|$, in $\mathcal{O}\left(\zeta_k^{-1/2}\right)$ calls to the function f. Here,

$$\zeta_k = \frac{|S_k|}{\binom{n}{k+1}} \qquad (12)$$

is the fraction of occupied k-simplices of the simplicial complex S. The mixed state

$$\rho_k = \frac{1}{|S_k|} \sum_{s_k \in S_k} |s_k\rangle \langle s_k|. \qquad (13)$$

is then constructed by including an additional n-qubit ancilla system, applying a CNOT gate to each qubit in $|S_k\rangle$ and subsequently tracing out (ignoring) the ancilla.

(2) Quantum phase estimation. The second step is to perform quantum phase estimation (QPE) with the operator $e^{i\Delta}$ on the state ρ_k. While directly exponentiating the Hodge Laplacian Δ might be costly in general, it is the square of a sparse operator B defined as

$$B = \begin{pmatrix} 0 & \partial_1 & 0 & \cdots & \cdots & 0 \\ \partial_1^\dagger & 0 & \partial_2 & & & \vdots \\ 0 & \partial_2^\dagger & 0 & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & 0 & \partial_{n-1} \\ 0 & \cdots & \cdots & \cdots & \partial_{n-1}^\dagger & 0 \end{pmatrix}. \qquad (14)$$

It is straightforward to verify that $B^*B = B^2 = \Delta$, and thus $\operatorname{Ker} B = \operatorname{Ker} \Delta$. Because B is the square root of a Laplacian, it is often referred to as *Dirac operator*. Furthermore, because ∂_k is a $\binom{n}{k} \times \binom{n}{k+1}$ matrix with $n - k$ nonzero entries in each row and $k + 1$ nonzero entries in each column, B is n-sparse. Thus standard protocols allow the Hamiltonian simulation of e^{iB} using $\mathcal{O}(n^3)$ gates. To get accurate results from QPE, the spectrum of B has to be considered: denote the smallest nonzero eigenvalue and the largest eigenvalue of Δ by $\lambda_{\min}$ and $\lambda_{\max}$, respectively. To avoid multiples of 2π in the exponent of e^{iB}, the Dirac operator has to be rescaled by $\lambda_{\max}^{-1}$. Therefore, the phase estimation has to be executed with precision at least $\kappa^{-1} = \lambda_{\min}/\lambda_{\max}$ to resolve whether an eigenvalue is zero or not, which yields a total cost of $\mathcal{O}(n^3\kappa)$ for the QPE subroutine.

(3) Kernel estimation.
Repeating the above procedure M times results in M samples $\{\lambda_i\}_i$ from the eigenvalue distribution of Δ_k. The relative frequency of the eigenvalue zero gives an additive estimate of the *normalized* Betti number c_k,

$$c_k = \frac{\beta_k}{|S_k|} = \frac{\dim \boldsymbol{H_k}}{|S_k|} \approx \frac{|\{i : \lambda_i = 0\}|}{M}. \qquad (15)$$

The Hoeffding inequality shows that $M = \mathcal{O}(\delta^{-2})$ samples suffice to estimate c_k up to additive error of

ε. Combined with the state preparation and quantum phase estimation, the LGZ algorithm computes an ε-additive approximation of the normalized kth Betti number in time

$$\mathcal{O}\left(\frac{n^3\kappa + nk^2\zeta_k^{-1/2}}{\varepsilon^2}\right). \tag{16}$$

This closely resembles the original runtime analysis of the LGZ algorithm carried out in Ref. [8], however, there ε was mistakenly labeled a multiplicative error. We will argue below that it is indeed preferable to state the runtime in terms of a multiplicative approximation error, as originally intended.

Note that this quantum algorithm not only estimates the k-th Betti number, but also supplies the corresponding eigenvectors, which are the harmonic representatives of the homology [25].

D. Analysis

The best-known classical algorithm computes the kth Betti number β_k in time $\mathcal{O}\left(\binom{n}{k+1}\right)$, where $0 \le k \le n-1$. At first glance, Eq. (16) thus seems to provide an exponential quantum advantage if k scales with n. Two issues might reduce this advantage [12,13]. The first concerns the fraction $\kappa = \lambda_{\max}/\lambda_{\min}$. While $\lambda_{\max}$ is bounded from above by the Gershgorin circle theorem [13], there are no known lower bounds on the smallest eigenvalue $\lambda_{\min}$ of the simplicial Laplacian Δ_k. In some cases $\lambda_{\min}$ is exponentially small, resulting in an exponential runtime of the LGZ algorithm. Classes of simplicical Laplacians that have a minimum eigenvalue scaling at least inverse polynomially in the number of vertices have been identified in, e.g., Refs. [9,12]. The second, more pressing issue concerns the fraction

$$\zeta_k = \frac{|S_k|}{\binom{n}{k+1}}. \tag{17}$$

If the number of occupied k-simplices $|S_k|$ is small compared to $\binom{n}{k+1}$, the runtime of the quantum algorithm will again provide only a quadratic speedup over the best classical algorithm. A necessary condition for a polynomial runtime is thus that the simplicial complex is *clique-dense* [8,9,12], i.e., that

$$\zeta_k^{-1} = \binom{n}{k+1}/|S_k| \in \mathcal{O}(\text{poly}(n)). \tag{18}$$

Note that the restrictiveness of the requirement Eq. (18) depends entirely on how k scales with the number of vertices n. For constant k, every graph fulfills Eq. (18) trivially. If k grows slowly with n, such as $k = \Theta(\log(n))$, the

fraction ζ_k^{-1} will still not grow superpolynomially in the generic case, such as the random Erdös-Renyi model [26]. The regime where classical algorithms require exponential time, and thus where an exponential quantum advantage is possible, is when k grows asymptotically as $k = \Theta(n)$. This is also the regime where Eq. (18) becomes restrictive, the regime originally considered in Ref. [8], and the regime where standard complexity-theoretic arguments apply. Throughout this paper, we will thus study the setting where $k = \Theta(n)$ scales linearly with n (potentially up to logarithmic factors), unless specifically noted otherwise. In this case, the relative size of the clique-dense regime for random Vietoris-Rips complexes as defined in Sec. V tends asymptotically towards 0.

Several classes of clique-dense complexes have been summarized in Ref. [12]. If both the clique-density requirement and the inverse-polynomial eigenvalue threshold are fulfilled, the algorithm can estimate the normalized Betti number up to inverse-polynomial additive error ε in polynomial time. In this work, we give complexity-theoretic and analytical evidence that even in this optimal regime, the additive accuracy achievable by quantum algorithms for TDA is not sufficient to solve the problem **Homology**, e.g., to decide whether the Betti numbers are zero or nonzero.

What practical information does an additive approximation to Betti numbers contain? We first note practical applications of TDA [27,28] require the actual Betti numbers β_k. More importantly, the requirement that the simplicial complex is clique-dense precisely means that the denominator of ζ_k, $|S_k|$, scales exponentially with n. Unless the Betti number β_k scales similarly, the normalized Betti number c_k will be exponentially small. In that case, any inverse-polynomial additive error ε cannot distinguish it from zero.

We will show in Sec. V that Betti numbers typically do not scale exponentially in the number of vertices. Indeed, they are asymptotically almost always bounded by $\mathcal{O}(n)$, independent of k. The LGZ algorithm thus requires exponential time to distinguish the Betti numbers from zero, which could be seen as a consequence of Theorem 2.

Because of the above discussion, we argue that it can be misleading to state the runtime in terms of the normalized [29] Betti numbers c_k. Instead, a more insightful description of the algorithm is the runtime required to compute the actual Betti number β_k up to multiplicative error δ. This can be achieved naively by choosing $\varepsilon = \delta(\beta_k/|S_k|)$, resulting in

$$\mathcal{O}\left(\frac{|S_k|^2}{\beta_k^2}\frac{(n^3\kappa + nk^2\zeta_k^{-1/2})}{\delta^2}\right). \tag{19}$$

A series of recent improvements to the original LGZ algorithm, most notably by utilizing quantum counting

[8,13] and an efficient construction of the boundary operator [9] via fermionic annihilation operators, yield an algorithm that naturally outputs a multiplicative approximation with a substantially improved runtime of

$$\mathcal{O}\left(\frac{1}{\delta}\left(n^2\sqrt{\frac{\binom{n}{k+1}}{\beta_k}} + n\kappa\sqrt{\frac{|S_k|}{\beta_k}}\right)\right). \quad (20)$$

We note that further improvements to the complexity Eq. (20) might be possible by, e.g., combining the results of Refs. [9–15]. However, it is to the best of our knowledge not possible to prepare the state ρ_k significantly faster than $\Omega(\zeta_k^{-1/2})$ or estimate the Betti number β_k given ρ_k significantly faster than $\Omega\left(\sqrt{|S_k|/\beta_k}\right)$, hence the total runtime is lower bounded by

$$T_q = \Omega\left(\sqrt{\frac{\binom{n}{k+1}}{\beta_k}}\right). \quad (21)$$

While this might suggest a polynomial runtime for large Betti numbers, we will show in Sec. V that this expression is exponential in n for asymptotically almost all inputs, in line with Theorems 1 and 2. We now prove both theorems.

III. COMPUTING BETTI NUMBERS IS #P-HARD

In this section, we prove the complexity-theoretic hardness of computing Betti numbers exactly. Let us recall the problem definition and theorem from the first section.

*Problem 3 (**Betti**).*—Given a clique complex S defined by its vertices and edges and an integer $k \geq 0$ as input, output the kth Betti number β_k of S.

Theorem 3.—The problem ***Betti*** is *(a)* #P-hard and *(b)* remains #P-hard when the input is restricted to clique-dense complexes.

Informally, the complexity class NP is the set of decision problems for which a solution can be verified in polynomial time. The complexity class #P is the set of counting problems associated to NP. That is, whereas a problem in NP asks whether a given instance has a solution or not, the corresponding problem in #P asks how many solutions the instance has. The archetypal #P-complete problem is #SAT, the problem of counting the number of satisfying truth assignments to a given SAT instance (more details on SAT can be found after Theorem 3). Clearly, solving a problem in #P is at least as hard as solving the corresponding problem in NP, as computing the number of solutions in particular determines whether this number is nonzero. On the other hand, there are problems where the decision version is easy, i.e., in P, but the counting version is #P-complete. The most prominent such example is the matrix permanent [30]. No known quantum algorithm solves an NP-complete problem in polynomial time and

it is conjectured that NP $\not\subset$ BQP, i.e., that NP-hard problems are not accessible to quantum computers. The same holds for #P-hard problems. It is worth noting that it is not known whether the NP-hardness of a problem implies the #P-hardness of the corresponding counting problem [31].

The main idea of our proof of Theorem 1 is to relate the computation of Betti numbers to the computation of the number of maximal cliques, which is a known #P-hard problem. To do so, we introduce the *Euler characteristic* $\chi(S)$ of a simplicial complex S defined on n vertices as

$$\chi(S) = \sum_{k=0}^{n-1}(-1)^k|S_k|, \quad (22)$$

where as before $|S_k|$ is the number of k-simplices in S. The index k is bounded by $n-1$ because the largest possible simplices on n vertices are $(n-1)$-simplices. The Euler characteristic χ allows us to relate the task of computing β_k to the task of computing $|S_k|$ via the Euler-Poincaré formula [32].

Theorem 4 (Euler-Poincaré).—The Euler characteristic of a simplicial complex S has an alternative representation as

$$\chi(S) = \sum_{k=0}^{n-1}(-1)^k\beta_k,$$

where β_k are the Betti numbers of S.

Proof.—By the Rank-nullity theorem, we have $|S_k| = \dim \text{Im } \partial_k + \dim \text{Ker } \partial_k$ and by the definition of the Betti number $\beta_k = \dim \mathcal{H}_k = \dim \text{Ker } \partial_k - \dim \text{Im } \partial_{k+1}$. Thus

$$\sum_{k=0}^{n-1}(-1)^k(|S_k| - \beta_k) = \sum_{k=0}^{n-1}(-1)^k(\dim \text{Im } \partial_k$$
$$+ \dim \text{Im } \partial_{k+1}) = 0,$$

where we have used $\text{Im } \partial_0 = \text{Im } \partial_n = 0$. ∎

The computation of the Euler characteristic thus reduces to the computation of the Betti number vector. In the following, we show that #SAT, the problem of counting the number of solutions to a given SAT problem, reduces to the computation of the Euler characteristic of a Vietoris-Rips complex defined by its vertices and edges. Recall that a SAT (or *boolean satisfiability problem*) instance with variables $X_1, \ldots, X_n$ and clauses $C_1, \ldots, C_k$ in conjunctive normal form (CNF) is a boolean formula $C_1 \wedge C_2 \wedge \cdots \wedge C_k$ where every clause C_j is a disjunction of literals. A literal is either a variable X_i or its negation $\neg X_i$. The hardness of computing the Euler characteristic of an abstract simplicial complex defined by its inclusion-maximal simplices has already been established in Ref. [33]. However, this proof does not directly apply here, because the complexes considered in Ref. [33] are not clique complexes and a much more extensive description of the complex is given

as input. Indeed, translating between both descriptions of a clique complex, i.e., computing the list of inclusion-maximal simplices from a list of edges and vertices, is by itself a #P-hard problem [34]. Our techniques are inspired by a proof of the hardness of computing independent sets [33,35].

Proof of Theorem 1(a).—Let $\mathcal{K}$ be a *SAT* instance in conjunctive normal form with variables $X_1, \ldots, X_n$ and clauses $C_1, \ldots, C_k$. We will construct a clique complex S on $3n + k$ vertices such that the number of satisfying truth assignments of $\mathcal{K}$ equals $(-1)^n(1 - \chi(S))$. ∎

To construct S we first build up the underlying graph G as follows: for every variable X_i, construct three vertices t_i, f_i and p_i. It will become clear later that t_i and f_i encode whether X_i is *true* or *false*, and p_i whether X_i appears [36] in $\mathcal{K}$. Connect every two points v_i, v_j for which $i \neq j$. Here, $v \in \{t, f, p\}$ is a placeholder. Consequently, add an additional vertex c_j for every clause C_j. Connect c_j to every vertex ***except*** to

(a) t_i if the clause C_j contains the literal X_i.
(b) f_i if the clause C_j contains the literal $\neg X_i$.

In Fig. 3 we illustrate the graph G corresponding to the 2-SAT instance $\mathcal{K} = (X_1 \vee \neg X_2) \wedge (\neg X_1 \vee X_2)$. The vertex set V of G is

$$V = \{t_i, f_i, p_i \mid 1 \leq i \leq n\} \cup \{c_j \mid 1 \leq j \leq k\}. \quad (23)$$

We can now build a clique complex $S = \mathrm{Cl}(G)$ on G in the usual way by identifying $(k + 1)$-cliques with k-simplices. It remains to show that the Euler characteristic of $\mathrm{Cl}(G)$ counts the number of solutions to $\mathcal{K}$. To do so, we introduce the notion of a *maximal clique,* which is a clique $s \subset S$ that is not contained in any other clique of S. Maximal cliques should not be confused with maximum cliques, which are cliques of maximal cardinality. While a maximum clique is always maximal, the converse must not hold. For the clique complex shown in Fig. 3, the maximal cliques are

$$\{c_1, p_1, p_2, c_2\}$$
$$\{c_1, p_1, t_2\}, \{c_1, p_2, f_1\}, \{c_1, f_1, t_2\},$$
$$\{c_2, p_2, t_1\}, \{c_2, p_1, f_2\}, \{c_2, f_2, f_1\}$$
$$\{t_1, t_2\}, \{f_1, f_2\}.$$

Let A be the set of all vertices labeled either t_i or f_i for any i and let B be the set of all vertices labeled either c_i or p_i. By construction, $A \cup B = V$. Let further $\Gamma(S)$ be the set of all maximal cliques of S that contain only vertices of A. In the above example, $\Gamma(S) = \{\{t_1, t_2\}, \{f_1, f_2\}\}$ We will show the following two results:

(i) The elements of $\Gamma(S)$ are in a one-to-one correspondence with the solutions of the *SAT* instance $\mathcal{K}$.
(ii) All remaining cliques $s \in S \backslash \Gamma(S)$ contribute trivially to the Euler characteristic, that is

$$\chi(S \backslash \Gamma(S)) := \sum_{s \in S \backslash \Gamma(S)} (-1)^{|s|-1} = 1.$$

As a consequence of (i) and (ii), the Euler characteristic of S equals

$$\chi(S) = \sum_{s \in S} (-1)^{|s|-1}$$
$$= (-1)^{n-1}(\textbf{number of solutions to } \mathcal{K}) + 1, \quad (24)$$

which proves Theorem 1. It remains to show (i) and (ii).

Proof of (i).—Let $s \in \Gamma(S)$. Suppose there exists a fixed i such that s contains neither t_i nor f_i. Then s can be extended to the strictly larger clique $s \cup \{p_i\}$, since p_i is connected to all other t_j, f_j for $j \neq i$. This contradicts that s is maximal. Likewise, s can not contain both t_i and f_i for any fixed i, as they are not connected in S. Therefore, any $s \in \Gamma(S)$ has cardinality n and corresponds to an assignment of the variables $X_1, \ldots, X_n$. Consider now a clause c_j. By the maximality of s, there is at least one vertex of s that is not connected to c_j and therefore the truth assignment corresponding to s fulfills the clause c_j. Since c_j was arbitrary, s satisfies $\mathcal{K}$. The same reasoning shows that any $z \in S$ corresponding to a satisfying truth assignment of $\mathcal{K}$ is maximal and contained in Γ. ∎

Proof of (ii).—Let $\widetilde{L} \subset S$ be the set of all cliques of S that contain only vertices of A and consider $L = \widetilde{L} \backslash \Gamma(S)$. For each $v \in L$, let us define the set

$$D_v = \{k \in S \mid v \subset k \text{ and } k \backslash v \subset B\} \quad (25)$$

of all extensions of v to B. Note that $\{D_v\}_{v \in L}$ together with $D_\emptyset = \{k \in S | k \subset B\}$ is a complete partition of $S \backslash \Gamma(S)$, that is, every $s \in S \backslash \Gamma(S)$ is contained in exactly one D_v. Let v be a clique contained in L. Thus v is not maximal and by (i) we can find a $p_j \in B$ such that $v \cup \{p_j\}$ is in D_v. Since all elements of B are pairwise connected, there is a bijection between the sets $E_v = \{k \in D_v \mid p_j \notin k\}$ and $F_v = \{k \in D_v \mid p_j \in k\}$:

$$E_v \rightarrow F_v, \quad k \mapsto k \cup \{p_j\}. \quad (26)$$

Because $E_v \cup F_v = D_v$ by definition, the Euler characteristic of D_v is $\chi(D_v) = \chi(E_v) + \chi(F_v) = 0$. Using once again that all elements of B are pairwise connected and the fact that the clique complex of a complete graph has Euler
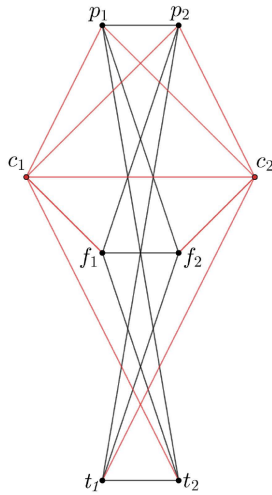
FIG. 3. The graph G corresponding to the *SAT* instance $\mathcal{K} = (X_1 \vee \neg X_2) \wedge (\neg X_1 \vee X_2)$. The edges connected to the two clauses c_1 and c_2 are colored in red for clarity. The Vietoris-Rips complex on G has 8 vertices, 18 edges, 10 triangles (2-simplices) and 1 tetrahedron (3-simplex). Its Euler characteristic is thus $\chi(S) = 8 - 18 + 10 - 1 = -1$ and the number of satisfying truth assignments $(-1)^2(1 - \chi(S)) = 2$.

characteristic 1, we get $\chi(D_\emptyset) = 1$ and

$$\sum_{s \in S \setminus \Gamma(S)} (-1)^{|s|-1} = \chi(D_\emptyset) + \sum_{v \in L} \chi(D_v) = 1, \quad (27)$$

which proves (ii). Because the elements of $\Gamma(S)$ have cardinality n, they contribute with a factor $(-1)^{n-1}$ to $\chi(S)$. This leads to expression Eq. (24) and concludes the proof. ∎

Before proving Theorem 1(b), we mention two alternative approaches to arrive at the #P-hardness of computing Betti numbers exactly. First, note that the only cliques contributing to the Euler characteristic of the above graph G are the cliques of size n. Building upon this idea, one can show that computing the nth Betti number of $\mathrm{Cl}(G)$ is already #P-hard, as opposed to computing all Betti numbers. Because we do not require this stronger statement, we omit the details. Second, one could also arrive at Theorems 1(a) and 1(b) building upon a result [37] that computing the Witten index of a certain local supersymmetric system is #P-hard. This is because the Witten index can be related to the independence polynomial (which is directly related to the Euler characteristic) of a clique complex via the so-called Fermionic hard-core model.

We can strengthen Theorem 1(a) by restricting the class of simplicial complexes allowed as input to the problem **Betti**. As argued in Sec. II, the LGZ quantum algorithm works best in the regime of clique-dense complexes. We now show that computing Betti numbers remains hard even in this regime.

Proof of Theorem 1(b).—In the proof of Theorem 1(a) we reduced an arbitrary #SAT instance to the computation of the Euler characteristic of the clique complex of a corresponding graph G. We will show that $\mathrm{Cl}(G)$ is already clique-dense, which evidently proves Theorem 1(b). Recall that for an instance $\mathcal{K}$ with n variables and s clauses, G is defined on $N = 3n + s$ vertices. If $\mathcal{K}$ belongs to ℓ-SAT, i.e., every clause $C \in \mathcal{K}$ has ℓ literals, the number of edges of G is

$$|E(G)| = \frac{9}{2}n(n-1) + 3ns - ls + \frac{1}{2}s(s-1). \quad (28)$$

Because SAT reduces to 3-SAT, we can fix $\ell = 3$. Moreover, as 3-SAT remains NP-hard [38] when restricted to instances where every variable appears 4 times [39,40], we can fix $s = \frac{4}{3}n$. The fraction γ of connected edges over the number of vertices squared is then

$$\gamma = \frac{|E(G)|}{N^2} = \frac{1}{2} - \frac{9n + 7s}{2N^2}$$
$$= \frac{1}{2}\left(1 - \frac{165}{169n}\right) > \frac{1}{2}\left(1 - \frac{1}{n}\right). \quad (29)$$

The clique-density theorem [41] states that the number of k-simplices in G is lower bounded by $\Omega(N^{k+1})$ if

$$\gamma > \frac{1}{2}\left(1 - \frac{1}{k}\right). \quad (30)$$

Thus the clique-density condition Eq. (18) is fulfilled for any $k \leq n$, in particular for $k = \Theta(n)$, and $\mathrm{Cl}(G)$ is clique-dense. ∎

This concludes the proof of Theorem 1. We emphasize that it is not clear whether either **Betti** or **Homology** are in #P or NP, respectively, which would make them #P- or NP-complete. The nontriviality of the kth Betti number could be witnessed by a cycle that is not a boundary, but such a cycle may be exponentially large in the size of the input.

IV. APPROXIMATING BETTI NUMBERS IS NP-HARD

In the previous section, we established that computing Betti numbers exactly is #P-hard. However, the LGZ algorithm only approximates Betti numbers. To be able to argue about the runtime of quantum algorithms for TDA, we need to determine the complexity of computing a multiplicative approximation to β_k. This motivates us to define the following problem, which we recall along with Theorem 2 from Sec. I.

*Problem 4 (**Homology**).*—Given a clique complex S defined by its vertices and edges and an integer $k \geq 0$ as input, output *true* if $\beta_k > 0$ and *false* if $\beta_k = 0$.

Theorem 5.—The problem **Homology** is *(a)* NP-hard and *(b)* remains NP-hard when the input is restricted to clique-dense complexes.

Theorem 2(a) is due to Adamaszek and Stacho [21], who proved that the decision problem **Homology** is NP-hard for the clique complexes of co-chordal graphs. Since the class of all graphs in particular includes the class of co-chordal graphs, the theorem follows. This by itself however is not enough to reason about the runtime of the LGZ algorithm, as co-chordal graphs tend to be sparse, whereas it has already been established that a necessary requirement for a polynomial runtime of the LGZ algorithm is that the input is clique-dense. To strengthen the results of Adamaszek and Stacho, we show that the problem **Homology** retains its hardness when restricted to the class of clique-dense complexes.

Proof of Theorem 2(b).—Let $G = (V, E)$ be any co-chordal graph with vertex set $V = \{x_1, \ldots, x_n\}$ and edges E. Building upon constructions in Refs. [21,42], we will relate the homology of $\mathrm{Cl}(G)$ to the homology of the complement $\bar{H}$ of a bipartite graph H defined via the Alexander dual of G. We will then show that $\mathrm{Cl}(\bar{H})$ is clique-dense, which reduces the homology problem of co-chordal graphs to the homology problem of clique-dense complexes and proves the theorem.

Denote by $\mathrm{Cl}(G)$ the clique complex of the co-chordal graph G. Consider the complex A_G defined on the same vertex set V consisting of simplices s with

$$s \in A_G \iff (V \backslash s) \notin \mathrm{Cl}(G). \qquad (31)$$

A_G is called the Alexander dual of G and consists of the complements of all noncliques in $\mathrm{Cl}(G)$. The homologies of A_G and $\mathrm{Cl}(G)$ are related via the Alexander duality [43]

$$H_k(\mathrm{Cl}(G)) = H^k(\mathrm{Cl}(G)) = H_{n-k-3}(A_G), \qquad (32)$$

where the first equality holds due to Corollary 2.2 in Ref. [21]. A_G is not a clique complex and thus not the type of complex that appears in TDA. However, we can relate the homology of A_G to the homology of the clique complex of a bipartite graph H, as follows. Denote by $\mathcal{M}(A_G)$ the set of inclusion-maximal faces of A_G. By construction,

$$\mathcal{M}(A_G) = \{V \backslash e \mid e \in V \times V \text{ and } e \notin E\} \qquad (33)$$

is the complement of all edges e that are not in the edge set of G, i.e., that are edges of $\bar{G}$. In particular, $\mathcal{M}(A_G)$ can be computed in polynomial time given G. Let us now construct a bipartite graph H with two parts of size n and m, respectively, where $m = |\mathcal{M}(A_G)|$. Each vertex in the first part corresponds to one of the vertices x_i of G, and each vertex in the second part corresponds to the inclusion-maximal faces $F_j \in \mathcal{M}(A_G)$ of the Alexander dual of G. Connect every pair (x_i, F_j) for which $x_i \notin F_j$. By Eq. (33),

each F_j is only connected to exactly two vertices. H is thus a graph on $n + m$ vertices with $2m$ edges and clearly sparse. By construction, $\mathrm{Cl}(\bar{H})$ is isomorphic to the suspension of A_G and the suspension isomorphism [21,42,44] relates the homology of both complexes via

$$H_k(A_G) = H_{k+1}(\mathrm{Cl}(\bar{H})), \qquad (34)$$

where the complement $\bar{H}$ is now dense. Combining Eqs. (32) and (34) yields

$$H_k(\mathrm{Cl}(G)) = H_{n-k-2}(\mathrm{Cl}(\bar{H})), \qquad (35)$$

which reduces the homology problem of co-chordal graphs G to the homology problem of dense co-bipartite graphs $\bar{H}$. For Eq. (35) to be well defined, we have to pick $0 \le k \le n - 2$, which does not matter for the reduction as the last Betti number β_{n-1} can be computed in linear time.

To conclude the proof, we show that $\mathrm{Cl}(\bar{H})$ obtained via this construction is clique-dense. Let $N = n + m$ be the number of vertices of $\bar{H}$. The fraction γ of connected edges over the number of vertices squared is

$$\gamma = \frac{|E(\bar{H})|}{N^2} = \frac{N(N-1) - 4m}{2N^2},$$
$$= \frac{1}{2}\left(1 - \frac{5N - 4n}{N^2}\right) > \frac{1}{2}\left(1 - \frac{5}{N}\right). \qquad (36)$$

By the clique-density theorem [41], $\mathrm{Cl}(\bar{H})$ contains $\Omega(N^{k+1})$ many k-simplices if

$$\gamma > \frac{1}{2}\left(1 - \frac{1}{k}\right). \qquad (37)$$

Hence $\bar{H}$ fulfills the requirement Eq. (18) for any $k \le N/5$ and in particular, since $n \le N$, for $k = \Theta(n)$. The clique complex $\mathrm{Cl}(\bar{H})$ is thus clique-dense. For completeness, we note that one could similarly show that Eq. (18) is fulfilled for any $k \le n$ by utilizing that the MaxClique problem required to establish NP-hardness for co-chordal graphs [21] remains hard for graphs with a constant fraction of edges, in which case $m = \Theta(n^2)$ and

$$\frac{5N - 4n}{N^2} = o\left(\frac{1}{n}\right). \qquad (38)$$

Having reduced the homology problem of the clique complexes of co-chordal graphs to the homology problem of clique-dense clique complexes, this concludes the proof. ■

Theorems 1 and 2 together describe the computational hardness of topological data analysis. It is widely believed that quantum computers cannot solve NP-hard or #P-hard problems in subexponential time. Our results therefore

imply that any quantum algorithm for TDA, including the LGZ algorithm, runs in exponential time in the worst case. Moreover, our proofs show that even in the optimal clique-dense regime, quantum algorithms for TDA require exponential time. Because the LGZ algorithm performs optimally in the clique-dense regime, and the relative size of this regime tends towards zero, this further implies that the LGZ algorithm runs in exponential time for asymptotically almost all inputs. The same holds for any quantum algorithm for TDA that uses similar techniques (Grover search) as LGZ to arrive at the mixed state ρ_k, which are only efficient in the clique-dense case.

V. TIME COMPLEXITY OF THE LGZ ALGORITHM

Having established the complexity-theoretic hardness of topological data analysis, we turn back to the original LGZ algorithm. We consider here the version of the LGZ algorithm that outputs a multiplicative approximation to β_k, as originally intended [8] and required by practical applications. Our previous results imply that the algorithm will have a runtime exponential in the input size for asymptotically almost all inputs. In this section, we verify this directly by considering random Vietoris-Rips complexes.

Disregarding other details—like potentially exponentially small eigenvalue gaps of the simplicial Laplacian—the optimal runtime of the LGZ algorithm as derived in Eq. (20) will be of order $\text{poly}(n)\xi$, where

$$\xi = \sqrt{\frac{\binom{n}{k+1}}{\beta_k}}. \tag{39}$$

Recall that the best-known classical algorithm for computing β_k takes time $\mathcal{O}\left(\binom{n}{k+1}\right)$. The LGZ algorithm will thus achieve a superquadratic speedup only if the kth Betti number is large. However, as we will discuss in the following, Betti numbers of Vietoris-Rips complexes are small in expectation and variance.

Consider a set of points $\{x_i\}_{i=1}^n$ randomly distributed IID in a metric space of dimension d according to any bounded probability density function. Let us denote by V^ϵ the simplicial complex that results from the TDA procedure outlined in Sec. II. That is, we choose a grouping scale $\epsilon > 0$, connect all (x_i, x_j) that are closer than ϵ and identify cliques in the corresponding graph with simplices. The resulting simplicial complex is called a Vietoris-Rips complex, and together with the assumption that $\{x_i\}_{i=1}^n$ were distributed randomly, a *random Vietoris-Rips* complex. Note that this is the most natural input distribution model for practical applications of topological data analaysis.

The Betti numbers of random Vietoris-Rips complexes were analyzed by Kahle in Refs. [45,46]. Because we are interested in the behavior of Betti numbers as $n \to \infty$, we allow the grouping scale ϵ to vary with n. Let us denote

by $\hat{r} = n^{-1/d}$ the average distance [47] between two vertices of V^ϵ. Kahle identified three regimes for ϵ that fully describe the asymptotic behavior of Betti numbers of V^ϵ.

(a) **Subcritical regime:** The subcritical regime occurs when $\epsilon = o(\hat{r})$ and the Vietoris-Rips complex is sparsely connected. In this regime, the Betti numbers experience a transition from vanishing to nonvanishing. In either case, they are upper bounded by

$$\mathbb{E}[\beta_k] = o(n). \tag{40}$$

(b) **Critical regime:** If $\epsilon = \Theta(\hat{r})$ is in the same order of magnitude as $\hat{r}$, V is said to be in the critical regime. This region is also called the thermodynamic limit and is the regime where Betti numbers take on their maximal value. The expectation value and variance of β_k are

$$\mathbb{E}[\beta_k] = \Theta(n), \quad \text{Var}[\beta_k] = \Theta(n). \tag{41}$$

(c) **Supercritical regime:** Here, $\epsilon = \omega(\hat{r})$ dominates the average separation $\hat{r}$ and large connected components emerge, which do not contribute to the homology of V. Consequently, Betti numbers grow sublinearly and

$$\mathbb{E}[\beta_k] = o(n). \tag{42}$$

In all three regimes, the expectation value of β_k grows at most linearly with n. Moreover, the variance $\text{Var}(\beta_k)$ is also bounded by $\mathcal{O}(n)$. Thus $\beta_k = \mathcal{O}(n)$ asymptotically almost surely (a.a.s.) over random Vietoris-Rips complexes and the LGZ algorithm applied to random Vietoris-Rips complexes almost always achieves only a quadratic speedup over classical approaches.

This average-case analysis does of course not exclude the possibility that specific families of graphs exhibit large enough Betti numbers to provide a much better quantum advantage. Indeed, a recent paper by Google Quantum AI [14] provided an example of a graph with a superpolynomial speedup. It has Betti numbers scaling as $\beta_{\log(n)-1} \sim (n/\log(n))^{\log(n)}$, which is sufficiently large to give a superpolynomial advantage over the best-known classical algorithm. Unlike the present paper, the authors of Ref. [14] study the regime where k scales logarithmic with n. This relaxes the clique-density condition Eq. (18) but removes the possibility for an exponential speedup. To the best of our knowledge, there is currently no known example of a graph that provides an exponential quantum advantage.

Upper bounds on the size of Betti numbers of Vietoris-Rips complexes imply lower bounds on the runtime of the LGZ algorithm. Goff analyses the maximal size of Betti

numbers of Vietoris-Rips complexes in Ref. [48]. For the kth Betti number, he proves an upper bound of $\mathcal{O}(n^k)$, which is slightly more prohibitive than the trivial upper bound $\mathcal{O}(n^{k+1})$. However, the best construction in Ref. [48] achieves only $\beta_k = \mathcal{O}(n^{k/2+1/2})$ (which corresponds to a quartic quantum advantage) and does not saturate this bound. Since the upper bound in Ref. [48] was obtained inductively, it might be loose. Whether or not it can be tightened is an open question that would provide insight into whether any superpolynomial quantum advantage for TDA can be achieved in the regime where k scales as $k = \Theta(n)$.

The above discussion implies that the LGZ algorithm achieves a Grover-like speedup over the best classical approach in almost all cases and up to a superpolynomial speedup in certain specific cases. This agrees with Theorem 2 and the widely believed conjecture that quantum computers cannot solve #P-hard or NP-hard problems in polynomial time. While a quadratic speedup is significant, this suggests that quantum algorithms for topological data analysis might not achieve a practical advantage for the noisy intermediate-scale quantum (NISQ) era and early generation fault-tolerant quantum computers [49]. In the next section, we argue that an exponential speedup can be recovered if computational problems beyond TDA are considered.

VI. QUANTUM ADVANTAGE BEYOND TDA

Up to this point, our complexity-theoretic analysis in Secs. III and IV as well as our direct analysis of the LGZ algorithm in Sec. V have exclusively focused on clique complexes, and even more specifically on Vietoris-Rips complexes. While all simplicial complexes appearing in the typical setting of TDA [50] are of this form, the LGZ algorithm and the ideas behind it can be applied to a much broader class of simplicical complexes. Computing Betti numbers of more general simplicial complexes is an important problem in computational topology [51–54]. In this section, we argue that an exponential quantum advantage can be recovered if certain computational problems beyond TDA are considered. To illustrate this, we emphasize three points:

(a) The limiting factor for quantum advantage in TDA is the construction of (and subsequent random sampling from) the list of k-simplices of a clique complex, given the underlying graph as input. This by itself is an NP-hard problem, thus quantum algorithms solve it only polynomially faster than the best classical algorithm.

(b) Given an oracle that provides random sampling from the k-simplices in an (abstract) simplicial complex, quantum algorithms for TDA can efficiently compute the kth Betti number of the simplicial

complex. Moreover, they can do so sometimes exponentially faster than the best-known classical algorithm with access to the same oracle.

(c) Such an oracle is not natural in TDA. Indeed, constructing such an oracle for a clique complex described by its underlying graph is again NP-hard. However, in some applications of computational topology, it is natural to describe simplicial complexes by a method of specifying the set of simplices. This enables the efficient construction of such an oracle.

Let us illustrate the first point. The main complexity of quantum algorithms for TDA arises from two factors. First, the clique complex S has to be constructed from the input graph. We call this the *construction* step. Taking again $k = \Theta(n)$ and surpressing factors polynomial in n, this step takes time

$$\mathcal{O}(\zeta_k^{-1/2}) = \mathcal{O}\left(\sqrt{\frac{\binom{n}{k+1}}{|S_k|}}\right) \qquad (43)$$

via Grover search, where $|S_k|$ is the number of k-simplices in the complex and n is the number of vertices. The complex construction by itself is a limiting factor in the runtime: efficient sampling from the k-simplices of a clique complex, given a description of the underlying graph, allows us to find a maximum clique of the graph. It is therefore an NP-hard problem, which in particular suggests that quantum computers cannot solve it significantly more efficiently than via Grover search.

The second step, which we call *estimation* step, is to estimate the Betti numbers given the ability to sample from the set of k-simplices, and to determine whether a given k-simplex is in the complex. The estimation step can be done via, e.g., quantum phase estimation and quantum counting, which has complexity proportional to

$$\mathcal{O}\left(\frac{1}{\delta}\sqrt{\frac{|S_k|}{\beta_k}}\right). \qquad (44)$$

Here, δ is the multiplicative error and factors of poly(n) are again suppressed. It is this second step where quantum algorithms for TDA have the potential to provide an exponential advantage over classical algorithms. To our knowledge, the best classical algorithm can compute the Betti numbers of a simplicial complex with inverse-polynomial spectral gap of the Laplacian $\lambda_{max}/\lambda_{min} = \text{poly}(n)$ from a list of simplices in time $\mathcal{O}(|S_k|)$. This is because determining the kernel dimension of sparse matrices scales at best linearly in the matrix dimension [55]. For a detailed discussion of the best classical algorithms for estimating Betti numbers, see, for example, Sec. IV in Ref. [14].

On the other hand, the LGZ algorithm can solve the same task up to constant multiplicative error exponentially faster if $\beta_k \approx |S_k|$. In the following subsection, we discuss regimes where the quantum algorithm solves the second step, *estimation*, efficiently.

A. Complexes beyond Vietoris-Rips

While Vietoris-Rips complexes do not have sufficiently large Betti numbers for a large quantum speedup, other types of complexes do. For Erdös-Renyi complexes, the LGZ algorithm solves the *estimation* step exponentially faster than the best-known classical approach in asymptotically almost all cases. Random Erdös-Renyi complexes are the clique complexes of random Erdös-Renyi graphs $G(n,p)$. The graph $G(n,p)$ on n vertices with probability parameter p is constructed by randomly connecting vertices with probability p, independently from every other edge. Kahle [56] showed that for random Erdös-Renyi complexes, the fraction in Eq. (44) is asymptotically almost always 1. Specifically, for any $0 \leq k < n$ and a probability parameter $n^{-1/k} < p < n^{-1/(k+1)}$, we have

$$\mathbb{E}(\beta_k) \approx \mathbb{E}(|S_k|) \quad \text{and} \quad \beta_k \approx |S_k|, \qquad (45)$$

i.e., the kth Betti number takes on its maximal possible value $|S_k|$ (the number of k-simplices in the complex). Consequently, the estimation step takes time $\mathcal{O}(\text{poly}(n))$ on a quantum computer and time $\mathcal{O}(|S_k|)$ on a classical computer. Since for random Erdös-Renyi complexes, $|S_k| \approx n^{k/2}$, this gives an exponential quantum advantage for the *estimation* step if $k = \Theta(n)$.

Similarly, an exponential quantum advantage may appear if we go beyond clique complexes to abstract simplicial complexes. Abstract simplicial complexes are not defined via the cliques of an underlying graph but via a collection of inclusion-maximal simplices that are closed under subset taking. In general, this collection might be exponentially large and reintroduce the data-loading problem. However, in some typical real-world examples such a description is compact. There are many well-known examples of abstract simplicial complexes with exponentially large Betti numbers. One example is the *k-skeleton* of the n-simplex, which is the simplicial complex on $n+1$ vertices containing all possible simplices of dimension less than or equal to k. These simplicial complexes have close to maximal Betti numbers: The kth Betti number of the k-skeleton of an n-simplex equals $\beta_k = \binom{n}{k+1}$. They are also compactly specified by a list of maximal nonsimplices. For these and similar abstract simplicial complexes with exponentially large Betti numbers, the LGZ algorithm solves the *estimation* step efficiently.

B. Oracle construction

Solving the *estimation* step efficiently enables the efficient computation of Betti numbers only if we have access to random sampling of k-simplices in the complex. While random sampling of simplices is in general a hard problem, we note that datasets that permit random simplicial sampling could be available in real-world applications.

For example, consider the problem of calculating the homology of Facebook, where $k + 1$ Facebook users form a simplex if they are members of the same Facebook group (the same $k + 1$ users can all be members of more than one group). The resulting simplicial complex is not a clique complex. In this setting, random sampling of simplices can be achieved efficiently by simply querying Facebook for a random element of their stored list of groups, assuming classical access to this list. If we moreover assume coherent [57] access to a membership function f_k (see Sec. II C), a quantum algorithm can efficiently solve the *estimation* step by following the steps of the LGZ algorithm: randomly sample a k-simplex from the complex, apply quantum phase estimation to it with the exponentiated simplicial Laplacian and record the relative frequency of the zero eigenvalue using quantum counting. In this model, a quantum algorithm can thus compute the simplicial homology of Facebook efficiently, provided the Betti numbers of Facebook are comparable to the number of simplices. At the same time, even with access to efficient simplex sampling, the best classical approach seems to require the computation of the nullity of a exponentially large matrix, which takes exponential time [55].

In general, if we are given a list of sets and their members, and define a k-simplex to be $k + 1$ points all of which lie in the same set, this describes an abstract simplicial complex from which we can efficiently sample simplices. If we further assume coherent access to a membership function for the simplices, a modification of the LGZ algorithm supplies an exponential advantage over classical algorithms for computing homology when the kth Betti number is large and comparable to the number of k-simplices, and the spectral gap is scaling inverse polynomially.

VII. COMPLEXITY OF ESTIMATING NORMALIZED BETTI NUMBERS

Motivated by practical applications of TDA, the previous chapters have focused on computing Betti numbers exactly or up to multiplicative error. From a quantum algorithms' perspective, a more natural output is an additive approximations to the *normalized kth Betti number*

$$c_k = \frac{\beta_k}{|S_k|}, \qquad (46)$$

i.e., to the kth Betti number divided by the number of k-simplices in the simplicial complex. Indeed, this is the original output of the LGZ-algorithm [8], although there mistakenly labeled a multiplicative approximation. Later work [13,14] focuses on multiplicative approximations. As

we discussed in Secs. II and V, in almost all cases where c_k can be estimated in polynomial time, the normalized Betti number is exponentially small. Any inverse polynomial additive error ε can therefore not distinguish it from zero, in agreement with the hardness of **Homology**.

Nevertheless, the complexity of estimating normalized Betti numbers (BNE) is a highly interesting topic offering a potentially exponential quantum advantage. In a recent paper [12], Dunjko *et al.* proved that a generalization of this task, called low-lying spectral density estimation (LLSD), is DQC1-hard. In another recent work, Cade *et al.* [58] proved that estimating normalized Betti numbers for arbitrary chain complexes (as opposed to clique complexes) is also DQC1-hard. The complexity class DQC1 (or deterministic quantum computation with one clean qubit) is the class of decision problems solvable by a one clean qubit machine in polynomial time with low error probability (e.g., at most 1/3 for all instances). A *one clean qubit machine* is a n-qubit system, where the first qubit starts out in the pure state $|0\rangle$, while all other $n-1$ qubits start out in the maximally mixed state. The initial density matrix of the one clean qubit machine is thus

$$\rho_{\mathrm{DQC1}} = |0\rangle \langle 0| \otimes \frac{1}{2^{n-1}} \mathbb{1}_{n-1}, \qquad (47)$$

where $\mathbb{1}_{n-1}$ is the identity matrix on $n-1$ qubits. We are then allowed to apply any polynomial-sized quantum circuit to ρ and subsequently measure the first qubit.

The results of Refs. [12,58] imply a relationship between the problem of estimating normalized Betti numbers and other well-known DQC1-complete problems, such as the *power-of-one-qubit* task of estimating normalized traces of unitaries [59] and the approximation of the Jones polynomial at a fifth root of unity for the trace closure, due to Shor and Jordan [60]. Theorems 1 and 2 further show that both the Betti number problem and the Jones polynomial problem are #P-hard in the exact case and that additive approximations to both problems are not value distinguishing [61]. It is widely believed that the one clean qubit model can not be simulated by a classical computer in polynomial time.

How do the DQC1-completeness results of Refs. [12,58] relate to Theorems 1 and 2? As emphasized by the respective authors, it is not clear whether either result actually applies to quantum algorithms for TDA, as they both concern (possibly strong) generalizations of the central task of TDA, which is estimating Betti numbers of Vietoris-Rips complexes. Let us first discuss the potential complexity-theoretic gap between BNE and LLSD, as defined in Ref. [12], Sec. 3A. BNE solves LLSD only if the Hermitian matrix that is used as input for LLSD is the simplicial Laplacian Δ_k of a simplicial complex. Whether LLSD remains DQC1-hard if restricted to simplicial Laplacians remains an open problem. Similarly, the results of Ref.

[58] apply to more general chain complexes, and it is not clear whether the DQC1-hardness remains if the input is restricted to clique complexes. For arbitrary chain complexes, the Betti number estimation problem is trivially NP-hard.

If BNE were shown to be DQC1-hard, this would greatly strengthen the claim that the version of the LGZ-algorithm targeting an additive approximation to the normalized Betti number provides an exponential quantum advantage. Strengthening the above hardness results is thus an exciting direction for further research. Nevertheless, it would not contradict Theorems 1 and 2, since a computational advantage for additively estimating normalized Betti numbers does not necessarily translate to a computational advantage for multiplicatively estimating Betti numbers.

VIII. DISCUSSION

In this paper, we investigated the computational complexity of computing Betti numbers in topological data analysis. We proved that the problem of computing Betti numbers of clique complexes exactly is #P-hard, while the problem of estimating them up to any multiplicative error is NP-hard. Moreover, we show that both problems retain their hardness if the input is restricted to clique-dense complexes, a regime of measure zero in which the LGZ-algorithm performs best. Because quantum computers are not expected to be able to solve #P-hard or NP-hard problems in subexponential time, our results imply that quantum algorithms for TDA run in exponential time in the worst case. Contrary to previous claims, this holds even in the optimal regime of clique-dense complexes. We verify our results through a detailed analysis of random Vietoris-Rips complexes, from which we conclude that the LGZ-algorithm achieves a quadratic speedup over the best classical algorithm for asymptotically almost all inputs.

Our results show that the bottleneck of quantum algorithms for TDA is the construction of a clique complex from a description of its underlying graph. If the simplicial complex is instead specified in a way that allows random sampling from the simplices in the complex, a modification of quantum algorithms for TDA can compute Betti numbers efficiently if certain conditions are met. Under the same conditions, classical algorithms for estimating Betti numbers still require a runtime exponential in the number of data points. In this case, an exponential quantum advantage is recovered. We note that data composed of lists of sets and their members takes the form required to retain an exponential quantum advantage, and provide examples where such input models appear naturally.

———————————

[1] Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd, Quantum machine learning, Nature **549**, 195 (2017).

[2] Aram W. Harrow, Avinatan Hassidim, and Seth Lloyd, Quantum algorithm for linear systems of equations, Phys. Rev. Lett. **103**, 150502 (2009).

[3] Maria Schuld, Ilya Sinayskiy, and Francesco Petruccione, An introduction to quantum machine learning, Contemp. Phys. **56**, 172 (2015).

[4] Maria Schuld and Nathan Killoran, Quantum machine learning in feature Hilbert spaces, Phys. Rev. Lett. **122**, 040504 (2019).

[5] Ewin Tang, in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, Phoenix, AZ, USA* (Association for Computing Machinery, New York, NY, USA, 2019), p. 217.

[6] Nai-Hui Chia, András Gilyén, Tongyang Li, Han-Hsuan Lin, Ewin Tang, and Chunhao Wang, in *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, Chicago, IL, USA* (Association for Computing Machinery, New York, NY, USA, 2020), p. 387.

[7] Scott Aaronson, Read the fine print, Nat. Phys. **11**, 291 (2015).

[8] Seth Lloyd, Silvano Garnerone, and Paolo Zanardi, Quantum algorithms for topological and geometric analysis of data, Nat. Commun. **7**, 1 (2016).

[9] Shashanka Ubaru, Ismail Yunus Akhalwaya, Mark S. Squillante, Kenneth L. Clarkson, and Lior Horesh, Quantum topological data analysis with linear depth and exponential speedup, Preprint ArXiv:2108.02811 (2021).

[10] Iordanis Kerenidis and Anupam Prakash, Quantum machine learning with subspace states, Preprint ArXiv: 2202.00054 (2022).

[11] Ryu Hayakawa, Quantum algorithm for persistent Betti numbers and topological data analysis, Preprint ArXiv:2111. 00433 (2021).

[12] Casper Gyurik, Chris Cade, and Vedran Dunjko, Towards quantum advantage via topological data analysis, Preprint ArXiv:2005.02607 (2020).

[13] Sam Gunn and Niels Kornerup, Review of a quantum algorithm for Betti numbers, Preprint ArXiv:1906.07673 (2019).

[14] Dominic W. Berry, Yuan Su, Casper Gyurik, Robbie King, Joao Basso, Alexander Del Toro Barba, Abhishek Rajput, Nathan Wiebe, Vedran Dunjko, and Ryan Babbush, Quantifying quantum advantage in topological data analysis, Preprint ArXiv:2209.13581 (2022).

[15] Sam McArdle, András Gilyén, and Mario Berta, A streamlined quantum algorithm for topological data analysis with exponentially fewer qubits, Preprint ArXiv:2209.12887 (2022).

[16] He-Liang Huang, Xi-Lin Wang, Peter P. Rohde, Yi-Han Luo, You-Wei Zhao, Chang Liu, Li Li, Nai-Le Liu, Chao-Yang Lu, and Jian-Wei Pan, Demonstration of topological data analysis on a quantum processor, Optica **5**, 193 (2018).

[17] Ismail Yunus Akhalwaya, Shashanka Ubaru, Kenneth L. Clarkson, Mark S. Squillante, Vishnu Jejjala, Yang-Hui He, Kugendran Naidoo, Vasileios Kalantzis, Lior Horesh, and Exponential advantage on noisy quantum computers, Preprint ArXiv:2209.09371 (2022).

[18] Volker Kaibel and Marc E. Pfetsch, in *Algebra, Geometry and Software Systems* (Springer, Berlin, Heidelberg, 2003), p. 23.

[19] Dorit Aharonov, Vaughan Jones, and Zeph Landau, in *Proceedings of the Thirty-Eighth Annual ACM Symposium on Theory of Computing, Seattle, WA, USA* (Association for Computing Machinery, New York, NY, USA, 2006), p. 427.

[20] Dorit Aharonov, Itai Arad, Elad Eban, and Zeph Landau, Polynomial quantum algorithms for additive approximations of the Potts model and other points of the Tutte plane, Preprint ArXiv:quant-ph/0702008 (2007).

[21] Michał Adamaszek and Juraj Stacho, Complexity of simplicial homology and independence complexes of chordal graphs, Comput. Geom. **57**, 8 (2016).

[22] Marcos Crichigno and Tamara Kohler, Clique homology is QMA1-hard, Preprint ArXiv:2209.11793 (2022).

[23] In algebraic topology, a simplex is defined as the linear span of an *ordered* set. The ordering will be fixed later on by an ordering of the qubits encoding the simplices into quantum states. The particular ordering chosen does not impact topological invariants such as Betti numbers. We will thus omit it in this preliminary introduction to TDA, which also facilitates the description of our proofs in Secs. III and IV.

[24] Because we have chosen to represent simplices as basis elements of a quantum mechanical Hilbert space, all homologies in this paper are computed over the field $\mathbb{C}$. In most applications of TDA, the data is however embedded in a *real* metric space, hence one is interested in the homology over the real $\mathbb{R}$. The simplicial Laplacian is real and symmetric in both cases, so this distinction does not matter.

[25] If the grouping scale ϵ varies reasonably slow during a filtration step, the harmonic representative of a persistent hole will not change by too much. Thus quantum phase estimation applied to the initial harmonic representative will have a high chance of projecting onto the new harmonic representative. This describes a quantum algorithm that can check whether a given hole persists.

[26] P. Erdös and A. Rényi, On random graphs I, Publicationes Mathematicae Debrecen, Hungary, **6**, 290 (1959).

[27] Afra Zomorodian, Topological data analysis, Adv. Appl. Comput. Topol. **70**, 1 (2012).

[28] Larry Wasserman, Topological data analysis, Annu. Rev. Stat. Appl. **5**, 501 (2018).

[29] The computational complexity of computing normalized Betti numbers is an interesting topic by itself. It was investigated in Ref. [12], which we briefly review in Sec. VI.

[30] Leslie G. Valiant, The complexity of computing the permanent, Theor. Comput. Sci. **8**, 189 (1979).

[31] Noam Livne, A note on #p-completeness of np-witnessing relations, Inf. Process. Lett. **109**, 259 (2009).

[32] Henri Poincaré, Second complément à l'analysis situs, Proc. London Math. Soc. **1**, 277 (1900).

[33] Bjarke Hammersholt Roune and Eduardo Sáenz-de Cabezón, Complexity and algorithms for Euler characteristic of simplicial complexes, J. Symb. Comput. **50**, 170 (2013).

[34] Richard M. Karp, in *Complexity of Computer Computations, The IBM Research Symposia Series*, edited by R.E. Miller, J.W. Thatcher, and J.D. Bohlinger (Springer, Boston, MA, 1972), p. 85.

[35] Michael R. Garey and David S. Johnson, Strong NP-completeness results: Motivation, examples, and implications, J. ACM **25**, 499 (1978).

[36] If every literal X_j appears in the *SAT* instance, the vertices p_j do not contribute to the Euler characteristic and they can be neglected, leading to a graph of size $2n + k$.

[37] P. Marcos Crichigno, Supersymmetry and quantum computation, Preprint ArXiv:2011.01239 (2020).

[38] Technically, we not only require that the problem remains NP-hard but that the corresponding counting problem remains #P-hard. However, these reductions are parsimonious and preserve the number of solutions. Thus they also establish #P-hardness of the counting problem.

[39] Craig A. Tovey, A simplified NP-complete satisfiability problem, Discrete Appl. Math. **8**, 85 (1984).

[40] Piotr Berman, Marek Karpinski, and Alexander D. Scott, Computational complexity of some restricted instances of 3-SAT, Discrete Appl. Math. **155**, 649 (2007).

[41] Christian Reiher, The clique density theorem, Ann. Math. **184**, 683 (2016).

[42] Jonathan Ariel Barmak, Star clusters in independence complexes of graphs, Adv. Math. (N. Y.) **241**, 33 (2013).

[43] Anders Björner and Martin Tancer, Combinatorial alexander duality - a short and elementary proof, Discrete Comput. Geom. **42**, 586 (2009).

[44] Dimitry Kozlov, *Combinatorial Algebraic Topology* (Springer Science & Business Media, Heidelberg, 2008), Vol. 21.

[45] Matthew Kahle, Random geometric complexes, Discrete Comput. Geom. **45**, 553 (2011).

[46] Matthew Kahle and Elizabeth Meckes, Limit theorems for Betti numbers of random simplicial complexes, Homol. Homotopy Appl. **15**, 343 (2013).

[47] By possibly normalizing the metric, we assume that no two points of V are further than 1 apart. Hence V^ϵ is disconnected for $\epsilon = 0$ and fully connected for $\epsilon = 1$.

[48] Michael Goff, Extremal Betti numbers of rips complexes, Preprint ArXiv:0910.0040 (2009).

[49] Ryan Babbush, Jarrod R. McClean, Michael Newman, Craig Gidney, Sergio Boixo, and Hartmut Neven, Focus beyond quadratic speedups for error-corrected quantum advantage, PRX Quantum **2**, 010103 (2021).

[50] In theory, a set of data points could also be represented by the related Čech complex. This representation has the advantage that it is homotopy equivalent to the union of ϵ balls centered around each data point. However, the construction of Čech complexes is computationally much more expensive than the construction of Vietoris-Rips complexes, hence they are not often used in practice. They also have similarly scaling Betti numbers as Vietoris-Rips complexes [45], thus we expect similar prohibitive bounds on quantum speedups.

[51] Herbert Edelsbrunner and John L. Harer, *Computational Topology: An Introduction* (American Mathematical Society, Providence, Rhode Island, 2022).

[52] Saugata Basu, Richard Pollack, and Marie-Françoise Roy, Betti number bounds, applications and algorithms, Comb. Comput. Geom. **52**, 87 (2005).

[53] Dave Bayer, Hara Charalambous, and Sorin Popescu, Extremal Betti numbers and applications to monomial ideals, Preprint ArXiv:math/9804052 (1998).

[54] Tim Römer, Generalized Alexander duality and applications, Osaka J. Math. **38**, 469 (2001).

[55] Shashanka Ubaru and Yousef Saad, in *Proceedings of the 33rd International Conference on Machine Learning*, edited by Balcan, Maria Florina and Weinberger, and Q. Kilian, Vol. 48 (PMLR, New York, New York, USA, 2016), p. 468.

[56] Matthew Kahle, Topology of random clique complexes, Discrete Math. **309**, 1658 (2009).

[57] In general, coherent access might require a large QRAM.

[58] Chris Cade and P. Marcos Crichigno, Complexity of supersymmetric systems and the cohomology problem, Preprint ArXiv:2107.00011 (2021).

[59] Emanuel Knill and Raymond Laflamme, Power of one bit of quantum information, Phys. Rev. Lett. **81**, 5672 (1998).

[60] Peter W. Shor and Stephen P. Jordan, Estimating Jones polynomials is a complete problem for one clean qubit, Quantum Inf. Comput. **8**, 681 (2008).

[61] Greg Kuperberg, How hard is it to approximate the Jones polynomial?, Theory Comput. **11**, 183 (2015).