The topology of data hides in quantum thermal states

Stefano Scali,[1] Chukwudubem Umeano,[1] and Oleksandr Kyriienko[1]
Department of Physics and Astronomy, University of Exeter, Stocker Road, Exeter EX4 4QL, United Kingdom
(*Electronic mail: s.scali@exeter.ac.uk)

We provide a quantum protocol to perform topological data analysis (TDA) via the distillation of quantum thermal states. Recent developments of quantum thermal state preparation algorithms reveal their characteristic scaling defined by properties of dissipative Lindbladians. This contrasts with protocols based on unitary evolution which have a scaling depending on the properties of the combinatorial Laplacian. To leverage quantum thermal state preparation algorithms, we translate quantum TDA from a real–time to an imaginary–time picture, shifting the paradigm from a unitary approach to a dissipative one. Starting from an initial state overlapping with the ground state of the system, one can dissipate its energy via channels unique to the dataset, naturally distilling its information. Therefore calculating Betti numbers translates into a purity estimation. Alternatively, this can be interpreted as the evaluation of the Rényi 2–entropy, Uhlmann fidelity or Hilbert–Schmidt distance relative to thermal states with the embedded topology of simplicial complexes. Our work opens the field of TDA toward a more physical interpretation of the topology of data.

Extracting useful information from very large datasets[1] is challenging given the tools available today, both from an algorithmic and a computational point of view. For this reason, approaches to analyze the crucial features of datasets have emerged. One of these approaches consists of extracting information from the "shape" of data, i.e. from its topological features, via the tools of topological data analysis (TDA)[2–5]. TDA finds applications in several areas spanning physics[6–11], medicine[12,13], and machine learning[14,15]. Further applications can be found in Refs.[16,17]. However, even TDA suffers from an unfavorable scaling with the system dimension. To bypass this problem, a natural extension comes in the form of quantum topological data analysis (QTDA). Since the first stages of QTDA[18], these quantum algorithms have exploited the unitary evolution generated by the combinatorial Laplacian to extract its kernel information, key to access the topological properties. This requires state preparation based on Grover's search[19] and quantum phase estimation (QPE)[20,21]. Successive protocols proposed improved scaling by finding efficient representations of the combinatorial Laplacian while replacing Grover's search and QPE[22–24]. Others found smart encoding strategies to provide, under certain conditions, an almost quintic advantage in space saving[25]. Alternative protocols based on cohomology approaches[26] or hybrid quantum–classical pipelines focused on near–term devices[27] have been proposed. While the existence of real instances of quantum advantage in QTDA is still debated[24,28], the estimation of (normalized) Betti numbers on general chain complexes was shown to be DQC1–hard, i.e. classically intractable[29,30]. Safe from known protocols of dequantization[31–33], problems involving clique complexes have been shown to be QMA_1–hard and contained in QMA[34,35].

In this manuscript, we reinterpret topological data analysis from a quantum thermal state perspective. We propose a distinct paradigm to perform quantum topological data analysis, dubbed thermal–QTDA, that relies on the dissipative process defined by the combinatorial Laplacian associated with the dataset. The thermal state built from this process reveals, at low temperature, the topological features of the dataset. In this way, the evaluation of Betti numbers reduces to a purity test on the low–temperature quantum thermal state of the combinatorial Laplacian. We further interpret this result as the Rényi 2–entropy of the thermal state and as the Uhlmann fidelity or the Hilbert–Schmidt distance between the maximally mixed state of the system and its imaginary time evolved version. Thermal–QTDA inherits the performance guarantees of the thermal state preparation protocol adopted and the efficiency of the purity test. This makes it a viable choice for early fault-tolerant quantum computers, with application in data analysis and machine learning.

I. BETTI NUMBERS IN QUANTUM THERMAL STATES

Let us briefly recall the theory behind QTDA and the evaluation of Betti numbers. Consider a simplicial complex Γ obtained from a dataset of dimension N, filtration distance, and a metric. Let S_k be the set of k–simplices of the complex $\Gamma = \{\mathrm{S}_k\}_{k=0}^{N-1}$. Let $\mathscr{H}_k$ be the $\binom{N}{k+1}$–dimensional Hilbert space spanned by all possible k–simplices. We refer to the single simplices $s_k \in \mathscr{H}_k$ with $s_k = j_0 \cdots j_k$ where j_i is the ith vertex in s_k. Consider the boundary operator (map) $\partial_k : \mathscr{H}_k \mapsto \mathscr{H}_{k-1}$ defined by its action on the single simplices as $\partial_k |s_k\rangle = \sum_{l=0}^{k-1} (-1)^l |s_{k-1}(l)\rangle$, where $|s_{k-1}(l)\rangle = j_0 \cdots \hat{j}_l \cdots j_k$ is the $(k-1)$–simplex obtained by removing the lth vertex from s_k. The action of the boundary operator on the k–simplices and their linear combinations determines the chain complex. Note that the operators and spaces just described can be restricted to the domain of the simplicial complex Γ and it is usually indicated by a "tilde" as $\tilde{\bullet}$. From the k–homology group of Γ defined as $\mathbb{H}_k = \ker(\tilde{\partial}_k)/\mathrm{im}(\tilde{\partial}_{k+1})$, we obtain the Betti number as $b_k = \dim(\mathbb{H}_k)$. As a result of Hodge theory[36], Betti numbers can also be evaluated as $b_k = \dim(\ker(\Delta_k)) = \dim(\tilde{\mathscr{H}}_k) - \mathrm{rank}(\Delta_k)$, where the combinatorial Laplacian Δ_k relates to the boundary operators via $\Delta_k = \tilde{\partial}_k^\dagger \tilde{\partial}_k + \tilde{\partial}_{k+1} \tilde{\partial}_{k+1}^\dagger$. From this expression, it is natural to look at spectral methods to estimate either the kernel or the rank of Δ_k. In this direction, since the seminal work by Lloyd[18], several techniques relying on QPE or alternative spectrum evaluations have been developed[22–25,27]. In this work, we propose an alternative approach built upon a dissipative process or, equivalently, imaginary time evolution.

Given a dataset of dimension N, a filtration distance, and a metric, we can construct the simplicial complex $\Gamma = \{\mathrm{S}_k\}_{k=0}^{N-1}$ defining the topological properties of the dataset at that scale. Here, S_k is the set of the k–simplices in the complex. From this simplicial complex, we can construct the kth combinatorial Laplacian Δ_k which encodes information of Γ via its topological boundaries. We propose to estimate the kth Betti number b_k of Γ via the preparation of the low-temperature thermal state ρ_β as

$$b_k = \lim_{\beta\to\infty} \mathrm{Tr}\left\{\rho_\beta^2\right\}^{-1}, \tag{1}$$

where $\beta = 1/T$ is the inverse temperature and the thermal state ρ_β is the Gibbs state $\rho_\beta = e^{-\beta\Delta_k}/\,\mathrm{Tr}\left\{e^{-\beta\Delta_k}\right\}$ (assuming Planck natural units). For Eq. (1) to be valid, the thermalization process must start with an initial state overlapping with the ground state of the system. A possible choice is the maximally mixed state $\rho_{\mathrm{mix}} = \rho_{\beta=0} = \mathbb{I}_k/|\mathrm{S}_k|$ living in the Hilbert space $\tilde{\mathscr{H}}_k$, spanned by the elements of S_k. Note that the overlap of the initial state with the ground state (single or multiple) of the system affects the thermal state preparation, that is, the smaller the overlap, the slower the thermalization. Once an adequate initial state is formed, the thermalization process leads to the ground state of Δ_k, with the inverse purity of such state being its degeneracy. In case of degenerate ground state, the thermalization process will guide the system towards a thermal state that is an equiprobable mixture of the non–unique ground states. The overlap between this final state and the initial fully mixed state leads to the Betti numbers. This shows how the evaluation of Betti numbers via Eq. (1) is related to the well–known expression $b_k = \dim(\ker(\Delta_k))$[18]. In the following, we refer to the interpretation of QTDA via Eq. (1) as thermal–QTDA.

One may wonder what the physics behind Eq. (1) is. Eq. (1) is equivalent to evaluating the degeneracy of the ground state, relying on the fact that b_k is defined as the kernel dimension of the kth homology group, as previously seen. To obtain this degeneracy, we start with an initial state with support on the ground state of the combinatorial Laplacian Δ_k, e.g. the maximally mixed state ρ_{mix}, and we cool it down to $\beta\to\infty$ via the channels of dissipation determined by Δ_k. In this limit, the state converges to the ground state of Δ_k. We show a schematic of the protocol in Fig. 1(a). In Fig. 1(b) we show an example of Betti number numerical evaluation via Eq. (1) and the corresponding simplicial complex (inset). Notably, the thermalization approach is particularly suited to study “the integer problem” of Betti numbers. In fact, given the monotonic nature of Eq. (1), we can approximate the Betti number as $\lfloor b_k \rfloor$ when the state is prepared at large but finite β. This suggests the possibility of using an approximate thermal state preparation to reach a low–enough finite temperature.

II. INTERPRETATIONS

Eq. (1) can be interpreted as the Uhlmann fidelity between the maximally mixed state ρ_{mix} and its imaginary–time evolved version $\rho_{\mathrm{mix}}(\tau)$. This becomes possible if one translates the unitary evolution problem from Minkowski space to Euclidean space by allowing time to take imaginary values and replacing $t = -i\tau$. Thus, the Uhlmann fidelity can be expressed as

$$\begin{aligned}\mathscr{F}(\rho_{\mathrm{mix}},\rho_{\mathrm{mix}}(\tau)) &= \mathrm{Tr}\left\{(\rho_{\mathrm{mix}}^{1/2}\rho_{\mathrm{mix}}(\tau)\rho_{\mathrm{mix}}^{1/2})^{1/2}\right\}^2\\ &= \mathrm{Tr}\left\{e^{-\Delta_k\tau}\right\}^2/(\mathrm{Tr}\left\{e^{-2\Delta_k\tau}\right\}|\mathrm{S}_k|)\\ &= \mathrm{Tr}\left\{\rho_\beta^2\right\}^{-1}/|\mathrm{S}_k|,\end{aligned} \tag{2}$$

with $\rho_{\mathrm{mix}}(\tau) = e^{-\Delta_k\tau}\rho_{\mathrm{mix}}e^{-\Delta_k\tau}/\big(\mathrm{Tr}\left\{\rho_{\mathrm{mix}}e^{-2\Delta_k\tau}\right\}\big)$, where the factor at the denominator makes sure the density matrix is normalized throughout the evolution. While the components populating the kernel contribute to the fidelity at all imaginary times, the remaining components decay to zero exponentially. We obtain the final line in Eq. (2) by means of Wick rotation, i.e. replacing $\tau = \beta$, thus reinterpreting the imaginary time as a temperature. These two pictures, imaginary time in quantum mechanics and temperature in statistical physics, are indeed formally related through analytic continuation[37] by the Osterwalder–Schrader theorem[38,39].

As a natural extension, Eq. (1) can also be interpreted as the quantum version[40,41] of the Rényi 2–entropy[42] of the thermal

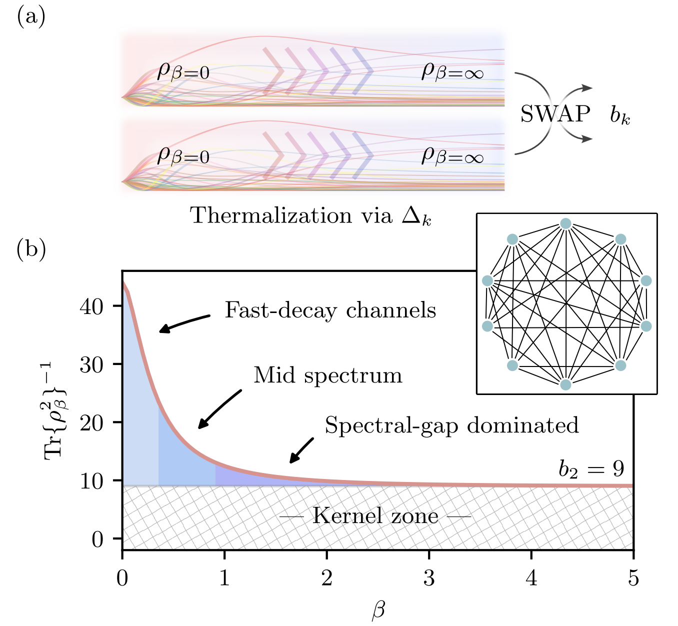

FIG. 1. **Schematic of the protocol and Betti number evaluation.** In (a), two copies of the state ρ_β are prepared at high temperature $\beta = 0$. The states are thermalized to low temperature $\beta = \infty$ along the trajectories of dissipation dictated by the combinatorial Laplacian Δ_k. The resulting states are then swapped to perform a purity test and the Betti number evaluated through Eq. (1). In (b), we show an example numerical evaluation of the Betti number b_2 for the simplicial complex shown in the inset. We qualitatively identify regions of thermalization where specific decaying channels dominate. The hatched area represents the dimensionality of the kernel that remains inaccessible to the protocol.

state ρ_β,

$$\mathscr{H}_2(\rho_\beta) = \log\left(\mathrm{Tr}\left\{\rho_\beta^2\right\}^{-1}\right). \tag{3}$$

Here, the base of the logarithm determines the unit of information. The Rényi 2–entropy is often referred to as the collision entropy. Rényi entropies are of crucial interest for the estimation of the statistical properties of quantum states, finding applications as entanglement measures[43–46], in the estimation of Gaussianity of quantum states[47,48], and as a measure of non–stabilizerness (also known as magic)[49]. Given their ubiquity, several quantum algorithms have been developed to estimate Rényi entropies[50–53].

Finally, Eq. (1) can be interpreted in terms of the Hilbert–Schmidt distance (Schatten 2–norm),

$$\begin{aligned}\mathscr{D}_{\mathrm{HS}}(\rho_{\mathrm{mix}},\rho_\beta) &= ||\rho_{\mathrm{mix}} - \rho_\beta||_2^2 \\ &= \mathrm{Tr}\left\{\rho_\beta^2\right\} - (2|\mathrm{S}_k|-1)/|\mathrm{S}_k|^2. \end{aligned} \tag{4}$$

This distance is often employed as the cost function in variational quantum algorithms (VQAs)[54–58].

III. ROUTINES

Assuming that we are given two copies of the thermal state ρ_β at low temperature, we only need to perform a purity test to implement Eq. (1) on a quantum computer. While this can be done in the form of a traditional SWAP test or its destructive variant[59], we note that some generalizations with short–depth circuit have been proposed[60–62]. Alternatively, purity can be evaluated via single distinct classical shadows of the thermal state[44–46,63]. This approach trades two identical copies of the thermal state and a simple measurement routine for a single copy of the thermal state and a larger number of (randomized) measurements. In the following, we consider a simple SWAP test. To perform it, we need an additional ancilla qubit. We place a Hadamard gate on the ancilla, a Fredkin (CSWAP) gate controlled on the ancilla between the two copies of qubits in the state registers, and then again a Hadamard gate on the ancilla. By measuring the ancilla qubit multiple times, we construct the probabilities P_0 and P_1 relative to the outcome 0 and 1 respectively. We obtain the purity of the thermal state from $\mathrm{Tr}\left\{\rho_\beta^2\right\} = P_0 - P_1$[64]. Once the SWAP test has been performed, the result is obtained by truncating (flooring) the Betti number in Eq. (1), $\lfloor b_k \rfloor$, as previously mentioned.

The thermal state can be prepared in several ways. A nature–inspired thermal state preparation can be found in Ref.[65] where the authors propose a protocol to simulate the Lindbladian (or the relative discriminant proxy) whose fixed point is approximately a quantum Gibbs state. By means of this construction and the introduction of an operator Fourier transform (FT) for the Lindblad operators, the authors give the recipes for incoherent and coherent implementations. See later for an example implementation using such coherent approach. Several variational quantum algorithms have been developed to prepare thermal states using imaginary time evolution[66], open system dynamics[67], and hybrid quantum circuits using classical neural networks[68]. Thermal–QTDA also opens the door to possible quantum–inspired implementations based on thermal tensor network (TTN) states. Examples of these are the realization of thermal states using minimally entangled typical thermal states (METTS) and imaginary time evolution[69] or the exponential tensor renormalization group (XTRG)[70] producing accurate low–temperature thermal states by exponentially evolving the matrix product operator (MPO) along a path of imaginary time evolution. In Ref.[71], an interesting approach to the estimation of linear functions of ρ_β avoids the direct construction of the mixed state ρ_β. The authors combine pure thermal quantum states[72,73] and classical shadow tomography[74–76] to estimate several Gibbs state expectation values. However, thermal–QTDA requires the estimation of nonlinear functions of ρ_β. In this direction, the authors envision improvements to their algorithm in the form of derandomization[77–79].

IV. RUNTIME

The cost of thermal–QTDA is inherited from the scaling of the two subroutines, the quantum thermal state preparation and the purity test. For the latter, the SWAP test requires two copies of the quantum thermal state and an ancilla qubit, with a final measurement on the ancilla qubit. The destructive alternative requires two copies of the quantum thermal state but trades the additional ancilla with a final measurement onto the full set of qubits. A test result with additive error ε requires $\mathscr{O}(\varepsilon^{-2})$ runs.

Now, the quantum thermal state preparation. When estimating Betti numbers, quantum phase estimation explicitly introduces an inverse linear dependence on the spectral gap of the combinatorial Laplacian, $\mathscr{O}(N^3\delta_{\mathrm{gap}}^{-1})$[18]. In contrast, the cost of estimating Betti numbers via quantum thermal state preparation varies depending on the chosen algorithm. For example, simulating an incoherent version of the Lindbladian via the operator FT costs $\tilde{\mathscr{O}}(\beta t_{\mathrm{mix}}^2/\varepsilon)$ while the coherent version via the discriminant proxy and an adiabatic path to low temperature costs $\tilde{\mathscr{O}}\left((\varepsilon\beta^2\|H\|+\beta)/\varepsilon\lambda_{\mathrm{gap}}^{3/2}\right)$[65]. Here, we use the soft-O notation $\tilde{\mathscr{O}}$, which corresponds to scaling that ignores constant and logarithmic factors. The mixing time of the Lindbladian, t_{mix}, and the minimum spectral gap of the discriminant proxies along the adiabatic path, λ_{gap}, are related by the approximate detailed balance introduced in Ref.[65]. In the perspective of open systems and Gibbs samplers, the spectral gap of the Lindbladian is also related to the spectral gap of the combinatorial Laplacian. However, this relation strongly depends on the characteristics of the simplicial complex (system Hamiltonian) considered[80–85]. As a general guideline, a faster convergence to the thermal state can be obtained in the presence of stronger dissipative coupling, ad–hoc transitions boosting convenient channels of dissipation, or higher finite temperatures of thermalization. Note that, producing thermal states for long mixing times or closing spectral gaps can be challenging. Hard instances of thermal state preparation are expected regardless of the algorithm chosen, since the ground state preparation problem is QMA–hard[86–88].

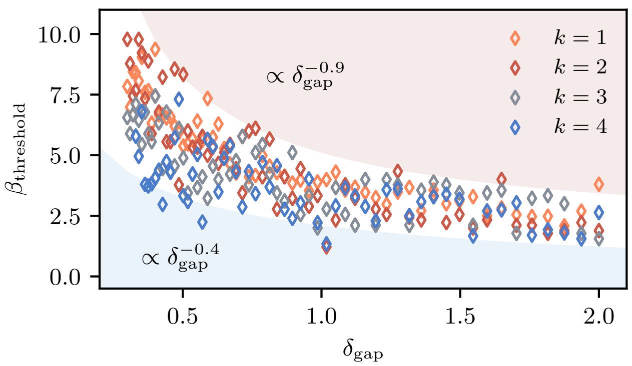

FIG. 2. **Spectral gap scaling.** We sample the imaginary time/inverse temperature needed to obtain $\frac{\mathrm{d}}{\mathrm{d}\tau}\,\mathrm{Tr}\left\{\rho_{\text{mix}}e^{-\Delta_k\tau}\right\} \leq 10^{-3}$ for random instances of simplicial complexes with $N=10$ and $k=1,2,3,4$. We find a corresponding inverse temperature dependence on the spectral gap of the combinatorial Laplacian of $\beta_{\text{threshold}} = \mathscr{O}(\delta_{\text{gap}}^{-(0.4\div 0.9)})$. Here, the notation $a\div b$ indicates a range of values between a and b.

In Fig. 2 we show the spectral gap scaling of thermal–QTDA numerically evaluated for random simplicial complexes chosen with $N=10$ and $k=1,2,3,4$. The initial state ρ_{mix} is thermalized until the stopping criterion is met, that is, $\frac{\mathrm{d}}{\mathrm{d}\tau}\,\mathrm{Tr}\left\{\rho_{\text{mix}}e^{-\Delta_k\tau}\right\} \leq 10^{-3}$. We refer to the smallest inverse temperature that satisfies such criterion as the threshold inverse temperature $\beta_{\text{threshold}}$. We find this to be $\beta_{\text{threshold}} = \mathscr{O}(\delta_{\text{gap}}^{-(0.4\div 0.9)})$, where the notation $a\div b$ indicates a range of values between a and b, suggesting an inverse–sub–linear dependence of thermal–QTDA on the spectral gap of the combinatorial Laplacian.

V. EXAMPLE IMPLEMENTATION

In Fig. 3, we show an example circuit to prepare the thermal state ρ_β. Following the coherent approach in Ref.[65], we implement the discriminant proxy D_β governing the dissipation dictated by the combinatorial Laplacian,

$$D_\beta := \frac{1}{|\mathscr{A}|}\sum_{j,\bar{\omega}} \sqrt{\gamma(\bar{\omega})\gamma(-\bar{\omega})}A_j(\bar{\omega})\otimes A_j^*(\bar{\omega}) - \frac{\gamma(\bar{\omega})}{2}\Big(A_j^\dagger(\bar{\omega})A_j(\bar{\omega})\otimes\mathbb{I} + \mathbb{I}\otimes A_j^{*\dagger}(\bar{\omega})A_j^*(\bar{\omega})\Big). \tag{5}$$

The top eigenvector of the discriminant proxy D_β is the canonical purification of the thermal state at inverse temperature $\beta/2$, that is $|\sqrt{\rho_\beta}\rangle \propto \sum_i e^{-\beta E_i/2}|\psi_i\rangle\otimes|\psi_i\rangle$, where the E_i are the eigenvalues of the Hamiltonian operator $H=\Delta_k$. By using two copies of this purified version of the thermal state, we can perform the SWAP test and evaluate $\mathrm{Tr}\left\{\rho_\beta^2\right\}$, getting Betti numbers via Eq. (1).

In Eq. (5), $A_j(\bar{\omega})$ are the Fourier–transformed versions of the Heisenberg time–evolved jump operators $A_j(\bar{t})$ (see later), $\gamma(\bar{\omega})$ are the jump weights, and $\mathscr{A}$ is the set of jump operators A_j. For convenience, we choose the A_j to be the self–adjoint single–site Pauli operators. We also choose the transition weights $\gamma(\bar{\omega})$ to be the Metropolis weights. This choice ensures that the transition weights satisfy $\gamma(\bar{\omega})/\gamma(-\bar{\omega}) = e^{\beta\bar{\omega}}$, consequently the detailed balance condition. The "bar" indicates discretized frequencies in the set of frequencies S_{ω_0}. These are multiples of a base frequency ω_0, which is defined by the relation $\omega_0 t_0 = 2\pi/M$, where t_0 is the base discretized time and M is the number of discretized points. To cover the full set of transition (Bohr) frequencies defined by H, the number of points and the base frequency are chosen such that $M\omega_0/2 \geq \|H\|$.

To build the circuit implementing Eq. (5), we rewrite the discriminant in terms of the reflection R and the isometry T' as $\mathbb{I}+D_\beta = T'^\dagger RT'$[65,89,90]. The isometry T' is a coherent sum of the two sub–isometries T_0 and T_1, $T' = |0\rangle\otimes T_0 + T_1\otimes|1\rangle$. Each of these sub–isometries returns a superposition of jump operators applied on the system register $|\text{sys}\rangle$ of interest. This is done via the operator FT $\mathscr{F}[A]$ controlled on the ancillary register $|j\rangle$ that carries the weight distribution of the jump operators. In T_0 and T_1, we also find the encoder for the transition rates, $Y_{\gamma(\bar{\omega})}$. R is a reflection ($R^2=\mathbb{I}$) that implements a bit flip X on the ancilla qubit $|+\rangle$ dedicated to the control of the sub–isometries T_0 and T_1 and a sign change to the Bohr frequency register. These operations are all controlled by the ancillary qubit register ($|0\rangle$) storing the transition rates. If the reflection is triggered, it generates a cross term between the two $|\text{sys}\rangle$ copies.

The operator FT $\mathscr{F}[A]$, introduced in Ref.[65], implements the following weighted transform on the jump operators, $A(\bar{\omega}) := \sum_{\bar{t}} e^{-i\bar{\omega}\bar{t}} f(\bar{t})A(\bar{t})$, where $A(\bar{t}) = e^{iH\bar{t}}Ae^{-iH\bar{t}}$ are the time evolved versions of the jump operators. Here, the unique channels of dissipation defined by the combinatorial Laplacian Δ_k enter into the picture. By equating H with Δ_k as previously stated, we build jumps that encode information of the combinatorial Laplacian and that dissipate the energy of the thermal state accordingly. We show the circuit implementing $\mathscr{F}[A]$ in Fig. 3(b). Given the weighted nature of the FT implemented, the time evolution of the jump operators is controlled by the weighting function $f(\bar{t})$. After the control, the weights are quantum Fourier transformed into the Bohr frequencies $|\bar{\omega}\rangle$ to then control the transition rates $\gamma(\bar{\omega})$ (Boltzmann weights). Notably, when the weighting function follows a Gaussian distribution, the sub–routine achieves the performance scaling of boosted phase estimation[65,91]. Then, the unitary $Y_{\gamma(\bar{\omega})}$ encodes the transition rates $\gamma(\bar{\omega})$ in the amplitudes of the ancillary qubit $|0\rangle$. The unitary is controlled by the output of the operator FT $\mathscr{F}[A]$, namely the Bohr frequency read–out $|\bar{\omega}\rangle$. Finally, F is the negation of the Bohr frequency register. This makes sure that the ensuing isometries $T_0^\dagger$ and $T_1^\dagger$ produce the correct jump operator, that is $A(\bar{\omega}) = A^\dagger(-\bar{\omega})$ (see details in Ref.[65]).

To finally find the top eigenvector of the discriminant proxy, i.e. the quantum thermal state, we need to perform quantum simulated annealing. Having access to the block encoding of D_β, we can build a unitary U_{D_β} with $\mathscr{O}(\lambda_{\text{gap}}^{-1/2}(D_\beta))$ calls of D_β. Starting from the maximally mixed state at very high temperature $\beta_0\approx 0$, we proceed with adiabatic

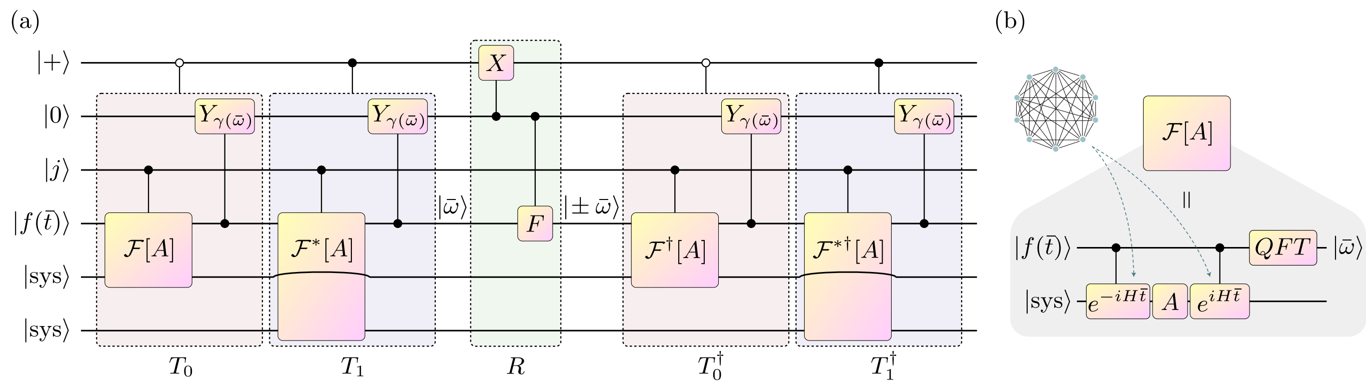

FIG. 3. **Discriminant proxy and operator Fourier transform circuits.** In (a) we show a circuit for preparing quantum thermal states ρ_β as a sub–routine of thermal–QTDA. These are prepared as the top eigenvector of the discriminant proxy D_β via quantum simulated annealing[65]. These copies are then used to estimate Betti numbers via a SWAP test and Eq. (1). The operator Fourier transform $\mathscr{F}[A]$ from the discriminant proxy implementation is shown in (b). This routine encodes the information of the simplicial complex Γ (and the relative combinatorial Laplacian Δ_k) into the jump operators A.

steps of unitaries through a path of decreasing temperatures, $U_{D_{\beta_0}} \to \cdots \to U_{D_{\beta_{\mathrm{target}}}}$, finally reaching a very low temperature β_{target}. At the end of this path, we obtain the top eigenvector of the discriminant, thus the thermal state. Once we have prepared two copies of the state, we can perform the SWAP test and extract the Betti numbers.

VI. CONCLUSIONS AND OUTLOOK

In this manuscript, we reinterpreted a pillar of quantum topological data analysis, that is, the estimation of Betti numbers, under the lens of quantum thermal states. We dubbed this interpretation thermal–QTDA. The protocol relies upon the preparation of two copies of quantum thermal states at low temperature and a purity test. The thermalization of the states happens through the unique channels of dissipation of the combinatorial Laplacian. The purity test is naturally related to other quantities of interest such as the Rényi 2–entropy, Uhlmann fidelity or Hilbert–Schmidt distance. The scaling of thermal–QTDA routines directly depends on the scaling of its two sub–routines, the thermal state preparation and the purity test. We presented a possible coherent Lindbladian approach for the former, while a simple SWAP test is sufficient for the latter.

Thermal–QTDA has theoretical and practical importance. On the theoretical side, it hints to possible quantum thermodynamic quantities as the tools of interest to perform QTDA tasks. Here, one may wonder if thermal states prepared from simplicial complexes can reveal additional topological features of the data. On the practical side, it opens the field of QTDA to an entire new set of possible implementations via quantum thermal state preparation protocols. Here, the crucial step is breaking away from the traditional scaling of quantum phase estimation, which may lead to an advantageous dependence on the spectral gap of the combinatorial Laplacian. From a perspective of applications, extensions of our approach to the evaluation of persistent Betti numbers are desirable. In fact, these enclose additional information on the persistence of features and can yield enhanced explainable AI protocols[92]. We expect advances in QTDA will facilitate the use of topological properties as features of datasets, aiding the building of models with high predictive power and excellent generalization.

VII. ACKNOWLEDGEMENTS

The authors thank Ravindra Mutyamsetty, Zhihao Lan and Okello Ketley for useful discussions. The authors acknowledge the support from the Innovate UK ISCF Feasibility Study project number 10030953 granted under the "Commercialising Quantum Technologies" competition round 3.

[1] P. Taylor, "Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025," (2023).
[2] Edelsbrunner, Letscher, and Zomorodian, "Topological persistence and simplification," Discrete & Computational Geometry **28**, 511–533 (2002).
[3] A. Zomorodian and G. Carlsson, "Computing persistent homology," Discrete & Computational Geometry **33**, 249–274 (2004).
[4] G. Carlsson, A. Zomorodian, A. Collins, and L. J. Guibas, "Persistence barcodes for shapes," International Journal of Shape Modeling **11**, 149–187 (2005).
[5] G. Carlsson, "Topology and data," Bulletin of the American Mathematical Society **46**, 255–308 (2009).
[6] M. Kramár, R. Levanger, J. Tithof, B. Suri, M. Xu, M. Paul, M. F. Schatz, and K. Mischaikow, "Analysis of kolmogorov flow and rayleigh–bénard convection using persistent homology," Physica D: Nonlinear Phenomena **334**, 82–98 (2016).
[7] F. A. Khasawneh and E. Munch, "Chatter detection in turning using persistent homology," Mechanical Systems and Signal Processing **70-71**, 527–541 (2016).
[8] Y. Lee, S. D. Barthel, P. Dłotko, S. M. Moosavi, K. Hess, and B. Smit, "Quantifying similarity of pore-geometry in nanoporous materials," Nature Communications **8** (2017), 10.1038/ncomms15396.
[9] P. Pranav, H. Edelsbrunner, R. van de Weygaert, G. Vegter, M. Kerber, B. J. T. Jones, and M. Wintraecken, "The topology of the cosmic web in terms of persistent betti numbers," Monthly Notices of the Royal Astronomical Society **465**, 4281–4310 (2016).

[10] A. Tirelli and N. C. Costa, "Learning quantum phase transitions through topological data analysis," Physical Review B **104** (2021), 10.1103/physrevb.104.235146.

[11] B. Olsthoorn, "Persistent homology of quantum entanglement," Physical Review B **107** (2023), 10.1103/physrevb.107.115174.

[12] M. Saggar, O. Sporns, J. Gonzalez-Castillo, P. A. Bandettini, G. Carlsson, G. Glover, and A. L. Reiss, "Towards a new approach to reveal dynamical organization of the brain using topological data analysis," Nature Communications **9** (2018), 10.1038/s41467-018-03664-4.

[13] T. Qaiser, Y.-W. Tsang, D. Taniyama, N. Sakamoto, K. Nakane, D. Epstein, and N. Rajpoot, "Fast and accurate tumor segmentation of histology images using persistent homology and deep convolutional features," Medical Image Analysis **55**, 1–14 (2019).

[14] V. Kovacev-Nikolic, P. Bubenik, D. Nikolić, and G. Heo, "Using persistent homology and dynamical distances to analyze protein binding," Statistical Applications in Genetics and Molecular Biology **15** (2016), 10.1515/sagmb-2015-0057.

[15] F. Hensel, M. Moor, and B. Rieck, "A survey of topological machine learning methods," Frontiers in Artificial Intelligence **4** (2021), 10.3389/frai.2021.681108.

[16] L. Wasserman, "Topological data analysis," Annual Review of Statistics and Its Application **5**, 501–532 (2018).

[17] F. Chazal and B. Michel, "An introduction to topological data analysis: Fundamental and practical aspects for data scientists," Frontiers in Artificial Intelligence **4** (2021), 10.3389/frai.2021.667963.

[18] S. Lloyd, S. Garnerone, and P. Zanardi, "Quantum algorithms for topological and geometric analysis of data," Nature Communications **7** (2016), 10.1038/ncomms10138.

[19] L. K. Grover, "A fast quantum mechanical algorithm for database search," arXiv (1996), 10.48550/arXiv.quant-ph/9605043.

[20] A. Y. Kitaev, "Quantum measurements and the abelian stabilizer problem," arXiv (1995), 10.48550/arXiv.quant-ph/9511026.

[21] A. Y. Kitaev, "Quantum computations: algorithms and error correction," Russian Mathematical Surveys **52**, 1191–1249 (1997).

[22] S. Ubaru, I. Y. Akhalwaya, M. S. Squillante, K. L. Clarkson, and L. Horesh, "Quantum topological data analysis with linear depth and exponential speedup," arXiv (2021), 10.48550/arXiv.2108.02811.

[23] I. Y. Akhalwaya, S. Ubaru, K. L. Clarkson, M. S. Squillante, V. Jejjala, Y.-H. He, K. Naidoo, V. Kalantzis, and L. Horesh, "Towards quantum advantage on noisy quantum computers," arXiv (2022), 10.48550/arXiv.2209.09371.

[24] D. W. Berry, Y. Su, C. Gyurik, R. King, J. Basso, A. D. T. Barba, A. Rajput, N. Wiebe, V. Dunjko, and R. Babbush, "Analyzing prospects for quantum advantage in topological data analysis," PRX Quantum **5** (2024), 10.1103/prxquantum.5.010319.

[25] S. McArdle, A. Gilyén, and M. Berta, "A streamlined quantum algorithm for topological data analysis with exponentially fewer qubits," arXiv (2022), 10.48550/arXiv.2209.12887.

[26] N. A. Nghiem, X. D. Gu, and T.-C. Wei, "Quantum algorithm for estimating betti numbers using a cohomology approach," arXiv (2023), 10.48550/arXiv.2309.10800.

[27] S. Scali, C. Umeano, and O. Kyriienko, "Quantum topological data analysis via the estimation of the density of states," arXiv (2023), 10.48550/arXiv.2312.07115.

[28] A. Schmidhuber and S. Lloyd, "Complexity-theoretic limitations on quantum algorithms for topological data analysis," arXiv (2022), 10.48550/arXiv.2209.14286.

[29] C. Gyurik, C. Cade, and V. Dunjko, "Towards quantum advantage via topological data analysis," Quantum **6**, 855 (2022).

[30] C. Cade and P. M. Crichigno, "Complexity of supersymmetric systems and the cohomology problem," arXiv (2021), 10.48550/arXiv.2107.00011.

[31] E. Tang, "A quantum-inspired classical algorithm for recommendation systems," in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, STOC '19 (ACM, 2019).

[32] A. Gilyén, S. Lloyd, and E. Tang, "Quantum-inspired low-rank stochastic regression with logarithmic dependence on the dimension," arXiv (2018), 10.48550/arXiv.1811.04909.

[33] N.-H. Chia, A. Gilyén, T. Li, H.-H. Lin, E. Tang, and C. Wang, "Sampling-based sublinear low-rank matrix arithmetic framework for dequantizing quantum machine learning," in *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, STOC '20 (ACM, 2020).

[34] M. Crichigno and T. Kohler, "Clique homology is qma1-hard," arXiv (2022), 10.48550/arXiv.2209.11793.

[35] R. King and T. Kohler, "Promise clique homology on weighted graphs is QMA_1-hard and contained in QMA," arXiv (2023), 10.48550/arXiv.2311.17234, 2311.17234.

[36] L.-H. Lim, "Hodge laplacians on graphs," arXiv (2019), 10.48550/arXiv.1507.05379.

[37] A. Zee, *Quantum field theory in a nutshell*, Vol. 7 (Princeton university press, 2010).

[38] K. Osterwalder and R. Schrader, "Axioms for euclidean green's functions," Communications in Mathematical Physics **31**, 83–112 (1973).

[39] K. Osterwalder and R. Schrader, "Axioms for euclidean green's functions ii," Communications in Mathematical Physics **42**, 281–305 (1975).

[40] D. Petz, "Quasi-entropies for finite quantum systems," Reports on Mathematical Physics **23**, 57–65 (1986).

[41] M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr, and M. Tomamichel, "On quantum rényi entropies: A new generalization and some properties," Journal of Mathematical Physics **54** (2013), 10.1063/1.4838856.

[42] A. Rényi, "On measures of entropy and information," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, Vol. 4 (University of California Press, 1961) pp. 547–562.

[43] Y.-X. Wang, L.-Z. Mu, V. Vedral, and H. Fan, "Entanglement rényi α–entropy," Physical Review A **93** (2016), 10.1103/physreva.93.022324.

[44] A. Elben, B. Vermersch, M. Dalmonte, J. Cirac, and P. Zoller, "Rényi entropies from random quenches in atomic hubbard and spin models," Physical Review Letters **120** (2018), 10.1103/physrevlett.120.050406.

[45] T. Brydges, A. Elben, P. Jurcevic, B. Vermersch, C. Maier, B. P. Lanyon, P. Zoller, R. Blatt, and C. F. Roos, "Probing rényi entanglement entropy via randomized measurements," Science **364**, 260–263 (2019).

[46] A. Elben, B. Vermersch, C. F. Roos, and P. Zoller, "Statistical correlations between locally randomized measurements: A toolbox for probing entanglement in many-body quantum states," Physical Review A **99** (2019), 10.1103/physreva.99.052323.

[47] G. Adesso, D. Girolami, and A. Serafini, "Measuring gaussian quantum information and correlations using the rényi entropy of order 2," Physical Review Letters **109** (2012), 10.1103/physrevlett.109.190502.

[48] J. Park, J. Lee, K. Baek, and H. Nha, "Quantifying non-gaussianity of a quantum state by the negative entropy of quadrature distributions," Physical Review A **104** (2021), 10.1103/physreva.104.032415.

[49] L. Leone, S. F. Oliviero, and A. Hamma, "Stabilizer rényi entropy," Physical Review Letters **128** (2022), 10.1103/physrevlett.128.050402.

[50] T. Li and X. Wu, "Quantum query complexity of entropy estimation," IEEE Transactions on Information Theory **65**, 2899–2921 (2019).

[51] J. Acharya, I. Issa, N. V. Shende, and A. B. Wagner, "Estimating quantum entropy," IEEE Journal on Selected Areas in Information Theory **1**, 454–468 (2020).

[52] S. Subramanian and M.-H. Hsieh, "Quantum algorithm for estimating α-renyi entropies of quantum states," Physical Review A **104** (2021), 10.1103/physreva.104.022428.

[53] Y. Wang, B. Zhao, and X. Wang, "Quantum algorithms for estimating quantum entropies," Physical Review Applied **19** (2023), 10.1103/physrevapplied.19.044041.

[54] R. LaRose, A. Tikku, É. O'Neel-Judy, L. Cincio, and P. J. Coles, "Variational quantum state diagonalization," npj Quantum Information **5** (2019), 10.1038/s41534-019-0167-6.

[55] A. Arrasmith, L. Cincio, A. T. Sornborger, W. H. Zurek, and P. J. Coles, "Variational consistent histories as a hybrid algorithm for quantum foundations," Nature Communications **10** (2019), 10.1038/s41467-019-11417-0.

[56] S. Khatri, R. LaRose, A. Poremba, L. Cincio, A. T. Sornborger, and P. J. Coles, "Quantum-assisted quantum compiling," Quantum **3**, 140 (2019).

[57] K. C. Tan and T. Volkoff, "Variational quantum algorithms to estimate rank, quantum entropies, fidelity, and fisher information via purity minimization," Physical Review Research **3** (2021), 10.1103/physrevresearch.3.033251.

[58] N. Ezzell, E. M. Ball, A. U. Siddiqui, M. M. Wilde, A. T. Sornborger, P. J. Coles, and Z. Holmes, "Quantum mixed state compiling," Quantum Science and Technology **8**, 035001 (2023).

[59] J. C. Garcia-Escartin and P. Chamorro-Posada, "swap test and hong-ou-mandel effect are equivalent," Physical Review A **87** (2013), 10.1103/phys-

reva.87.052330.

[60] S. Johri, D. S. Steiger, and M. Troyer, "Entanglement spectroscopy on a quantum computer," Physical Review B **96** (2017), 10.1103/physrevb.96.195136.

[61] L. Cincio, Y. Subaşı, A. T. Sornborger, and P. J. Coles, "Learning the quantum algorithm for state overlap," New Journal of Physics **20**, 113022 (2018).

[62] Y. Subaşı, L. Cincio, and P. J. Coles, "Entanglement spectroscopy with a depth-two quantum circuit," Journal of Physics A: Mathematical and Theoretical **52**, 044001 (2019).

[63] A. Elben, S. T. Flammia, H.-Y. Huang, R. Kueng, J. Preskill, B. Vermersch, and P. Zoller, "The randomized measurement toolbox," Nature Reviews Physics **5**, 9–24 (2022).

[64] H. Kobayashi, K. Matsumoto, and T. Yamakami, "Quantum merlin-arthur proof systems: Are multiple merlins more helpful to arthur?" in *Lecture Notes in Computer Science* (Springer Berlin Heidelberg, 2003) p. 189–198.

[65] Chi-Fang, Chen, M. J. Kastoryano, F. G. S. L. Brandão, and A. Gilyén, "Quantum thermal state preparation," arXiv (2023), 10.48550/arXiv.2303.18224.

[66] S. McArdle, T. Jones, S. Endo, Y. Li, S. C. Benjamin, and X. Yuan, "Variational ansatz-based quantum simulation of imaginary time evolution," npj Quantum Information **5** (2019), 10.1038/s41534-019-0187-2.

[67] S. Endo, J. Sun, Y. Li, S. C. Benjamin, and X. Yuan, "Variational quantum simulation of general processes," Physical Review Letters **125** (2020), 10.1103/physrevlett.125.010501.

[68] J.-G. Liu, L. Mao, P. Zhang, and L. Wang, "Solving quantum statistical mechanics with variational autoregressive networks and quantum circuits," Machine Learning: Science and Technology **2**, 025011 (2021).

[69] M. Motta, C. Sun, A. T. K. Tan, M. J. O'Rourke, E. Ye, A. J. Minnich, F. G. S. L. Brandão, and G. K.-L. Chan, "Determining eigenstates and thermal states on a quantum computer using quantum imaginary time evolution," Nature Physics **16**, 205–210 (2019).

[70] B.-B. Chen, L. Chen, Z. Chen, W. Li, and A. Weichselbaum, "Exponential thermal tensor network approach for quantum lattice models," Physical Review X **8** (2018), 10.1103/physrevx.8.031082.

[71] L. Coopmans, Y. Kikuchi, and M. Benedetti, "Predicting gibbs-state expectation values with pure thermal shadows," PRX Quantum **4** (2023), 10.1103/prxquantum.4.010305.

[72] S. Sugiura and A. Shimizu, "Thermal pure quantum states at finite temperature," Physical Review Letters **108** (2012), 10.1103/physrevlett.108.240401.

[73] S. Sugiura and A. Shimizu, "Canonical thermal pure quantum state," Physical Review Letters **111** (2013), 10.1103/physrevlett.111.010401.

[74] M. Ohliger, V. Nesme, and J. Eisert, "Efficient and feasible state tomography of quantum many-body systems," New Journal of Physics **15**, 015024 (2013).

[75] S. Aaronson, "Shadow tomography of quantum states," in *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, STOC '18 (ACM, 2018).

[76] H.-Y. Huang, R. Kueng, and J. Preskill, "Predicting many properties of a quantum system from very few measurements," Nature Physics **16**, 1050–1057 (2020).

[77] H.-Y. Huang, R. Kueng, and J. Preskill, "Efficient estimation of pauli observables by derandomization," Physical Review Letters **127** (2021), 10.1103/physrevlett.127.030503.

[78] T. Zhang, J. Sun, X.-X. Fang, X.-M. Zhang, X. Yuan, and H. Lu, "Experimental quantum state measurement with classical shadows," Physical Review Letters **127** (2021), 10.1103/physrevlett.127.200501.

[79] C. Hadfield, S. Bravyi, R. Raymond, and A. Mezzacapo, "Measurements of quantum hamiltonians with locally-biased classical shadows," Communications in Mathematical Physics **391**, 951–967 (2022).

[80] M. J. Kastoryano and K. Temme, "Quantum logarithmic sobolev inequalities and rapid mixing," Journal of Mathematical Physics **54** (2013), 10.1063/1.4804995.

[81] K. Temme, "Lower bounds to the spectral gap of davies generators," Journal of Mathematical Physics **54** (2013), 10.1063/1.4850896.

[82] M. J. Kastoryano and F. G. S. L. Brandao, "Quantum gibbs samplers: the commuting case," (2014).

[83] Á. Capel, C. Rouzé, and D. S. França, "The modified logarithmic sobolev inequality for quantum spin systems: classical and commuting nearest neighbour interactions," (2020).

[84] C.-F. Chen and F. G. S. L. Brandão, "Fast thermalization from the eigenstate thermalization hypothesis," (2021).

[85] D. S. Wild, D. Sels, H. Pichler, C. Zanoci, and M. D. Lukin, "Quantum sampling algorithms, phase transitions, and computational complexity," Physical Review A **104** (2021), 10.1103/physreva.104.032602.

[86] A. Kitaev, A. Shen, and M. Vyalyi, *Classical and Quantum Computation* (American Mathematical Society, 2002).

[87] J. Kempe and O. Regev, "3-local hamiltonian is qma-complete," arXiv (2003), 10.48550/arXiv.quant-ph/0302079.

[88] J. Kempe, A. Kitaev, and O. Regev, "The complexity of the local hamiltonian problem," in *Lecture Notes in Computer Science* (Springer Berlin Heidelberg, 2004) p. 372–383.

[89] M. Szegedy, "Quantum speed-up of markov chain based algorithms," in *45th Annual IEEE Symposium on Foundations of Computer Science* (IEEE, 2004).

[90] P. Wocjan and K. Temme, "Szegedy walk unitaries for quantum maps," arXiv (2021), 10.48550/arXiv.2107.07365.

[91] D. Nagaj, P. Wocjan, and Y. Zhang, "Fast amplification of qma," arXiv (2009), 10.48550/arXiv.0904.1549.

[92] N. Neumann and S. den Breeijen, "Limitations of clustering using quantum persistent homology," arXiv (2019), 10.48550/arXiv.1911.10781.