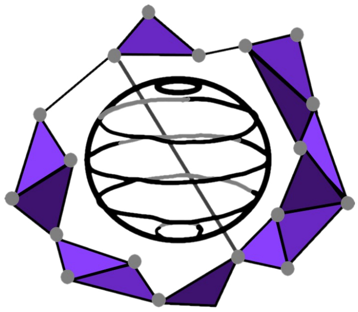

QISKIT HACKATHON EUROPE 2021

Quantum Topological Data Analysis

Andreas Woitzik, Eric Brunner, Kathrin König

andreas.woitzik@physik.uni-freiburg.de, eric.brunner@physik.uni-freiburg.de, kathrin.koenig@iaf.fraunhofer.de

Abstract

Persistent homology looks for topological features (Betti numbers, topological spectra) on all scales arbitrary (high-dimensional) data clouds and how they persist as the scale parameter is varied. Technically, this is done by considering a filtration of simplicial complexes generated from the data points. We implemented a recently suggested quantum algorithm that shows exponential speedup over all classical counterparts in Qiskit. The algorithm shows the expected results for small examples when using a state vector simulator. We show that for an implementation on real hardware a further gate depth optimization is necessary.

1 Motivation

The availability of data on a very large scale leads to a growing interest in information retrieval algorithms and related fields [1]. As modern computers hit physical limits already within the next few decades [2], much attention focuses on quantum hardware, allowing to speed up certain computational tasks, further. Some quite evolved examples for possible quantum speed ups are found in the field of machine learning or data analysis [3, 4].

In data analysis, topological and geometric features of the data play an important role as they unveil interpretable, global information of large scale data clouds [5]. Applications range from the representation of molecular structures in chemistry [6] over image analysis [7] to natural language processing [8].

In our research project, we bring a recently suggested algorithm [9] to life, by implementing it in Qiskit. We investigate its behaviour using both a simulator as well as real quantum hardware.

2 Implementation

Topological features of data clouds can be extracted from filtrating the data by simplicial complexes, which are multi-dimensional combinatorial generalizations of graphs that encode information about the data's shape in different dimension, see fig. 1 (a) and (b). In our project, we implement an efficient quantum algorithm for computing *Betti numbers* β_k, which count the number of holes in the complex on dimension $k > 0$, respectively β_0 counts the number of connected components (see fig. 1 (b)). Technically, these are extracted from the *topological spectra* of the simplical complexes, i.e. the spectra of the *combinatorial Laplacian* Δ_k which is an operator describing diffusion across the simplicial complex in dimension k [10]. By building a series of simplicial complexes from data points for different scale parameter, we obtain a *filtration* of the data (see fig. 1 (c)), which contains information of the *persistence* of topological features at different orders during the scaling. The implemented algorithm shows an exponential speedup over all known classical algorithms [9], and allows the efficient topological analysis of data in terms of persistent homology, once the appropriate quantum hardware is available.

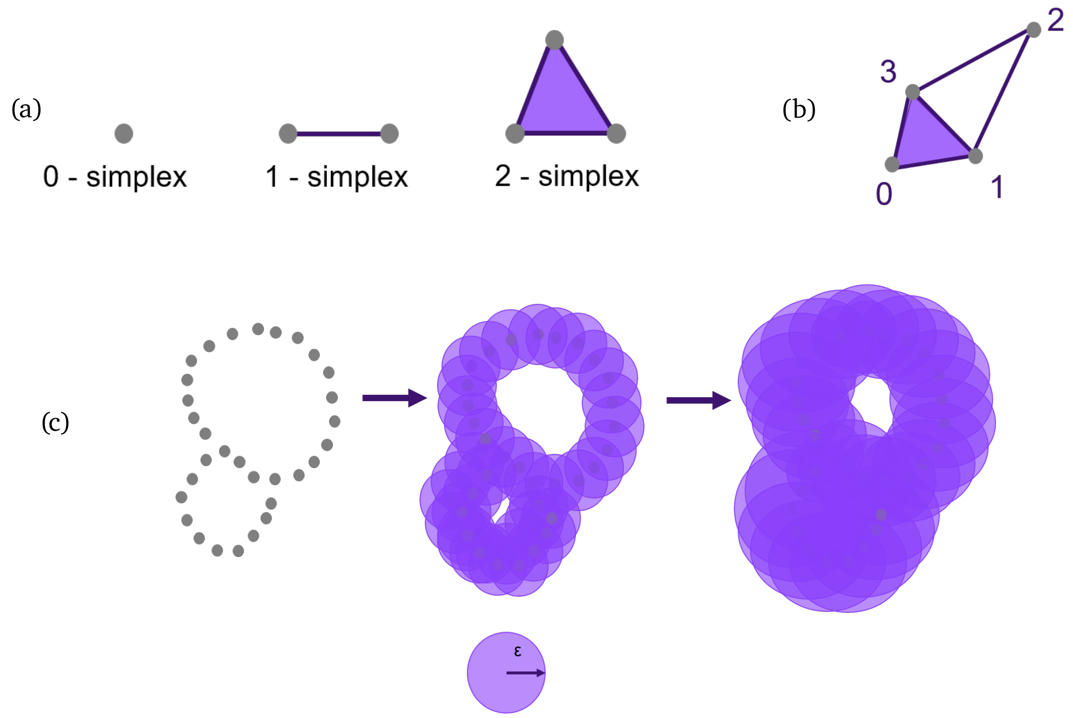

Figure 1: Simplices, simplicial complexes and data filtration: (a) Simplices of different orders: vertices, edges, triangles. (b) Simplical complex on four vertices: The contained simplices are four vertices (0-dim.), five edges (1-dim.) and one triangle (2-dim.). It has one connected component ($\beta_0 = 1$), a single 1-dim. hole surrounded by the edges (1,2),(2,3),(3,1) ($\beta_1 = 1$), and no higher order topological features ($\beta_k = 0,\ k > 1$). If the second (purple) triangle was be empty too, the simplicial complex would have two 1-dim. holes ($\beta_1 = 2$). (c) Data filtration: We draw balls of radius ε around every data point. For any $k+1$ points lying in one ε-ball, we draw the connecting k-simplex. By varying ε, topological features (holes) of different dimension emerge and terminate again, when they are filled by emerging higher dimensional simplices as the scale parameter grows. At $\varepsilon = 0$ we have the $\beta_0^\varepsilon = 31$ connected components (the data points), which are connected via edges, as soon as ε is large enough. At intermediate filtration scale, all points are connected with at least two other points ($\beta_0^\varepsilon = 0$), while the holes in the two circles are still present and not completely filled by 1-simplices ($\beta_1^\varepsilon = 2$). For larger ε the hole in the small circle dies, while the larger hole of the second circle *persists*. At sufficient large ε, the topology of the complex becomes trivial as all possible simplices of the data points are included. Persistent homology looks at the *persistence* of these topological features on different dimensions as the scale parameter is increased.

In a first step, we got familiar with certain Qiskit libraries. Since Kathrin and Andreas work already in the field of quantum computation, we were quickly familiar with the most important libraries. Eric, who already worked himself through the algorithm, gave a course on it and was then taught to use Qiskit. To implement the quantum algorithm, we were closely working together, often coding with two people in order to testify the code as early as possible. Besides coding, we also worked on small examples by hand, to understand the step by step behavior of the algorithm. This thorough study led us also to think a lot about how our results can be shared with the community.

3 Project results

To our knowledge, this is the first work that implements a persistent homology algorithm in Qiskit for a superconducting quantum computer. With this proof of principle we set ground for subsequent work on quantum persistent homology and further quantum routines for topological data analysis. Thereby, this project benefits the whole quantum data science and Qiskit community.

In this contribution we implemented the quantum algorithm for the computation of Betti numbers of arbitrary simplicial complexes, suggested in [9]. It is directly integrated in the Qiskit library as a subclass of the `qiskit.QuantumCircuit` class. This allows for a convenient usage as a building block in further implementation of quantum topological data analysis routines in Qiskit. Additionally, building upon the Python library GUDHI for topological data analysis, we combine this with a filtration routine to generate the full data filtration by means of simplicial complexes. This filtration is fed into the quantum algorithm to obtain the topological spectra of the combinatorial Laplacian Δ_k, and from those the *Betti numbers* β_k, of the data across the full filtration scaling. We verified the capabilities of the algorithm with small datasets using a quantum simulator. Thereafter, we changed the backend to a real quantum machine, the IBM Q Brooklyn for a set of three data points. Our algorithm reached a gate depth of 1225 and therefore made it impossible to derive meaningful results. These findings are summarized in the three Jupyter-notebooks in the project repository.

We developed a pedagogical Jupyter-notebook in Qiskit textbook fashion to introduce the unfamiliar reader to the basic ideas of data filtration, its topological analysis and the functioning of the quantum algorithm by means of small toy examples. These examples hopefully help to understand the structure of the algorithm and the subroutines on an intuitive level. Nonetheless, the full quantum analysis of topological features of the data filtration is provided as a complete input-output based algorithm (easily integrable in Qiskit) for the convenient application to real data sets.

4 Impact-outlook

This project sets ground for several research directions. As the algorithm failed to converge on real quantum hardware, a more hardware efficient ansatz needs to be developed. No optimization of the circuit was conducted due to the limited time horizon of the project, but we see a large potential in circuit optimization.

The quantum phase estimation is a known powerful quantum routine, but so far it withstands an implementation on NISQ devices on medium scale. With quantum error rates dropping, at some point in future quantum error correction schemes may allow us to apply the quantum phase estimation to a large set of qubits. By then, the very favourable scaling of the problem size with the number of qubits allows for data analysis on an unprecedented level.

The k-simplex state $|\Psi_k\rangle$ can also be constructed by Grover's quantum search algorithm, to decide whether a state is included in the complex. Grover's algorithm will further speed up the quantum topological data analysis, although the circuit depth will further increase.

The topological spectra, i.e. the eigen-decomposition of the combinatorial Laplacian Δ_k, contain far more information about the shape of the data then only the number of holes in dimension k [10]. In fact, these also encode geometric information about the data, which is readily available from the implemented quantum algorithm which determines the full spectrum of Δ_k. This opens the perspective to further research questions and applications of the algorithm to topological and geometric data analysis.

From a didactical point of view we created an introductory jupyter notebook that can be used to understand the exponential speedup certain quantum algorithms provide in a geometric, intuitive manner. This work can easily be adapted to a Qiskit textbook chapter.

References

[1] G. G. Chowdhury, *Introduction to modern information retrieval*. Facet publishing, 2010.

[2] M. P. Frank, "Approaching the physical limits of computing," in *35th International Symposium on Multiple-Valued Logic (ISMVL'05)*, pp. 168–185, IEEE, 2005.

[3] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, "Quantum machine learning," *Nature*, vol. 549, no. 7671, pp. 195–202, 2017.

[4] M. Schuld, I. Sinayskiy, and F. Petruccione, "An introduction to quantum machine learning," *Contemporary Physics*, vol. 56, no. 2, pp. 172–185, 2015.

[5] A. Zomorodian and G. Carlsson, "Computing persistent homology," *Discrete & Computational Geometry*, vol. 33, no. 2, pp. 249–274, 2005.

[6] J. Townsend, C. P. Micucci, J. H. Hymel, V. Maroulas, and K. D. Vogiatzis, "Representation of molecular structures with persistent homology for machine learning applications in chemistry," *Nature communications*, vol. 11, no. 1, pp. 1–9, 2020.

[7] W. Bae, J. Yoo, and J. Chul Ye, "Beyond deep residual learning for image restoration: Persistent homology-guided manifold simplification," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 145–153, 2017.

[8] C. S. Pun, K. Xia, and S. X. Lee, "Persistent-homology-based machine learning and its applications– a survey," *arXiv preprint arXiv:1811.00252*, 2018.

[9] S. Lloyd, S. Garnerone, and P. Zanardi, "Quantum algorithms for topological and geometric analysis of data," *Nature communications*, vol. 7, no. 1, pp. 1–7, 2016.

[10] D. Horak and J. Jost, "Spectra of combinatorial Laplace operators on simplicial complexes," *arXiv:1105.2712 [math]*, Nov. 2011. arXiv: 1105.2712.