7th May 2020

**Article for Physical Review B – “Machine learning topological phases in real space”**

Dear Editors,

In this paper we report the discovery of Shannon information entropy signatures from the analysis of real space lattice data in the bulk of two topological condensed matter systems (Su-Schrieffer-Heeger (SSH) systems with nearest and second-nearest neighbor hoppings). Our discovery was made possible by the use of a machine learning algorithm designed for the task of learning phase transitions from local lattice data and with potentially many more applications in data-driven physics.

The Shannon information entropy signatures reported in this paper were discovered by analyzing real space eigenvector data from several thousand Hamiltonians of systems that exhibit topological phase transitions. The task of learning local patterns in topological systems is hard due to the fact that the topological state of a system is a global property emerging from complex interactions between its components. Therefore, even defining a local topological signal distributed along the components of the system is a challenging task. Strikingly, we demonstrate that topological information can be retrieved from local data in the bulk of such topological systems.

Our algorithm combines decision trees and random forests with eigenvector ensembling to analyze the Shannon entropy of ensembles of eigenstates of SSH systems. This allowed us to apply machine learning explainability tools to statistically recover from the analysis of several thousand SSH systems an information entropy signature describing how topological information is distributed in a lattice, a feat that would be infeasible with the usual approach in physics based on theoretical modeling. Our work therefore represents a true advancement in the understanding of quantum materials that could only have been made possible by machine learning.

In recent years it has become clear to the condensed matter physics and quantum materials communities that future discovery and development of new quantum materials will rely heavily on machine learning tools (see Zdeborová, *Nature Physics* volume 13, pages 420–421 (2017) for a high-level discussion of the applications of machine learning to physics). This is due to the increasing complexity of the systems studied that render traditional analytical approaches infeasible. As progress in cornerstone technologies such as quantum computers, spintronics and advanced electronics depends heavily on the development of new quantum materials with robust and exotic properties, the stakes could not be higher (as an example, the worldwide semiconductor industry alone is estimated to be worth around £200 billion). The arrival of machine learning to the physicist's toolbox is indeed great news to our community.

Physics is famously known among scientists to be hermetic to the non-specialist. In an era when many of the most relevant scientific problems require a multidisciplinary effort, this situation can have very detrimental consequences to the development of the field. This is increasingly the case in the fields of condensed matter and quantum materials, where machine learning experts shall soon become highly demanded. Being aware of the diverse readership of your journal, we have made the effort to frame the problem of learning topological phase transitions as a machine learning problem. The relevant physics and machine learning concepts are pedagogically explained in the Supplementary Material to bridge the gap between the two communities. We therefore expect that our paper will have a wide audience among your readers and contribute to introduce many of them to the exciting possibilities of applying machine learning to physics problems.

Referees with the required background to assess the manuscript include Roger Melko (University of Waterloo), Juan Carrasquilla (University of Waterloo), Eun-Ah Kim (Cornell University), Mathias Scheurer (Harvard University) and Sebastian Huber (ETH Zürich).

Yours faithfully,

N. L. Holanda and M. A. R. Griffith