Research Article Vol. 5, No. 2 / February 2018 / Optica 193

Demonstration of topological data analysis on a quantum processor

He-Liang Huang,[1,2,3,4] Xi-Lin Wang,[1,2,3] Peter P. Rohde,[5] Yi-Han Luo,[1,2,3] You-Wei Zhao,[1,2,3] Chang Liu,[1,2,3] Li Li,[1,2,3,*] Nai-Le Liu,[1,2,3] Chao-Yang Lu,[1,2,3] and Jian-Wei Pan[1,2,3]

[1]Hefei National Laboratory for Physical Sciences at Microscale and Department of Modern Physics, University of Science and Technology of China, Hefei, Anhui 230026, China
[2]CAS Centre for Excellence and Synergetic Innovation Centre in Quantum Information and Quantum Physics, University of Science and Technology of China, Hefei, Anhui 230026, China
[3]CAS-Alibaba Quantum Computing Laboratory, Shanghai 201315, China
[4]Henan Key Laboratory of Quantum Information and Cryptography, Zhengzhou, Henan 450000, China
[5]Centre for Quantum Software & Information (QSI), Faculty of Engineering & Information Technology, University of Technology Sydney, NSW 2007, Australia
*Corresponding author: eidos@ustc.edu.cn

Received 28 November 2017; revised 17 January 2018; accepted 18 January 2018 (Doc. ID 314506); published 12 February 2018

Topological data analysis offers a robust way to extract useful information from noisy, unstructured data by identifying its underlying structure. Recently, an efficient quantum algorithm was proposed [Nat. Commun. 7, 10138 (2016)] for calculating Betti numbers of data points—topological features that count the number of topological holes of various dimensions in a scatterplot. Here, we implement a proof-of-principle demonstration of this quantum algorithm by employing a six-photon quantum processor to successfully analyze the topological features of Betti numbers of a network including three data points, providing new insights into data analysis in the era of quantum computing.

OCIS codes: (270.0270) Quantum optics; (270.5585) Quantum information and processing.

https://doi.org/10.1364/OPTICA.5.000193

1. INTRODUCTION

In exploratory data analysis and data mining, our data often encodes extremely valuable information but is typically large, unstructured, noisy, and incomplete, such that extracting useful information from the data is an important yet challenging task. Topological data analysis (TDA) [1] provides a general framework for studying such data in a manner that is insensitive to the particular metric and robust against noise. In particular, persistent homology [2,3] has been well established as a technique for extracting useful information by identifying topological features of data. One essential feature is the number of k-dimensional holes and voids in datasets, that is, the kth Betti number β_k (a topological invariant). For instance, the first three Betti numbers, β_0, β_1, and β_2, represent, respectively, the number of connected components, one-dimensional holes, and two-dimensional voids. The Betti numbers abstract away the actual data, reducing it to a purely topological representation, which is valuable for understanding the underlying structure of datasets. The field of using topological data analysis to analyze Betti numbers of data has been growing rapidly in recent years, yielding applications in image recognition [4], signal processing [5], network science [6,7], sensor analysis [8–11], brain connectomics [12,13], and fMRI data analysis [14,15], just to name a few.

Practically, however, when facing the issue of computational complexity, classical topological methods pose a formidable task: a set of n data points possesses 2^n potential subsets that could contribute to the topology, quickly overwhelming even the most powerful classical computers, even for not-so-large datasets. So far, the best classical algorithm for estimating Betti numbers to all orders with accuracy δ takes time $O(2^n \log(1/\delta))$ [16–21]. Moreover, exact calculation of Betti numbers is known to be PSPACE-hard for some classes of topologies [22].

Recently, Lloyd *et al.* [23,24] extended methods from quantum machine learning to TDA for efficiently estimating Betti numbers to all orders. Indeed, if the proportion of k-simplices generated from a dataset is large enough, the quantum algorithm for calculating Betti numbers to all orders with accuracy δ has runtime $O(n^5/\delta)$—exponentially faster than the best known classical algorithms. Furthermore, the algorithm does not require a large-scale quantum random access memory (qRAM) [25]—just $O(n^2)$ bits is sufficient for the algorithm to store the information of all pairwise distances between the n data points. The potential computational speedup and its practicality will likely make quantum TDA a promising application for future quantum computers, in addition to Shor's algorithm [26–29], quantum

simulation [30–33], solving linear systems [34,35], and classification of linear vectors [36–38].

Here we report a proof-of-principle demonstration of the quantum TDA algorithm on a small-scale photonic quantum processor for the first time, to our knowledge. The topological features of Betti numbers of three data points are revealed and monitored at two different topological scales in our experiment. Our experiment successfully demonstrates the viability of the algorithm and suggests that data analytics may be an important future application for quantum computing, with widespread applications in our increasingly data-centric world.

2. THEORY

To calculate Betti numbers, we first represent data topologically in terms of relationships between data points. Using a cutoff distance ϵ, we group data points into *simplices* [see Fig. 1(a)]—fully-connected subsets of data points. The set of simplices forms a *simplicial complex*, the topological structure from which features such as Betti numbers can be extracted. This topological construction is shown in Figs. 1(b)–1(d).

By determining the complete set of Betti numbers over the full range of ϵ, we can then construct the *barcode* [see Fig. 1(e)] [39], a parameterized version of Betti numbers in a distance-dependent manner. Each bar in the region of $\mathbf{H}_k$ represents a k-dimensional hole, and the length of the bar indicates its persistence in the parameter ϵ. With the barcode, we can qualitatively filter out the short bars as topological noise and capture the long bars as significant features, since the length of bars is indicative of their persistence against changes in distance ϵ. In Fig. 1(e), a bar in the region of $\mathbf{H}_1$ persists for a long range, leading us to determine that the underlying topological feature of the unstructured data [Fig. 1(b)] is a circle.

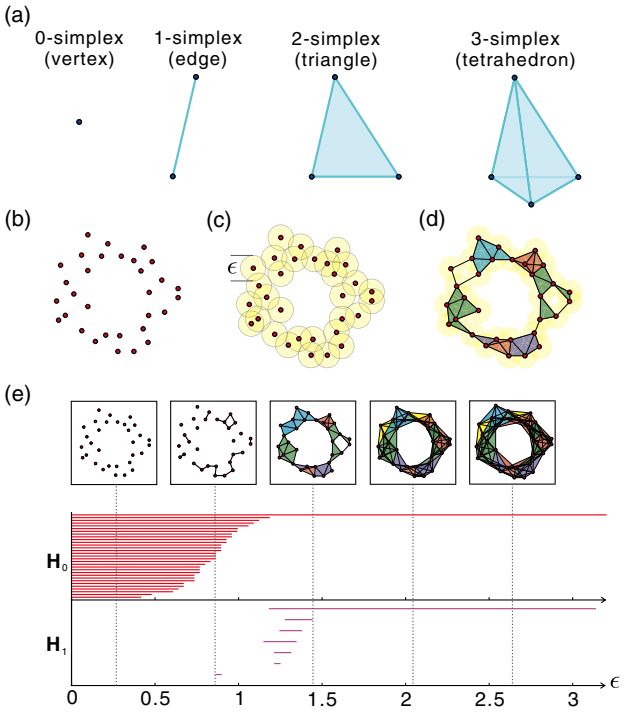

Fig. 1. (a) k-simplices (shown for $k = 0, 1, 2, 3$) are fully connected sets of $k + 1$ data points. (b) Scatterplot of data points. (c) Using some arbitrary metric for quantifying *distance* ϵ between data points, data points within ϵ of one another receive an edge between them. (d) The simplicial complex is formed as the set of simplices. The colored regions indicate the different simplices within the complex. (e) Construction of the barcode. The horizontal axis represents the distance ϵ. The bars are constructed such that the number of bars that intersect the vertical line through any ϵ in the area of $\mathbf{H}_k$ equals the Betti number β_k.

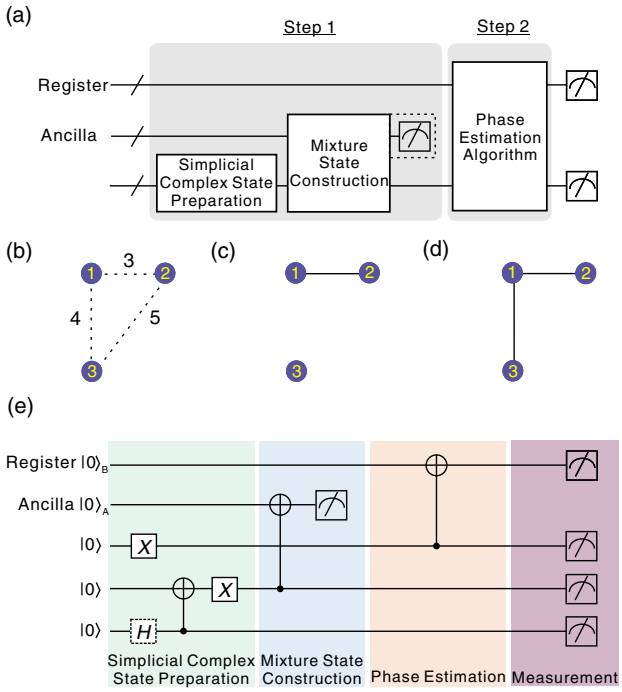

Fig. 2. Quantum circuit for quantum TDA. (a) Outline of the original quantum circuit. (b) A scatterplot including three data points. (c) Graph representation of the 1-simplices state $|\varphi\rangle_1^{\epsilon_1} = |110\rangle$ for $3 < \epsilon_1 < 4$. The first and second data points are connected by an edge. (d) Graph representation of 1-simplices state $|\varphi\rangle_1^{\epsilon_2} = (|110\rangle + |101\rangle)/\sqrt{2}$ for $4 < \epsilon_2 < 5$. The first data point is connected to the second and third points by two edges. (e) Optimized circuit with 5 qubits. The blocks with different colors represent the four basic stages.

In general, the quantum TDA algorithm has two main steps [see Fig. 2(a)]. First, one accesses the data to construct the uniform mixture of the k-simplices that encode the desired topological structure. The time of this step is in the worst case exponential and in fact depends on the proportion of k-simplices. In cases where this fraction is large enough, this step can be implemented efficiently either classically or by using Grover's algorithm, yielding a further quadratic algorithmic enhancement. In the quantum algorithm, this step could be realized via two small steps, namely, (1a) simplicial complex state preparation; (1b) uniform mixed state construction. Second, one implements Step (2) to reveal the topological invariants of the structure. This step is realized using the phase-estimation algorithm [40], which provides an exponential speedup over known classical procedures on a quantum computer; in fact [23,24], showed that this can be executed in time $O(n^5/\delta)$, with accuracy δ. The steps of the quantum algorithm are now described in more detail.

Implementing Step (1a) constructs the simplicial complex. For a scatterplot including n data points, a k-simplex s_k consists of $k + 1$ points $V_{j_0}, V_{j_1}, \ldots, V_{j_k}$, together with $k(k + 1)/2$ edges,

creating a fully connected subset of the data. We can encode a k-simplex as an n-qubit quantum state $|s_k\rangle$ with $k+1$ 1s at positions $j_0, j_1, \ldots, j_k$ and 0s at the other remaining positions.

The Vietoris–Rips simplicial complex S_k^ϵ is the set of k-simplices where all points are within distance ϵ of each other. In the quantum implementation, we can construct the simplicial complex state $|\psi\rangle_k^\epsilon$ as the uniform superposition of k-simplices in the complex:

$$|\psi\rangle_k^\epsilon = \frac{1}{\sqrt{|S_k^\epsilon|}} \sum_{s_k \in S_k^\epsilon} |s_k\rangle. \tag{1}$$

Classically verifying whether all points in each of the s_k are within distance ϵ of each other could help us construct the simplicial complex state. We can also implement a multi-target Grover algorithm [41] with a membership oracle function $\{f_k^\epsilon(s_k) = 1 \text{ if } s_k \in S_k^\epsilon\}$ to verify whether $s_k \in S_k^\epsilon$, yielding a quadratic speedup. Let H_k^ϵ be the Hilbert space spanned by $|s_k\rangle$, where $s_k \in S_k^\epsilon$. The construction of $|\psi\rangle_k^\epsilon$ also reveals the number of k-simplices, $|S_k^\epsilon| = \dim H_k^\epsilon$, and takes time $O(n^2(\zeta_k^\epsilon)^{-1/2})$, where $\zeta_k^\epsilon = |S_k^\epsilon|/\binom{n}{k+1}$ is the proportion of k-simplices that are actually in this complex at scale ϵ, and $(\zeta_k^\epsilon)^{-1/2} = (|S_k^\epsilon|/\binom{n}{k+1})^{-1/2}$ is the number of iterations of the multi-target Grover algorithm. When the proportion is too small, the quantum search procedure will fail to find the simplices [23,24].

In Step (1b), we construct the mixed state

$$\rho_k^\epsilon = \frac{1}{|S_k^\epsilon|} \sum_{s_k \in S_k^\epsilon} |s_k\rangle\langle s_k|, \tag{2}$$

the uniform mixture over the set of simplices in the complex. This procedure can be easily realized by adding an n-qubit ancillary register, performing controlled-NOT (CNOT) operations to copy $|\psi\rangle_k^\epsilon$ to construct $\frac{1}{\sqrt{|S_k^\epsilon|}} \sum_{s_k \in S_k^\epsilon} |s_k\rangle \otimes |s_k\rangle$, and, finally, tracing out the ancillary register to obtain ρ_k^ϵ.

Step (2) acts on the simplicial complex to reveal topological features—the core of exponential speedup in the algorithm. Define the boundary map ∂_k^ϵ that operates from H_k^ϵ to H_{k-1}^ϵ by

$$\partial_k^\epsilon |s_k\rangle = \sum_l (-1)^l |s_{k-1}(l)\rangle, \tag{3}$$

where $|s_{k-1}(l)\rangle$ is obtained from s_k with vertices $j_0 \ldots j_l \ldots j_k$ by omitting the lth point j_l from s_k. The kth Betti number is defined as [17–20]

$$\beta_k^\epsilon = \dim(\mathrm{Ker}\partial_k^\epsilon / \mathrm{Im}\partial_{k+1}^\epsilon). \tag{4}$$

Classical algorithms for calculating Betti numbers to all orders with accuracy δ require time $O(2^n \log(1/\delta))$ [16–21]. In quantum TDA, an exponential speedup is achieved by employing the phase-estimation algorithm. For this purpose, the boundary map is embedded into a Hermitian matrix:

$$B_k^\epsilon = \begin{pmatrix} 0 & \partial_k^\epsilon \\ \partial_k^{\epsilon\dagger} & 0 \end{pmatrix}. \tag{5}$$

Now, by applying phase estimation to decompose ρ_k^ϵ in terms of the eigenvectors and eigenvalues of B_k^ϵ, one obtains the probability η_k^ϵ of projecting onto the kernel by measuring the eigenvalue register. Then the dimension of the kernel of ∂_k^ϵ can be calculated as $\dim(\mathrm{Ker}\partial_k^\epsilon) = \eta_k^\epsilon \cdot |S_k^\epsilon|$. When both $\dim(\mathrm{Ker}\partial_k^\epsilon)$ and $\dim(\mathrm{Ker}\partial_{k+1}^\epsilon)$ are determined, we can reconstruct the kth Betti number by

$$\begin{aligned} \beta_k^\epsilon &= \dim(\mathrm{Ker}\partial_k^\epsilon) - \dim(\mathrm{Im}\partial_{k+1}^\epsilon) \\ &= \dim(\mathrm{Ker}\partial_k^\epsilon) + \dim(\mathrm{Ker}\partial_{k+1}^\epsilon) - |S_{k+1}^\epsilon|. \end{aligned} \tag{6}$$

We note that, for some special cases for ∂_k^ϵ, it is trivial to calculate $\dim(\mathrm{Ker}\partial_k^\epsilon)$. For example, if a k-simplex does not exist, $\dim(\mathrm{Ker}\partial_k^\epsilon) = |S_k^\epsilon| = 0$, while $\dim(\mathrm{Ker}\partial_0)$ is always equal to the number of points.

Careful evaluation indicates that Step (2) can estimate Betti numbers to all orders with accuracy δ in time $O(n^5/\delta)$ [23,24]. Hence, while in the worst case that their proportion is too small, Step (1) will fail to find the k-simplices, since both the classical and quantum algorithm will take exponential time. There are specific cases, in particular where Step (1) can be implemented efficiently, where the overall quantum algorithm can provide exponential savings. In fact, we have tested a particular case using data points with random distances between them and showed that indeed Step (1) can be implemented efficiently (see Supplement 1 for details), either by a classical algorithm or further improving the time by a square root factor through Grover's algorithm.

3. EXPERIMENTAL IMPLEMENTATION

To experimentally demonstrate the quantum TDA algorithm, we choose the simplest meaningful instance: estimating the Betti numbers for three data points at two different scales. Assume the distances between the three points are 3, 4, and 5 [see Fig. 2(b)]. For scales in the ranges $3 < \epsilon_1 < 4$ and $4 < \epsilon_2 < 5$, the corresponding states for 1-simplices (the k-simplex for $k > 1$ does not exist since not all three data points can be connected at ϵ_1 and ϵ_2) are $|\varphi\rangle_1^{\epsilon_1} = |110\rangle$ [Fig. 2(c)] and $|\varphi\rangle_1^{\epsilon_2} = (|110\rangle + |101\rangle)/\sqrt{2}$ [Fig. 2(d)], respectively, which means $|S_1^{\epsilon_1}| = 1$ and $|S_1^{\epsilon_2}| = 2$. A simple quantum circuit is designed to prepare $|\varphi\rangle_1^{\epsilon_1}$ ($|\varphi\rangle_1^{\epsilon_2}$) directly by removing (adding) a Hadamard gate marked by dashed lines at Step (1) in Fig. 2(e).

To construct the corresponding uniform mixed states, we do not actually need to generate a complete copy of $|\varphi\rangle_1^{\epsilon_1}$ ($|\varphi\rangle_1^{\epsilon_2}$). Instead, we need only perform a CNOT operation between the auxiliary qubit $|0\rangle_A$ and the second qubit of $|\varphi\rangle_1^{\epsilon_1}$ ($|\varphi\rangle_1^{\epsilon_2}$) to partially copy the state of simplices. After tracing out the ancillary qubit, the uniform mixed states ρ^{ϵ_1} and ρ^{ϵ_2} are obtained.

Next, we apply quantum phase estimation to reveal information related to Betti numbers. Since there are only three data points, k-dimensional holes for $k > 1$ cannot exist. Therefore, only the 0th and 1st Betti numbers need to be calculated. We note that the algorithm cares not about the exact eigenvalue spectrum, but the probability of detecting $|0\rangle$ in the eigenvalue register. We can exploit this property to reduce the number of qubits required in the eigenvalue register. A particular treatment for boundary matrices is utilized to greatly simplify the complex circuit (see Supplement 1 for details)—a single CNOT operation between the eigenvalue register comprising only one qubit $|0\rangle_B$ and the first bit of ρ^{ϵ_1} (ρ^{ϵ_2}) is sufficient for realizing phase estimation. Finally, the information related to Betti numbers will be read out by measuring the eigenvalue register. Note that since the quantum TDA algorithm depends only on how the points are connected, not the precise distances between points, our circuit works for all nontrivial cases of three points (where one or two edges are present). The cases where zero or three edges are present are trivial, since we could clearly know the Betti numbers in the cases that the N points are all disconnected ($\beta_0 = N$, and $\beta_k = 0$

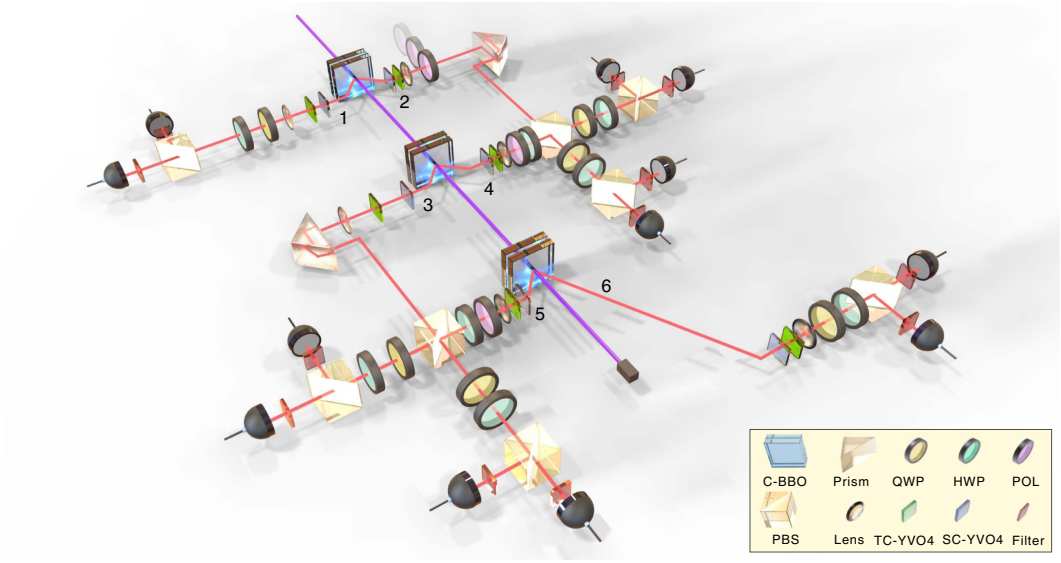

Fig. 3. Experimental setup. Ultraviolet laser pulses with a central wavelength of 394 nm, pulse duration of 150 fs, and repetition rate of 80 MHz pass through three HWP-sandwiched β-barium borate (BBO) crystals [42] to produce three entangled photon pairs $(|H\rangle|V\rangle + |V\rangle|H\rangle)/\sqrt{2}$ (see Supplement 1 for details) in spatial modes 1-2, 3-4, and 5-6. Photons 2(3) and 4(5) are temporally and spatially superposed on a PBS. All photons are spectrally filtered with 3 nm bandwidth filters. C-BBO, sandwich-like BBO + HWP + BBO combination; QWP, quarter-wave plate; POL, polarizer; SC-YVO4, YVO_4 crystal for spatial compensation; TC-YVO4, YVO_4 crystal for temporal compensation.

for $k > 0$) or all connected ($\beta_0 = 1$, and $\beta_k = 0$ for $k > 0$) for N points without calculating.

Figure 3 shows the setup of our experiment. We use single photons as qubits, where the logical qubits $|0\rangle$ and $|1\rangle$ are encoded into horizontal (H) and vertical (V) polarization, respectively. With these settings, the step of simplices state preparation becomes straightforward. $|\varphi\rangle_1^{\epsilon_1} = |V\rangle_3|V\rangle_2|H\rangle_1$ and $|\varphi\rangle_1^{\epsilon_2} = (|V\rangle_3|V\rangle_2|H\rangle_1 + |V\rangle_3|H\rangle_2|V\rangle_1)/\sqrt{2}$ can be prepared directly by adding or removing the polarizer in path 2, respectively, where the index i in $|H(\text{or } V)\rangle_i$ denotes the spatial mode. Photons 4 (ancilla) and 5 (eigenvalue register) are both disentangled by polarizers into $|H\rangle$, and then photons 3 and 6 (trigger) immediately collapse into $|V\rangle$. Note that the CNOT gates can be simulated using combinations of a polarizing beam splitter (PBS) and a half-wave plate (HWP) [27], since the target qubits are fixed at $|H\rangle$. This setup, in principle, suffices to demonstrate the underlying conceptual principles of quantum TDA.

Before running the algorithm, we first characterized the performance of the optical quantum circuit. In the case of $4 < \epsilon_2 < 5$, a three-photon entangled state $|\phi\rangle = (|H\rangle_1|V\rangle_2|V\rangle_4 + |V\rangle_1|H\rangle_2|H\rangle_4)/\sqrt{2}$ is generated after implementing the CNOT gate in Step (1a). We measured the fidelity of the experimentally prepared state (see Supplement 1 for details) as $F = 0.954(6)$, which exceeds the threshold of 0.5 for the entanglement witness to confirm genuine multi-partite entanglement [43]. To the best of our knowledge, such a high fidelity for three-photon entanglement has never been achieved before [44].

After tracing out the ancilla in the Pauli-Z basis, the uniform mixed states

$$\begin{aligned}\rho^{\epsilon_1} &= |V\rangle_3|V\rangle_2|H\rangle_1\langle V|_3\langle V|_2\langle H|_1,\\ \rho^{\epsilon_2} &= (|V\rangle_3|V\rangle_2|H\rangle_1\langle V|_3\langle V|_2\langle H|_1\\ &\quad + |V\rangle_3|H\rangle_2|V\rangle_1\langle V|_3\langle H|_2\langle V|_1)/2 \end{aligned} \tag{7}$$

are created at the scales of $3 < \epsilon_1 < 4$ and $4 < \epsilon_2 < 5$, respectively. We characterized these states using quantum state tomography to reconstruct the density matrices (see Supplement 1 for details.) The fidelity $F_p = \left(\text{Tr}\sqrt{\rho^{1/2}\rho_{\text{exp}}\rho^{1/2}}\right)^2$ and trace distance $D(\rho, \rho_{\text{exp}}) = \text{Tr}|\rho - \rho_{\text{exp}}|/2$ between the reconstructed (ρ_{exp}) and ideal (ρ) matrices were calculated as $F_p^{\rho^{\epsilon_1}} = 0.9817(9)$, $F_p^{\rho^{\epsilon_2}} = 0.9819(10)$ and $D_{\rho^{\epsilon_1}} = 0.0183(5)$, $D_{\rho^{\epsilon_2}} = 0.0181(5)$, respectively. Furthermore, the fidelity F_p and trace distance D are related by the inequality $1 - \sqrt{F_p} \le D \le \sqrt{1 - F_p}$ [40]. In our experiment, both $D_{\rho^{\epsilon_1}}$ and $D_{\rho^{\epsilon_2}}$ are located in the range of $0.009 \le D \le 0.135$ and close to the lower bound.

4. EXPERIMENTAL RESULTS

The final results were read out via sixfold coincidence events. Figures 4(a) and 4(b) show the measurement results of the eigenvalue register at the scales of $3 < \epsilon_1 < 4$ and $4 < \epsilon_2 < 5$, respectively. In the case of $3 < \epsilon_1 < 4$, with a probability of $\eta_1^{\epsilon_1} = 0.045(14)$, we measure $|0\rangle$ in the eigenvalue register, from which we calculate the dimension of the kernel space as $\dim(\text{Ker}\partial_1^{\epsilon_1}) = \eta_1^{\epsilon_1} \cdot |S_1^{\epsilon_1}| = 0.045(14)$. Since $\dim(\text{Ker}\partial_0^{\epsilon}) = 3$ and $\dim(\text{Ker}\partial_2^{\epsilon}) = |S_2^{\epsilon}| = 0$ for $\epsilon = \epsilon_1(\text{or}\epsilon_2)$, we finally obtain the 0th Betti number $\beta_0^{\epsilon_1} = 2.045(14)$ and 1st Betti number $\beta_1^{\epsilon_1} = 0.045(14)$, following Eq. (4), which can be rounded to $\beta_0^{\epsilon_1} = 2$ and $\beta_1^{\epsilon_1} = 0$. In the case of $4 < \epsilon_2 < 5$, the probability of measuring $|0\rangle$ in the eigenvalue register is 0.038(9). Using the same approach, we calculate the 0th and 1st Betti numbers as $\beta_0^{\epsilon_2} = 1.076(18)$ and $\beta_1^{\epsilon_2} = 0.038(9)$, respectively, which can be rounded to $\beta_0^{\epsilon_2} = 1$ and $\beta_1^{\epsilon_2} = 0$. That is to say, we have revealed and tracked the topological features of the dataset in Fig. 2(b) at two different scales: the number of connected components at scales of ϵ_1 and ϵ_2 are 2 and 1, respectively, and no k-dimensional holes for $k > 1$ exist. From these results, the barcode is constructed as shown in Fig. 4(c).

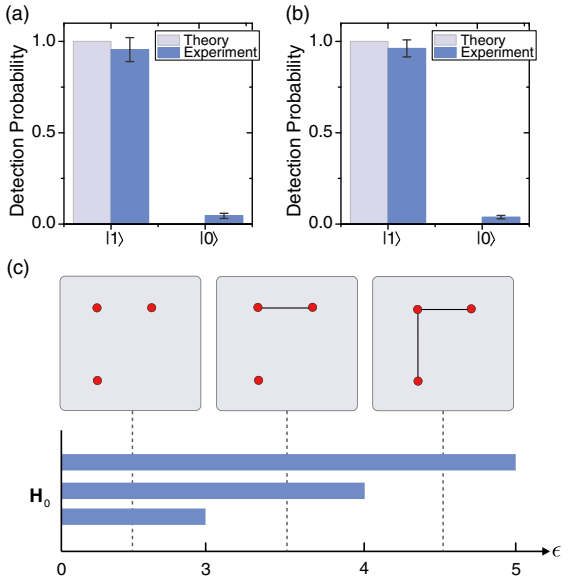

Fig. 4. Final experimental results. The output is determined by measuring the eigenvalue register in the Pauli-Z basis. Measured expectation values (blue bars) and theoretically predicted values (gray bars) are shown for two different 1-simplices state inputs: (a) $|\varphi\rangle_1^{\epsilon_1} = |110\rangle$, and (b) $|\varphi\rangle_1^{\epsilon_2} = (|110\rangle + |101\rangle)/\sqrt{2}$. Error bars represent one standard deviation, deduced from propagated Poissonian counting statistics of the raw detection events. (c) The barcode for $0 < \epsilon < 5$. Since no k-dimensional holes for $k \geq 1$ exist at these scales, only the 0th Betti barcode is given here. For $0 < \epsilon < 3$, there is no connection between each point, so the 0th Betti number is equal to the number of points. That is, there are three bars at $0 < \epsilon < 3$. At scales of $3 < \epsilon_1 < 4$ and $4 < \epsilon_2 < 5$, the 0th Betti numbers are 2 and 1.

To further quantify the experimental performance, we use the similarity measure $\gamma = (\sum_{k=0}^{1} \sqrt{e_k t_k})^2$ [45] to characterize the overlap between experimental and theoretical values, where e_k and t_k are the experimental and theoretical output probabilities of the state $|k\rangle$, respectively. The data in Fig. 4 shows the results as $\gamma_{\epsilon_1} = 0.955(3)$ and $\gamma_{\epsilon_2} = 0.962(2)$, indicating near perfect experimental accuracy, confirming that the algorithm is successful.

5. DISCUSSION

We note that for the quantum TDA algorithm, the results are read out by measuring the eigenvalues. In general, the eigenvalue register requires only a few qubits for the quantum TDA algorithm (1 qubit in the current work), since we care about only the proportion of $|0\rangle$ in the eigenvalue register, rather than the exact value of all eigenvalues. Thus, a small amount of measurements is sufficient for obtaining reliable results, an important feature for the scalability of the algorithm.

In addition, theoretically, for the quantum TDA algorithm, only the qubits in the eigenvalue register need to be measured, rather than having to measure all qubits. In our experiment, since the photons generated by SPDC are probabilistic, to ensure that all qubits in the circuit have been generated, and the quantum circuits have been fully implemented, we need to measure sixfold coincidence events. In fact, this is a common problem encountered in the current linear optical quantum computing. Fortunately, with the development of the deterministic quantum dot single-photon source [46], and other techniques [47], we believe this problem can eventually be overcome. We anticipate that with more qubits (more photons [42,48] or higher dimensional states [49,50]), our proposal could be extended to the analysis of much larger datasets in the future.

6. CONCLUSION

In summary, we have presented what we believe to be the first proof-of-principle demonstration of quantum TDA on a small-scale photonic quantum processor. The topological features of a dataset comprising three data points are revealed and tracked at two different topological scales, fully reproducing the Betti numbers associated with the topology of the data. Future advances in the field could open up new frontiers in data analysis for quantum computing, including signal and image analysis, astronomy, network and social media analysis, behavioral dynamics, biophysics, oncology, and neuroscience.

Funding. National Natural Science Foundation of China (NSFC); Chinese Academy of Sciences (CAS); National Fundamental Research Program; Australian Research Council (ARC) Future Fellowship (FT160100397).

Acknowledgment. We thank R.-Z. Liu, Michele Cirafici, and T. L. for enlightening discussions.

See Supplement 1 for supporting content.

REFERENCES

1. G. Carlsson, "Topology and data," Bull. Am. Math. Soc. **46**, 255–308 (2009).
2. H. Edelsbrunner, D. Letscher, and A. Zomorodian, "Topological persistence and simplification," Discrete Comput. Geom. **28**, 511–533 (2002).
3. A. Zomorodian and G. Carlsson, "Computing persistent homology," Discrete Comput. Geom. **33**, 249–274 (2005).
4. G. Carlsson, T. Ishkhanov, V. De Silva, and A. Zomorodian, "On the local behavior of spaces of natural images," Int. J. Comput. Vis. **76**, 1–12 (2008).
5. J. A. Perea and J. Harer, "Sliding windows and persistence: an application of topological methods to signal analysis," Found. Comput. Math. **15**, 799–838 (2015).
6. G. Petri, M. Scolamiero, I. Donato, and F. Vaccarino, "Networks and cycles: a persistent homology approach to complex networks," in *Proceedings of the European Conference on Complex Systems* (Springer, 2013), pp. 93–99.
7. G. Petri, M. Scolamiero, I. Donato, and F. Vaccarino, "Topological strata of weighted complex networks," PLoS One **8**, e66506 (2013).
8. V. De Silva and R. Ghrist, "Homological sensor networks," Not. Am. Math. Soc. **54**, 10–17 (2007).
9. V. De Silva and R. Ghrist, "Coverage in sensor networks via persistent homology," Algebr. Geom. Topol. **7**, 339–358 (2007).
10. V. De Silva and G. E. Carlsson, "Topological estimation using witness complexes," in *Proceedings of the First Eurographics Conference on Point-Based Graphics (SPBG)* (2004), Vol. **4**, pp. 157–166.
11. R. Ghrist and A. Muhammad, "Coverage and hole-detection in sensor networks via homology," in *International Symposium on Information Processing in Sensor Networks* (IEEE, 2005), pp. 254–260.
12. C. Giusti, R. Ghrist, and D. S. Bassett, "Two's company, three (or more) is a simplex," J. Comput. Neurosci. **41**, 1–14 (2016).
13. C. Giusti, E. Pastalkova, C. Curto, and V. Itskov, "Clique topology reveals intrinsic geometric structure in neural correlations," Proc. Natl. Acad. Sci. USA **112**, 13455–13460 (2015).
14. G. Petri, P. Expert, F. Turkheimer, R. Carhart-Harris, D. Nutt, P. J. Hellyer, and F. Vaccarino, "Homological scaffolds of brain functional networks," J. R. Soc. Interface **11**, 20140873 (2014).

15. L.-D. Lord, P. Expert, H. M. Fernandes, G. Petri, T. J. Van Hartevelt, F. Vaccarino, G. Deco, F. Turkheimer, and M. L. Kringelbach, "Insights into brain architectures from the homological scaffolds of functional connectivity networks," Front. Syst. Neurosci. **10**, 85 (2016).
16. D. Cohen-Steiner, H. Edelsbrunner, and J. Harer, "Stability of persistence diagrams," Discrete Comput. Geom. **37**, 103–120 (2007).
17. S. Basu, "On bounding the Betti numbers and computing the Euler characteristic of semi-algebraic sets," Discrete Comput. Geom. **22**, 1–18 (1999).
18. S. Basu, "Different bounds on the different Betti numbers of semi-algebraic sets," Discrete Comput. Geom. **30**, 65–85 (2003).
19. S. Basu, "Computing the top Betti numbers of semialgebraic sets defined by quadratic inequalities in polynomial time," Found. Comput. Math. **8**, 45–80 (2008).
20. S. Basu, "Algorithms in real algebraic geometry: a survey," arXiv:1409.1534 (2014).
21. J. Friedman, "Computing Betti numbers via combinatorial Laplacians," Algorithmica **21**, 331–346 (1998).
22. P. Scheiblechner, "On the complexity of deciding connectedness and computing Betti numbers of a complex algebraic variety," J. Complexity **23**, 359–379 (2007).
23. S. Lloyd, S. Garnerone, and P. Zanardi, "Quantum algorithms for topological and geometric analysis of big data," arXiv:1408.3106 (2014).
24. S. Lloyd, S. Garnerone, and P. Zanardi, "Quantum algorithms for topological and geometric analysis of data," Nat. Commun. **7**, 10138 (2016).
25. V. Giovannetti, S. Lloyd, and L. Maccone, "Quantum random access memory," Phys. Rev. Lett. **100**, 160501 (2008).
26. P. W. Shor, "Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer," SIAM J. Comput. **26**, 1484–1509 (1997).
27. C.-Y. Lu, D. E. Browne, T. Yang, and J.-W. Pan, "Demonstration of a compiled version of Shor's quantum factoring algorithm using photonic qubits," Phys. Rev. Lett. **99**, 250504 (2007).
28. B. P. Lanyon, T. J. Weinhold, N. K. Langford, M. Barbieri, D. F. V. James, A. Gilchrist, and A. G. White, "Experimental demonstration of a compiled version of Shor's algorithm with quantum entanglement," Phys. Rev. Lett. **99**, 250505 (2007).
29. H.-L. Huang, Q. Zhao, X. Ma, C. Liu, Z.-E. Su, X.-L. Wang, L. Li, N.-L. Liu, B. C. Sanders, C.-Y. Lu, and J.-W. Pan, "Experimental blind quantum computing for a classical client," Phys. Rev. Lett. **119**, 050503 (2017).
30. R. P. Feynman, "Simulating physics with computers," Int. J. Theor. Phys. **21**, 467–488 (1982).
31. S. Lloyd, "Universal quantum simulators," Science **273**, 1073–1078 (1996).
32. C.-Y. Lu, W.-B. Gao, O. Gühne, X.-Q. Zhou, Z.-B. Chen, and J.-W. Pan, "Demonstrating anyonic fractional statistics with a six-qubit quantum simulator," Phys. Rev. Lett. **102**, 030502 (2009).
33. B. P. Lanyon, J. D. Whitfield, G. G. Gillett, M. E. Goggin, M. P. Almeida, I. Kassal, J. D. Biamonte, M. Mohseni, B. J. Powell, M. Barbieri, A. Aspuru-Guzik, and A. G. White, "Towards quantum chemistry on a quantum computer," Nat. Chem. **2**, 106–111 (2010).
34. A. W. Harrow, A. Hassidim, and S. Lloyd, "Quantum algorithm for linear systems of equations," Phys. Rev. Lett. **103**, 150502 (2009).
35. X.-D. Cai, C. Weedbrook, Z.-E. Su, M.-C. Chen, M. Gu, M.-J. Zhu, L. Li, N.-L. Liu, C.-Y. Lu, and J.-W. Pan, "Experimental quantum computing to solve systems of linear equations," Phys. Rev. Lett. **110**, 230501 (2013).
36. P. Rebentrost, M. Mohseni, and S. Lloyd, "Quantum support vector machine for big data classification," Phys. Rev. Lett. **113**, 130503 (2014).
37. S. Lloyd, M. Mohseni, and P. Rebentrost, "Quantum principal component analysis," Nat. Phys. **10**, 631–633 (2014).
38. X.-D. Cai, D. Wu, Z.-E. Su, M.-C. Chen, X.-L. Wang, L. Li, N.-L. Liu, C.-Y. Lu, and J.-W. Pan, "Entanglement-based machine learning on a quantum computer," Phys. Rev. Lett. **114**, 110504 (2015).
39. R. Ghrist, "Barcodes: the persistent topology of data," Bull. Am. Math. Soc. **45**, 61–75 (2008).
40. M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University, 2010).
41. L. K. Grover, "Quantum mechanics helps in searching for a needle in a haystack," Phys. Rev. Lett. **79**, 325–328 (1997).
42. X.-L. Wang, L.-K. Chen, W. Li, H.-L. Huang, C. Liu, C. Chen, Y.-H. Luo, Z.-E. Su, D. Wu, Z.-D. Li, H. Lu, Y. Hu, X. Jiang, C.-Z. Peng, L. Li, N.-L. Liu, Y.-A. Chen, C.-Y. Lu, and J.-W. Pan, "Experimental ten-photon entanglement," Phys. Rev. Lett. **117**, 210502 (2016).
43. O. Gühne and G. Tóth, "Entanglement detection," Phys. Rep. **474**, 1–75 (2009).
44. D. R. Hamel, L. K. Shalm, H. Hübel, A. J. Miller, F. Marsili, V. B. Verma, R. P. Mirin, S. W. Nam, K. J. Resch, and T. Jennewein, "Direct generation of three-photon polarization entanglement," Nat. Photonics **8**, 801–807 (2014).
45. C. A. Fuchs, "Distinguishability and accessible information in quantum theory," Ph.D. dissertation (University of New Mexico, 1996).
46. Y.-M. He, J. Liu, S. Maier, M. Emmerling, S. Gerhardt, M. Davanço, K. Srinivasan, C. Schneider, and S. Höfling, "Deterministic implementation of a bright, on-demand single-photon source with near-unity indistinguishability via quantum dot imaging," Optica **4**, 802–808 (2017).
47. F. Kaneda, B. G. Christensen, J. J. Wong, H. S. Park, K. T. McCusker, and P. G. Kwiat, "Time-multiplexed heralded single-photon source," Optica **2**, 1010–1013 (2015).
48. H. Wang, Y. He, Y.-H. Li, Z.-E. Su, B. Li, H.-L. Huang, X. Ding, M.-C. Chen, C. Liu, J. Qin, J.-P. Li, Y.-M. He, C. Schneider, M. Kamp, C.-Z. Peng, S. Höfling, C.-Y. Lu, and J.-W. Pan, "High-efficiency multiphoton boson sampling," Nat. Photonics **11**, 361–365 (2017).
49. R. Fickler, R. Lapkiewicz, W. N. Plick, M. Krenn, C. Schaeff, S. Ramelow, and A. Zeilinger, "Quantum entanglement of high angular momenta," Science **338**, 640–643 (2012).
50. X.-L. Wang, X.-D. Cai, Z.-E. Su, M.-C. Chen, D. Wu, L. Li, N.-L. Liu, C.-Y. Lu, and J.-W. Pan, "Quantum teleportation of multiple degrees of freedom of a single photon," Nature **518**, 516–519 (2015).