Towards efficient and robust detection of entanglement structure

Jue Xu* and Qi Zhao†
(Dated: August 10, 2022)

Verification (detection) of entanglement structure is an indispensable step for practical quantum computation (communication). In this work, we compare complexity and performance of several recently-developed methods, including entanglement witness methods, shadow tomography, classical machine learning, and quantum algorithms (circuits). We illustrate the advantages and limitations of machine learning and quantum algorithms.

CONTENTS

I. INTRODUCTION

Entanglement [1] is the key ingredient of quantum computation [], quantum communication [], and quantum cryptography []. For practical purpose, it is essential to benckmark (characterize) multipartite entanglement structures of target states. We review the recently developed methods to entanglement detection: entanglement witness [2], shadow tomography [3], classical machine learning [4], and quantum (variational/circuit) algorithms [5].

II. PRELIMINARIES

Notations: The hats on the matrices such as $\hat{A}$, $\hat{H}$, ρ (omitted), $\hat{O}$, $\hat{W}$, emphasize that they play the roles of operators (Hermitian matrices). Denote vector (matrix) $\mathbf{x}$, $\mathbf{K}$ by boldface font. A simle (undirected, unweighted) graph $G = (V, E)$ is described by vertices V and edges E.

* juexu@cs.umd.edu
† email

For specific purpose, we use different basis (representations) for quantum states. One is the computational basis $\{|z\rangle\}$ with $z \in [2^n]$ where n is the number of qubits, while another useful one is the binary representation of computational basis $\{|\mathbf{x}\rangle \equiv |x_1\rangle |x_2\rangle \dots |x_n\rangle\}$ with $x_j \in \{0,1\}$. For simplicity, we let $N \equiv 2^n$ and $|\mathbf{0}\rangle \equiv |0^n\rangle \equiv |0\rangle^{\otimes n}$ if no ambiguity. shorthand $|\psi_A\rangle |\psi_B\rangle \equiv |\psi_A\rangle \otimes |\psi_B\rangle$. Hadamard basis $|+\rangle := (|0\rangle + |1\rangle)/\sqrt{2}$.

A. Entanglement detection

1. *Bipartite entanglement*

Large scale entanglement is the (main) resource of quantum advantages in quantum computation and communication. Firstly, we consider the simplest entanglement structure: bipartite separable case.

Definition 1 (density matrix)**.** pure state $|\psi\rangle$; density matrix ρ (trace one, Hermitian, PSD) ...

Definition 2 (partial trace)**.** partial trace; reduced density matrix $\rho_A = \mathrm{Tr}_B(\rho_{AB})$

Many methods [...] have been developed to determine whether a state is separable.

Definition 3 (bipartite separable)**.** A pure state is (bi-)separable if it is in a tensor product form $|\psi_b\rangle = |\phi_A\rangle \otimes |\phi_{\bar{A}}\rangle$, where $\mathcal{P}_2 = \{A, \bar{A}\}$ is a bipartition of the qubits in the system. A mixed state is separable if it can be written as a mixing of pure separable states. Note that each separable state $|\psi_b\rangle$ in the summation can have different bipartitions. The separable state set is denoted as S_b. There is another restricted way for the extension to mixed states. A state is $\mathcal{P}_2$-separable, if it is a mixing of pure separable states with a same partition $\mathcal{P}_2$, and we denote the state set as $S_b^{\mathcal{P}_2}$. entangled state?...

Rather than qualitatively determining (bi)separability, there are measures to quantify entanglement

Definition 4 (Schmidt measure)**.** Consider the following bipartite pure state, written in Schmidt form:

$$|\psi\rangle = \sum_i^r \sqrt{p_i}\, |\phi_i^A\rangle \otimes |\phi_i^B\rangle \tag{1}$$

where $\{|\phi_i^A\rangle\}$ is a basis for $\mathcal{H}_A$ and $\{|\phi_i^A\rangle\}$ for $\mathcal{H}_B$. The strictly positive values $\sqrt{p_i}$ in the Schmidt decomposition are its *Schmidt coefficients*. The number of Schmidt coefficients, counted with multiplicity, is called its *Schmidt rank*, or Schmidt number. (Schmidt rank ?? $\mathrm{SR}^A(\psi) = \mathrm{rank}(\rho_\psi^A)$) Schmidt measure is minimum of $\log_2 r$ where r is number of terms in an expansion of the state in product basis.

Definition 5 (entropy)**.** In quantum mechanics (information), the von Neumann *entropy* of a density matrix is $H_N(\rho) := -\mathrm{Tr}(\rho \log \rho) = -\sum_i \lambda_i \log(\lambda_i)$; In classical information (statistical) theory, the Shannon entropy of a probability distribution P is $H_S(P) := -\sum_i P(x_i) \log P(x_i)$. relative entropy (divergence)

Remark 1. (def partial trace ...) a pure (bipartite) state is entangled iff the reduced state $\rho^A = \mathrm{Tr}_A(\rho)$ is mixed. The mixedness of this reduced state allows one to quantify the amount of entanglement in this state.

Definition 6 (entanglement entropy)**.** The bipartite *von Neumann entanglement entropy* S is defined as the von Neumann entropy of either of its reduced density matrix ρ_A. For a pure state $\rho_{AB} = |\Psi\rangle\langle\Psi|_{AB}$, it is given by

$$E(\Psi_{AB}) = S(\rho_A) = -\mathrm{Tr}(\rho_A \log \rho_A) = -\mathrm{Tr}(\rho_B \log \rho_B) = S(\rho_B) \tag{2}$$

where $\rho_A = \mathrm{Tr}_B(\rho_{AB})$ and $\rho_B = \mathrm{Tr}_A(\rho_{AB})$ are the reduced density matrices for each partition.

Definition 7 (maximally entangled)**.** a state vector is maximually entangled iff the reduced state at one qubit is maximally mixed, i.e., $\mathrm{Tr}_a(|\psi\rangle\langle\psi|) = \frac{1}{2}$.

2. *Multipartite entanglement structures*

For multipartite quantum systems, it is crucial to identify not only the presence of entanglement but also its detailed structure. An identification of the entanglement structure may thus provide us with a hint about where imperfections in the setup may occur, as well as where we can identify groups of subsystems that can still exhibit strong quantum-informationprocessing capabilities.

Given a n-qubit quantum system and its partition into m subsystems, the *entanglement structure* indicates how the subsystems are entangled with each other. In some specific systems, such as distributed quantum computing[] quantum networks[] or atoms in a lattice, the geometric configuration can naturally determine the system partition. Therefore, it is practically interesting to study entanglement structure under partitions.

Definition 8 (fully entangled)**.** An n-qubit quantum state ρ is a *fully entangled*, if it is outside of the separable state set $S_b^{\mathcal{P}_2}$ for any partition, $\rho \notin S_b^{\mathcal{P}_2}, \forall \mathcal{P}_2 = \{A, \bar{A}\}$.

GME is the strongest form of entanglement, that is, all qubits in the system are indeed entangled with each other.

Definition 9 (genuine multipartite entanglement)**.** A state possesses *genuine multipartite entanglement* (GME) if it is outside of S_2, and is (fully) n-separable if it is in S_n. A state possesses $\mathcal{P}$-genuine entanglement if it is outside of $S_b^{\mathcal{P}}$. A state ρ possesses $\mathcal{P}$-genuine entanglement iff $\rho \notin S_b^{\mathcal{P}}$.

Compared with genuine entanglement, multipartite entanglement structure still lacks a systematic exploration, due to the rich and complex structures of n-partite system. Unfortunately, it remains an open problem of efficient entanglement-structure detection of general multipartite quantum states.

Definition 10 (Multipartite state)**.** denote the partition $\mathcal{P}_m = \{A_i\}$ and omit the index m when it is clear from the context.

define fully- and biseparable states with respect to a *specific partition* $\mathcal{P}_m$

Definition 11 (fully separable)**.** An n-qubit pure state $|\psi_f\rangle$ is *fully separable* iff. An n-qubit pure state $|\psi_f\rangle$ is $\mathcal{P}$-fully separable iff it can be written as $|\psi_f\rangle = \otimes_i^m |\phi_{A_i}\rangle$. An n-qubit mixed state ρ_f is $\mathcal{P}$-fully separable iff it can be decomposed into a convex mixture of $\mathcal{P}$-fully separable pure states.

$$\rho_f = \sum_i p_i \left|\psi_f^i\middle\rangle\!\middle\langle\psi_f^i\right|, (\forall i)(p_i \geq 0, \sum_i p_i = 1). \tag{3}$$

P-bi-separable... $S_f^{\mathcal{P}} \subset S_b^{\mathcal{P}}$

By going through all possible partitions, one can investigate higher level entanglement structures, such as entanglement intactness (non-separability), which quantifies how many pieces in the n-partite state are separated.

Remark 2. P-... can be viewed as generalized versions of regular fully separable, biseparable, and genuinely entangled states, respectively. In fact, when $m = n$, these pairs of definitions are the same. By definitions, one can see that if a state is P_m-fully separable, it must be m-separable. Of course, an m-separable state might not be P_m-fully separable, for example, if the partition is not properly chosen.

entanglement structure measures. To benchmark our technological progress towards the generation of largescale genuine multipartite entanglement, it is thus essential to determine the corresponding entanglement depth.

Definition 12 (Entanglement intactness, depth)**.** the entanglement intactness of a state ρ to be m, iff$\rho \notin S_{m+1}$ and $\rho \in S_m$. When the entanglement intactness is 1, the state possesses genuine multipartite entanglement; and when the intactness is n, the state is fully separable. k-producible.

Example 1. The Schmidt measure for any multi-partite GHZ states is 1, because there are just two terms. Schmidt measure for 1D, 2D, 3D-cluster state is $\lfloor \frac{N}{2} \rfloor$. Schmidt measure of tree is the size of its minimal vertex cover[??].

Example 2 (GHZ)**.** bipartite: Bell states; nontrivial multipartite: tripartite. GHZ state: $|\mathrm{GHZ}\rangle := \frac{1}{\sqrt{2}}(|0\rangle^{\otimes n} + |1\rangle^{\otimes n})$ (eight-photon) produce the five different entangled states (one from each entanglement structure/partition?):

$$|\mathrm{GHZ}_8\rangle, |\mathrm{GHZ}_{62}\rangle, |\mathrm{GHZ}_{44}\rangle, |\mathrm{GHZ}_{422}\rangle, |\mathrm{GHZ}_{2222}\rangle.$$

Schmidt rank, PPT criteria, entanglement witness

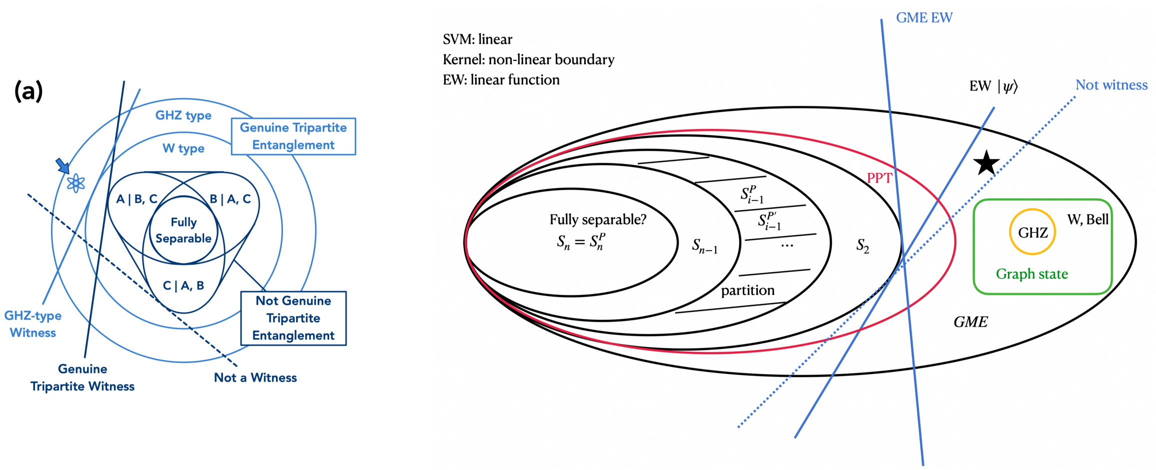


FIG. 1: (a) entanglement witness, PPT criteria, SVM (kernel)?. (c) convex hull...

3. Entanglement witness

Classically,

Theorem 1 ([6])**.** *The weak membership problem for the convex set of separable normalized bipartite density matrices is* NP*-Hard.* ***Input****: unknown state?? formal definition of the problem*

However, we do not know approximately correct complexity? quantum complexity? machine learning (data)? for entanglement () problem

Theorem 2 (PPT criterion)**.** *the positive partial transpose (*PPT*) criterion, saying that a separable state (bipartite separable) must have PPT?. Note, it is only necessary and sufficient when $d_A d_B \leq 6$.*

see Fig. 1 for relations. entanglement detection [7].

Definition 13 (entanglement witness)**.** Given an (unknown? known target state) quantum state (density matrix) ρ, the *entanglement witness* $\hat{W}$ is an obseverable such that

$$\mathrm{Tr}\left(\hat{W}\rho\right) \geq 0, \forall \text{ separable }; \quad \mathrm{Tr}\left(\hat{W}\rho\right) < 0, \text{ for some entangled} \tag{4}$$

In a typical experiment one aims to prepare a pure state, $|\psi\rangle$, and would like to detect it as true multipartite entangled. While the preparation is never perfect, it is still expected that the prepared mixed state is in the proximity of $|\psi\rangle$. The usual way to construct entanglement witnesses using the knowledge of this state is

$$\hat{W}_\psi = c\mathbb{1} - |\psi\rangle\langle\psi| \tag{5}$$

where c is the smallest constant such that for every product state $\mathrm{Tr}\left(\rho\hat{W}\right) \geq 0$

Remark 3. In order to measure the witness in an experiment, it must be decomposed into a sum of locally measurable operators. The number of local measurements in these decompositions seems to increase exponentially with the number of qubits.

Example 3 (entanglement witness for GHZ)**.** three-qubit GHZ state [8]

$$\hat{W}_{\mathrm{GHZ}_3} := \frac{3}{2}\mathbb{1} - \hat{\sigma}_x^{(1)}\hat{\sigma}_x^{(2)}\hat{\sigma}_x^{(3)} - \frac{1}{2}\left[\hat{\sigma}_z^{(1)}\hat{\sigma}_z^{(2)} + \hat{\sigma}_z^{(2)}\hat{\sigma}_z^{(3)} + \hat{\sigma}_z^{(1)}\hat{\sigma}_z^{(3)}\right] \tag{6}$$

This witness requires the measurement of the $\left\{\hat{\sigma}_x^{(1)}, \hat{\sigma}_x^{(2)}, \hat{\sigma}_x^{(3)}\right\}$ and $\left\{\hat{\sigma}_z^{(1)}, \hat{\sigma}_z^{(2)}, \hat{\sigma}_z^{(3)}\right\}$ settings. The projector based witness $\hat{W}_{\mathrm{GHZ}_3} = \mathbb{1}/2 - |\mathrm{GHZ}\rangle\langle\mathrm{GHZ}|$ requires four measurement settings. detect genuine n-qubit entanglement close to GHZ_n

$$\hat{W}_{\mathrm{GHZ}_n} = (n-1)\mathbb{1} - \sum_{k=1}^{n} S_k^{(\mathrm{GHZ}_n)} \tag{7}$$

S_k is the stabilizer ... Detecting Genuine Multipartite Entanglement with Two Local Measurements [8]

It is natural to ask nonlinear entanglement witness [9] and the kernel method in machine learning.

Proposition 1. *Given a state $|\psi\rangle$, the entanglement witness operator $\hat{W}_\psi$ can witness genuine multipartite entanglement near $|\psi\rangle$ with $c = 5/8$ in Eq. (5) that is, $\left\langle \hat{W}_\psi \right\rangle \geq 0$ for any separable state in S_b.*

If the fidelity (quantum kernel?) of the prepared state ρ_{pre} with the target state $|\psi\rangle$, i.e., $\text{Tr}(\rho_{pre} |\psi\rangle\langle\psi|)$, exceeds 5/8, ρ_{pre} possesses GME. It is generally difficult to evaluate the quantity $\text{Tr}(\rho_{pre} |\psi\rangle\langle\psi|)$ by the direct projection on $|\psi\rangle$, as it is an entangled state.

A usual approach for detecting entanglement is using Bell inequalities [??]

Definition 14 (Bell inequality)**.** CHSH inequality (game); Bell inequalities for graph states $\left|\sum_{\sigma\in S} \langle\sigma\rangle\right| \leq C$?

Another approach for detecting multipartite entanglement is using entanglement witnesses. Different Bell inequalities can be regarded as entanglement witness for different types of entanglement in a multi-party entangled state. These witnesses can be quite useful to detect entanglement in the vicinity of graph states.

Problem 1 (Entanglement witness with prior)**.** with/out prior knowledge

- **Input**: a **known** state $|\psi\rangle$, with noise
- **Output**: separable or not ??? $S_f^{\mathcal{P}}$? $S_b^{\mathcal{P}}$

Problem 2 (Certify entanglement)**.** Multipartite entanglement-structure detection

- **Input**: an (actual) state ρ' from experiment that is close to a **known/target** (general multipartite) state $|\psi\rangle$, certain partition?
- **Output**: the certified lower-order entanglement among several subsystems could be still useful for some quantum information tasks. entanglement structure

4. Graph state

graph state is an important (large?) class of multipartite states in quantum information. Typical graph states include cluster (lattice) states, GHZ states, and the states involved in error correction (toric code?). It worth noting that 2D cluster state is the universal resource for the measurement based quantum computation (MBQC) [10].

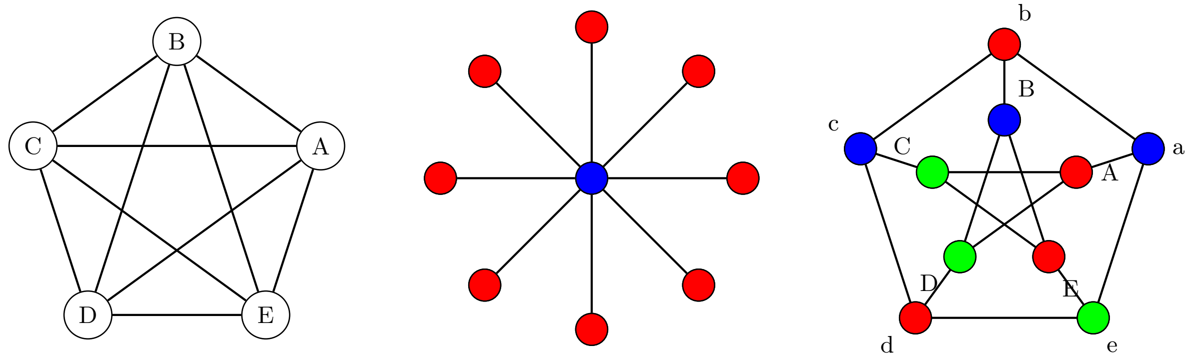
Definition 15 (graph state)**.** Given a simple graph (undirected, unweighted, no loop and multiple edge) $G = (V, E)$, a graph state is constructed as from the initial state $|+\rangle^{\otimes n}$ corresponding to n vertices. Then, apply controlled-Z gate to every edge, that is

$$|G\rangle := \prod_{(i,j)\in E} \mathsf{cZ}_{(i,j)} |+\rangle^{\otimes n} \tag{8}$$

Remark 4. An n-partite(qubit) graph state can also be uniquely determined by n independent stabilizers, $S_i := X_i \bigotimes_{j\in n} Z_j$, which commute with each other and $\forall i, S_i |G\rangle = |G\rangle$.?? The graph state is the unique eigenstate wtih eignevalue of $+1$ for all the n stabilizers. As a result, a graph state can be written as a product of stailizer projectors, $|G\rangle\langle G| = \prod_{i=1}^n \frac{S_i + \mathbb{1}}{2}$. stabilizer formalism?;

Example 4 (graph states)**.** Any connected graph state is fully entangled state. The GHZ state corresponds to the star graph and the complete gtaph (Fig. 2). This is easily seen by applying Hadamard unitaries $\hat{U}_{\text{H}}^{V\setminus a}$ to all but one qubit a in the GHZ-state, which yields the star graph state with a as the central qubit. (line, ring; hypercube, Petersen graph; cluster state in two dimensinos, which corresponds to a rectangular lattice.) The Petersen graph is not LC-equivalent to its isomorphism (exchanging the labels at each ed of the five "spokes"). However, the lists of Schmidt ranks (or, equivalently, the connectivity functions) of these graphs coincide. The class of CSS (error correction) states corresponds to the class of 2-colorable graphs. [11]

Remark 5 (??)**.** The entanglement entropy $S(\rho_A)$ equals the rank of the adjacency matrix of the underlying bipartite graph, which can be efficiently calculated. For graph states, the reduced density matrices can be represented efficiently in terms of their stabilizer elements or their adjacency matrix.



(a) complete graph K_5 (b) star graph (2-colorable) - GHZ (c) Petersen graph

FIG. 2: graph states

Proposition 2 ([2]). *Given a graph state $|G\rangle$ and a partition $\mathcal{P} = \{A_i\}$, the fidelity between $|G\rangle$ and any fully separable is upper bounded by*

$$\mathrm{Tr}(|G\rangle\langle G|\,\rho_f) \leq \min_{\{A,\bar{A}\}} 2^{-S(\rho_A)} \tag{9}$$

where $S(\rho_A)$ is the von Neumann entropy of the reduced density matrix $\rho_A = \mathrm{Tr}_{\bar{A}}(|G\rangle\langle G|)$.

Remark 6. LU, LC equivalence, local operations and classical communication (LOCC)

the entanglement in a graph state is related to the topology of its underlying graph.

Proposition 3 (Entanglement of graph state). *[11]. witness; bounds; graph property? vertex cover? Hamiltonian cycle of a graph state?*

generalize [12] stabilizer state, neural network state [13]?

Proposition 4 (Entanglement witness for graph state). *Bell inequality*

$$\hat{W} = \frac{C}{2^N}\mathbb{1}_V - |G\rangle\langle G| \tag{10}$$

Let $|G\rangle$ be a graph state corresponding to a connected graph. Then

$$\hat{W}_1^{ab} = \mathbb{1}_V - K_a - K_b \tag{11}$$

is an entanglement witness for the $|G\rangle$ that detects entanglement in the reduced state $\rho_G^A(A = N_a \cup N_b \cup \{a,b\})$ with only two measurement settings and thus can rule out full separability of the total graph state???. The entanglement witness

$$\hat{W}_2 = (N-1)\mathbb{1}_V - \sum_{a\in V} K_a \tag{12}$$

detects genuine multipartite entanglement.

Question 1. *for which case, C is hard to compute? non-stabilizer state? SWAP?*

Theorem 3. *k local measurements. Here, k is the chromatic number (minimal colorable) of the corresponding graph, typically, a small constant independent of the number of qubits.*

Proposition 5 (Bounds to the Schmidt measure of graph states). *For any graph state $|G\rangle$, the Schmidt measure E_A is bounded from below by the maximal Schmidt rank $\mathrm{SR}_{\max}$ and from above by the Pauli persistency PP or the minimal vertex cover, i.e.*

$$\mathrm{SR}_{\max}(G) \leq E_S(|G\rangle) \leq \mathrm{PP}(G) \leq \mathrm{VC}(G). \tag{13}$$

???

B. Tomography and trace estimation

Intuitively, a general tomography [14] that extract (recover) all information about a state requires exponential copies (samples/measurements).

Problem 3 (full tomography)**.** In contrast to shadow tomography, we refer to *full tomography* here

- **Input**: Given a **unknown** N-dimensional mixed state ρ
- **Output**: a complete description? of ρ (decomposition coefficients) with error? Stokes parameter $S_i \equiv \mathrm{Tr}(\hat{\sigma}_i \rho)$

$$\rho = \frac{1}{2^n} \sum_{i_1, i_2, \ldots, i_n = 0}^{3} S_{i_1, i_2, \ldots, i_n} \hat{\sigma}_{i_1} \otimes \hat{\sigma}_{i_2} \otimes \cdots \otimes \hat{\sigma}_{i_n} \tag{14}$$

Theorem 4 (lower bound of full tomography?[15])**.** *Known fundamental lower bounds [66, 73] state that classical shadows of exponential size (at least) $T = \Omega(2^n/\epsilon^2)$ are required to ϵ-approximate ρ in trace distance.*

Definition 16 (fidelity)**.** Given a pair of states (target ρ and prepared ρ'), Uhlmann fidelity $F(\rho, \rho') := \mathrm{Tr}\left(\sqrt{\sqrt{\rho}\rho'\sqrt{\rho}}\right)$

Definition 17 (distance)**.** trace distance $d_{\mathrm{tr}}(\rho, \rho') := \frac{1}{2}\|\rho - \rho'\|_1$

Definition 18 (norm)**.** Schatten p-norm $\|x\|_p := (\sum_i |x_i|^p)^{1/p}$. Euclidean norm l_2 norm; Spectral (operator) norm $\|\mathbf{x}\|_\infty$; Trace norm $\|A\|_{\mathrm{Tr}} \equiv \|A\|_1 := \mathrm{Tr}\left(\sqrt{A^\dagger A}\right)$, $p = 1$; Frobenius norm $\|A\|_F := \sqrt{\mathrm{Tr}(A^\dagger A)}$, $p = 2$; Hilbert-Schmidt norm $\|A\|_{HS} := \sqrt{\sum_{i,j} A_{ij}^2}$

In quantum mechanics, interesting properties are often linear functions of the underlying density matrix ρ. For example, the fidelity with a pure target state, entanglement witnesses fit this framework.

Problem 4 (trace estimation)**.** related problems defined as follows

- **Input**: Given an observable (Hermitian) $\hat{O}$ and (copies of) a mixed state ρ $(\rho', \ldots, \rho_m)$,
- **Output**: with ϵ in trace distance
  - linear function: the expectation value $\left\langle \hat{O} \right\rangle = \mathrm{Tr}\left(\hat{O}\rho\right)$; entanglement witness $\mathrm{Tr}\left(\hat{W}\rho\right)$; shadow tomography $\mathrm{Tr}(E_M \rho) = \mathbb{E}[E_M] = \mathbb{P}[E_i \text{ accept } \rho]$??; full tomography; two-point correlation $\langle O_i O_j \rangle$
  - nonlinear function: entropy (non-linear); fidelity $F(\rho, \rho')$, distance??;
  - multivariate: quantum kernel $\mathrm{Tr}(\rho\rho')$; multivariate $\mathrm{Tr}(\rho_1 \cdots \rho_m)$, nonlinear function?? linear;

A nonlinear function of ρ such as entanglement entropy, may also be of interest.

Nevertheless, we usually only need specific properties of a target state rather than full classical descriptions about the state. This enables the possibility to shadow tomography.

Problem 5 (shadow tomography)**.** *shadow tomography*

- **Input**: an **unknown** N-dimensional mixed state ρ, M known 2-outcome measurements $E_1, \ldots, E_M$
- **Output**: estimate $\mathbb{P}[E_i \text{ accept } \rho]$ to within additive error ϵ, $\forall i \in [M]$, with $\geq 2/3$ success probability.

Theorem 5 ([16])**.** *It is possible to do shadow tomography using $\tilde{\mathcal{O}}(\frac{\log^4 M \cdot \log N}{\epsilon^4})$ copies. [no construction algorithm?] sample complexity lower bound $\Omega(\log(M) \cdot \epsilon^{-2})$,*

more details in Section III A 1

Remark 7 ([3])**.** While very efficient in terms of samples, Aaronson's procedure is very demanding in terms of quantum hardware — a concrete implementation of the proposed protocol requires exponentially long quantum circuits that act collectively on all the copies of the unknown state stored in a quantum memory.??

III. CLASSICAL, DATA-POWERED, AND QUANTUM ALGORITHMS

We consider the problem, entanglement structure dectection for graph states

Problem 6 (?data)**.** problem without training data

- **Input**: a graph G encoding in a graph state $|G\rangle$; adjacency matrix A?
- **Output**: entanglement structure

with training data: **features**: classical shadow? raw data? quantum data, label: entangled? input encoding (model):

- amplitude encoding: given a normalized vector $\mathbf{x} \in \mathbb{R}^d$, the quantum state $|\mathbf{x}\rangle = \sum_z^d x_z |z\rangle$. need $\log(d)$ qubits for a data point; dequantization [17]. qubit (basic) encoding (inefficient? space)
- unitary encoding: quantum simulation (Hamiltonian); quantum random walk (adjacency matrix); oracle (controlled) unitary, quantum phase estimation; graph state, discrete, efficient? space (time), isomorphism?
- quantum data $|\psi\rangle$ or ρ: quantum state from real-world experiments or quantum circuits $\hat{U}$. no input problem? more efficient?

Definition 19 (graph property)**.** monotone

Example 5 (colorable)**.** k-colorable is a graph property, i.e., allow for a coloring of the vertices with k colors such that no two adjacent vertices have the same color. bipartite graph iff 2-colorable

Problem 7 (graph property test)**.**

quantum algorithms (bounds) for graph properties [18]

A. Classical shadow and machine learning

separability classifier by neural network [19]. rigorous quantum advantage of quantum kernel method in SVM [20]. classical machine learning with classical shadow [21].

1. *Classical shadow*

Inspired by Aaronson's shadow tomography [16], Huang et. al [3] introduce classical shadow. The classical shadow attempts to approximate this expectation value by an empirical average over T independent samples, much like Monte Carlo sampling approximates an integral.

Definition 20 (classical shadow)**.** classical shadow

$$\rho_{cs} = \mathcal{M}^{-1}\left(U^\dagger \left|\hat{b}\middle\rangle\middle\langle\hat{b}\right| U\right) \tag{15}$$

such that we can predict the linear function with classical shadows

$$o_i = \mathrm{Tr}(O_i \rho_{cs}) \text{ obeys } \mathbb{E}[o] = \mathrm{Tr}(O_i \rho) \tag{16}$$

$\sigma_T(\rho(x_l))$ is the classical shadow representation of $\rho(x_l)$, a $2^n \times 2^n$ matrix that reproduces $\rho(x_l)$ in expectation over random Pauli measurements.

$$\{x_l \to \sigma_T(\rho(x_l))\}_{l=1}^N \tag{17}$$

The classical shadow size required to accurately approximate all reduced r-body density matrices scales exponentially in subsystem size r, but is independent of the total number of qubits n.

Algorithm III.1: Classical Shadow (tomography)

input : a density matrix ρ (many copies), observables $\hat{O}$...
output: classical shadow ρ_{cs}

```
for i = 1, 2, ..., m do
    random Pauli measurements                                  // a comment
    ... return?
return?
```

Lemma 1. *the variance*

$$\mathrm{Var}[o] = \mathbb{E}[(o - \mathbb{E}[o])^2] \leq \left\| O - \frac{\mathrm{Tr}(O)}{2^n} \mathbb{1} \right\|^2_{shadow} \tag{18}$$

sample complexity

$$N_{tot} = \mathcal{O}\left(\frac{\log(M)}{\epsilon^2} \max_{1 \leq i \leq M} \left\| O_i - \frac{\mathrm{Tr}(O_i)}{2^n} \mathbb{1} \right\|^2_{\text{shadow}} \right) \tag{19}$$

Theorem 6 (Pauli/Clifford measurements)**.** *additive error ϵ, M arbitrary k-local linear function $\mathrm{Tr}\left(\hat{O}_i \rho\right)$, lower bound $\Omega(\log(M) 3^k / \epsilon^2)$ copies of the state ρ.*

2. training data and classical kernel methods

Graphs is another kind of data which is fundamentally different from a real value vector because of vertex-edge relation, graph isomorphism. So, graph kernel [22] need additional attention.

Definition 21 (graph kernel)**.** given a pair of graphs (G, G'), *graph kernel* is $k(G, G') =$. quantum graph kernel $k(G, G') = |\langle G | G' \rangle|^2$?? [23]

Definition 22 (shadow kernel)**.** given two density matrices (quantum states) ρ and ρ', *shadow kernel* [3] is

$$k_{\text{shadow}}(S, S') := \exp\left(\sum \exp\left(\sum \right) \right) \tag{20}$$

similarity measures? advantages? why? (isomorphism?)

Definition 23 (divergence)**.** KL divergence (relative entropy): measure the distance (similarity) between two probability distributions:

$$D_{\text{KL}}(P || Q) := \sum P(x) \log(P(x)/Q(x)) \tag{21}$$

symmetric version: Jensen-Shannon divergence (machine learning)

$$D_{\text{JS}}(P || P') := \frac{1}{2}(D_{\text{KL}}(P || M) + D_{\text{KL}}(P' || M)) \equiv H_S(M) - \frac{1}{2}(H_S(P) - H_S(P')) \tag{22}$$

where $M = (P + P')/2$ and Shannon entropy H_S. Analogously, quantum Jensen-Shannon divergence D_{QJS} can be defined...

power of data -

Proposition 6 ([4])**.** *If a classical algorithm without training data can compute $f(x) = \langle x | \hat{U}^\dagger_{\text{QNN}} \hat{O} U_{\text{QNN}} | x \rangle$ (amplitude encoding) efficiently (poly time in ...) for any $\hat{U}_{\text{QNN}}$ and $\hat{O}$, then* BPP $=$ BQP *(which is believed unlikely).*

Proposition 7 ([4])**.** *Training an arbitrarily deep quantum neural network $\hat{U}_{\text{QNN}}$ with a trainable observable $\hat{O}$ is equivalent to training a quantum kernel method with kernel $k_Q(\mathbf{x}, \mathbf{x}') = \mathrm{Tr}(\rho(\mathbf{x}) \rho'(\mathbf{x}'))$*

Proposition 8 ([4])**.** *exist quantum advantage in machine learning (not significant, practical) ... discrete log, factoring...*

Algorithm III.2: Classical learning (SVM) + classical shadow

input : classical shadow? (features) with label (training data)
output: entanglement structure? decision

for $i = 1, 2, \ldots, m$ **do**
  kernel estimation // classical kernel
  SVM // SVM
  return "?"
return w // parameters of the separating hyperplane in the feature space

B. Quantum trace (kernel) estimation

1. *Entanglement witness*

Algorithm III.3: entanglement witness by ... quantum trace estimation

input : (copies of) density matrix (graph state?) ρ, an entanglement witness (observable) $\hat{W}$
output: determine entangled structure??

for $i = 1, 2, \ldots, m$ **do**
  $\hat{W}_i$ // estimate entanglement witness by quantum circuit
  return $\mathrm{Tr}\left(\hat{W}\rho\right)$
return entangled ? GME ? separable with certain partition?

2. *Quantum kernel SVM*

[24] [20] [25] [26]

Definition 24 (quantum kernel)**.** quantum kernel with quantum feature map $\phi(\mathbf{x}) : \mathcal{X} \to |\phi(\mathbf{x})\rangle\langle\phi(\mathbf{x})|$

$$k_Q(\rho, \rho') := |\langle\phi(\mathbf{x})|\phi(\mathbf{x}')\rangle|^2 = \left| \langle 0|\hat{U}^\dagger_{\phi(\mathbf{x})}\hat{U}_{\phi(\mathbf{x}')}|0\rangle \right|^2 = \mathrm{Tr}(\rho\rho') \tag{23}$$

where $\hat{U}_{\phi(\mathbf{x})}$ is a quantum circuit or physics process that encoding an input $\mathbf{x}$. In quantum physics, quantum kernel is also known as transition amplitude (quantum propagator);

Theorem 7 ([5])**.** *multivariate trace estimation can be implemented in constant quantum depth, with only linearly-many controlled two-qubit gates and a linear amount of classical pre-processing*

C. Variational (hybrid) quantum algorithms

1. *Variational entanglement witness (ansatz)*

an ansatz for entanglement witness (graph state entanglement)

$$\hat{W}_a := \sum_{\{i\ldots\}} w_{..} \bigotimes_i^n \hat{\sigma}^{(i)}, \quad \hat{\sigma} \in \{\hat{\sigma}_x, \hat{\sigma}_y, \hat{\sigma}_z, \mathbb{1}_{2\times 2}\} \tag{24}$$

c.f. full tomography (Stokes parameters) Eq. (14)

2. *Variational trace estimate (direct)*

find optimal entanglement witness (qunatum circuit?)

D. Theoretic upper bounds and lower bounds

[4] [3] [16] [27] [20]

	gate/depth/computation	measurements/samples	query?	necessary?sufficient
shadow tomography		Theorems 5 and 6	N/A	
indirect? direct (no prior), promise				
entanglement witness		Proposition 1 (constant?)	convex?	
classical ML + classical shadow	??			
quantum (variational) circuits	Theorem 7 (c-depth?)			

TABLE I: complexity measures of different methods

1. *Separations (complexity)*

contrived problem (engineered dataset)? for exponential speedup

2. *Obstacles (practical)*

IV. NUMERICAL SIMULATION

A. Classification accuracy

1. *Data preparation*

multi-partite entangled state: generate synthetic (engineered) data from (random graph?). separable state from randomly ...

2. *Hyperparameters and settings*

We consider a set of different regularization parameters,...

3. *Results*

performance of different methods:

FIG. 3: comparison of

B. Robustness to noise

tradeoff between (white noise) tolerance (robustness) and efficiency (number of measurements).

$$\rho'_{\text{noise}} = (1 - p_{\text{noise}}) |G\rangle\langle G| + p_{\text{noise}} \frac{\mathbb{1}}{2^n} \tag{25}$$

p_{noise} indicates the robustness of the algorithm (witness).

Remark 8 ([2]). the largest noise tolerance p_{limit} just related to the **chromatic number** (graph property) of the graph.

FIG. 4: robustness

V. CONCLUSION AND DISCUSSION

future: experimental implementation with a few qubits (generation, verification) [28]

Acknowledgements

---

[1] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Rev. Mod. Phys. **81**, 865 (2009), arXiv:quant-ph/0702225.
[2] Y. Zhou, Q. Zhao, X. Yuan, and X. Ma, npj Quantum Inf **5**, 83 (2019).
[3] H.-Y. Huang, R. Kueng, and J. Preskill, Nat. Phys. **16**, 1050 (2020), arXiv:2002.08953 [quant-ph].
[4] H.-Y. Huang, M. Broughton, M. Mohseni, R. Babbush, S. Boixo, H. Neven, and J. R. McClean, Nat Commun **12**, 2631 (2021), arXiv:2011.01938 [quant-ph].
[5] Y. Quek, M. M. Wilde, and E. Kaur, Multivariate trace estimation in constant quantum depth (2022), arXiv:2206.15405 [hep-th, physics:quant-ph].
[6] L. Gurvits, Classical deterministic complexity of Edmonds' problem and Quantum Entanglement (2003), arXiv:quant-ph/0303055.
[7] O. Gühne and G. Toth, Physics Reports **474**, 1 (2009), arXiv:0811.2803 [cond-mat, physics:physics, physics:quant-ph].
[8] G. Toth and O. Guehne, Phys. Rev. Lett. **94**, 060501 (2005), arXiv:quant-ph/0405165.
[9] O. Gühne and N. Lütkenhaus, Phys. Rev. Lett. **96**, 170502 (2006).
[10] H. J. Briegel, D. E. Browne, W. Dür, R. Raussendorf, and M. V. den Nest, Nature Phys **5**, 19 (2009), arXiv:0910.1116.
[11] M. Hein, W. Dür, J. Eisert, R. Raussendorf, M. V. den Nest, and H.-J. Briegel, Entanglement in Graph States and its Applications (2006), arXiv:quant-ph/0602096.
[12] Y. Zhang, Y. Tang, Y. Zhou, and X. Ma, Phys. Rev. A **103**, 052426 (2021), arXiv:2012.07606 [quant-ph].
[13] X. Gao and L.-M. Duan, Nat Commun **8**, 662 (2017), arXiv:1701.05039 [cond-mat, physics:quant-ph].
[14] J. Altepeter, E. Jeffrey, and P. Kwiat, in *Advances In Atomic, Molecular, and Optical Physics*, Vol. 52 (Elsevier, 2005) pp. 105–159.
[15] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu, IEEE Trans. Inform. Theory , 1 (2017).
[16] S. Aaronson, in *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2018 (Association for Computing Machinery, New York, NY, USA, 2018) pp. 325–338, arXiv:1711.01053.
[17] E. Tang, Phys. Rev. Lett. **127**, 060503 (2021), arXiv:1811.00414 [quant-ph].
[18] S. Ben-David, A. M. Childs, A. Gilyén, W. Kretschmer, S. Podder, and D. Wang, 2020 IEEE 61st Annu. Symp. Found. Comput. Sci. FOCS , 649 (2020), arXiv:2006.12760.
[19] S. Lu, S. Huang, K. Li, J. Li, J. Chen, D. Lu, Z. Ji, Y. Shen, D. Zhou, and B. Zeng, Phys. Rev. A **98**, 012315 (2018), arXiv:1705.01523 [quant-ph].
[20] Y. Liu, S. Arunachalam, and K. Temme, Nat. Phys. **17**, 1013 (2021), arXiv:2010.02174 [quant-ph].
[21] H.-Y. Huang, R. Kueng, G. Torlai, V. V. Albert, and J. Preskill, Provably efficient machine learning for quantum many-body problems (2021), arXiv:2106.12627 [quant-ph].
[22] N. M. Kriege, F. D. Johansson, and C. Morris, Appl Netw Sci **5**, 6 (2020), arXiv:1903.11835 [cs, stat].
[23] L. Bai, L. Rossi, A. Torsello, and E. R. Hancock, Pattern Recognition **48**, 344 (2015).
[24] J. R. Glick, T. P. Gujarati, A. D. Corcoles, Y. Kim, A. Kandala, J. M. Gambetta, and K. Temme, Covariant quantum kernels for data with group structure (2021), arXiv:2105.03406 [quant-ph].
[25] V. Havlicek, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, Nature **567**, 209 (2019), arXiv:1804.11326.
[26] M. Schuld and N. Killoran, Phys. Rev. Lett. **122**, 040504 (2019), arXiv:1803.07128 [quant-ph].
[27] H.-Y. Huang, R. Kueng, and J. Preskill, Phys. Rev. Lett. **126**, 190505 (2021), arXiv:2101.02464 [quant-ph].
[28] H. Lu, Q. Zhao, Z.-D. Li, X.-F. Yin, X. Yuan, J.-C. Hung, L.-K. Chen, L. Li, N.-L. Liu, C.-Z. Peng, Y.-C. Liang, X. Ma, Y.-A. Chen, and J.-W. Pan, Phys. Rev. X **8**, 021072 (2018).
[29] A. Jacot, F. Gabriel, and C. Hongler, Neural Tangent Kernel: Convergence and Generalization in Neural Networks (2020), arXiv:1806.07572 [cs, math, stat].

Appendix A: Machine learning background

Notations: The (classical) training data (for supervised learning) is a set of m data points $\left\{(\mathbf{x}^{(i)}, y^{(i)})\right\}_{i=1}^{m}$ where each data point is a pair $(\mathbf{x}, y)$. Normally, the input (e.g., an image) $\mathbf{x} := (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ is a vector where d is the number of *features* and its *label* $y \in \Sigma$ is a scalar with some discrete set Σ of alphabet/categories. For simplicity and the purpose of this paper, we assume $\Sigma = \{-1, 1\}$ (binary classification).

1. Support vector machine

Definition 25 (SVM)**.** Given a set of (binary) labeled data, support vector machine (SVM) is designed to find a hyperplane (a linear function) such that maximize the margin between two partitions...

$$\max_{\mathbf{w}} \tag{A1}$$

a. kernel method

nonlinear boundary. map to a higher dimensional (feature) space, in which data is linearly separable. kernel

Definition 26 (kernel)**.** In general, the kernel function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ measures the similarity between two input data points by an inner product

$$k(\mathbf{x}, \mathbf{x}') := \langle \phi(\mathbf{x}), \phi(\mathbf{x}') \rangle \tag{A2}$$

If the input $\mathbf{x} \in \mathbb{R}^d$ (conventional machine learning task, e.g., image classification), the feature map $\phi(\mathbf{x}) : \mathbb{R}^d \to \mathbb{R}^n$ $(d < n)$ from a low dimensional space to a higher dimensional space. The corresponding kernel (Gram) matrix $\mathbf{K}$ should be a positive, semidefinite (PSD) matrix.

Example 6 (kernels)**.** Some common kernels: the polynomial kernel $k_{\text{poly}}(\mathbf{x}, \mathbf{x}') := (1 + \mathbf{x} \cdot \mathbf{x}')^q$ with feature map $\phi(\mathbf{x})$ ... The Gaussian kernel $k_{\text{gaus}}(\mathbf{x}, \mathbf{x}') := \exp\Big(-\gamma \|\mathbf{x} - \mathbf{x}'\|_2^2\Big)$ with an infinite dimensional feature map $\phi(\mathbf{x})$. An important feature of kernel method is that kernels can be computed efficiently without evaluating feature map (might be infinite dimension) explicitly.

2. Neural network

a. neural network and kernel

Definition 27 (neural tagent kernel)**.** neural tangent kernel [29]: proved to be equivalent to deep neural network [13] in the limit ...

$$k_{\text{NT}}\Big(S_T(\rho_l), \tilde{S}_T(\rho_{l'})\Big) = \Big\langle \phi^{(\text{NT})}(S_T(\rho_l)), \phi^{(\text{NT})}(\tilde{S}_T(\rho_l)) \Big\rangle \tag{A3}$$

b. Stabilizer formalism

denote a group by $\mathbb{G}$ and a subgroup $\mathbb{H}$.

Definition 28 (Pauli group)**.**

Definition 29 (Clifford group)**.**

Definition 30 (Stabilizer)**.**

c. quantum neural network

Appendix B: Hardness assumptions

Definition 31 (NP)**.** NP, NP-hard, NP-complete

Definition 32 (#P)**.** #P

Definition 33 (QMA)**.** QMA

Definition 34 (BPP)**.** BPP

Definition 35 (BQP)**.** BQP