# Quantum sketching protocols for Hamming distance and beyond

João Fernando Doriguello[1, 2, *] and Ashley Montanaro[1, †]

[1] *School of Mathematics, University of Bristol, Bristol BS8 1TW, United Kingdom*
[2] *Quantum Engineering Centre for Doctoral Training, University of Bristol, Bristol BS8 1TW, United Kingdom*

In this work we use the concept of quantum fingerprinting to develop a quantum communication protocol in the simultaneous message passing model that calculates the Hamming distance between two $n$-bit strings up to relative error $\epsilon$. The number of qubits communicated by the protocol is polynomial in $\log n$ and $1/\epsilon$, while any classical protocol must communicate $\Omega(\sqrt{n})$ bits. Motivated by the relationship between Hamming distance and vertex distance in hypercubes, we apply the protocol to approximately calculate distances between vertices in graphs that can be embedded into a hypercube such that all distances are preserved up to a constant factor. Such graphs are known as $\ell_1$-graphs. This class includes all trees, median graphs, Johnson graphs and Hamming graphs. Our protocol is efficient for $\ell_1$-graphs with low diameter, and we show that its dependence on the diameter is essentially optimal. Finally, we show that our protocol can be used to approximately compute $\ell_1$ distances between vectors efficiently.

## I. INTRODUCTION

Imagine that two separated parties (Alice and Bob) each have some data, and would like to determine how alike their data is, using the minimal amount of communication possible. Also imagine that they are not allowed to communicate with each other, but are each only allowed to send a single message to a third party ("referee"), and do not share any prior information with each other. This communication model is known as the simultaneous message passing (SMP) model with private randomness [1]. It encapsulates, for example, a scenario where it is not clear in advance whose data sets are to be compared. Another motivation comes from cryptographic scenarios. For example, it could be that the inputs to the two parties are controlled by an adversary, who has access to any previously shared randomness and can choose the inputs such that the protocol fails [2]; alternatively, Alice and Bob may simply want to find an efficient protocol which hides their data from the referee.

A natural strategy for completing this task is for each of Alice and Bob to compress their data to some kind of "sketch" [3, 4], and send the sketches to the referee, who uses them to determine the distance between the corresponding original data sets. Unfortunately, even for one of the simplest distance measures possible – testing equality of $n$-bit strings – and even if Alice and Bob are allowed a small probability of failure, this task requires $\Omega(\sqrt{n})$ bits of classical communication to the referee [5]. In comparison, if Alice and Bob are allowed access to a shared random bit-string, this complexity drops to $O(1)$ [6].

Remarkably, the use of quantum information allows an exponential reduction in the complexity of equality-testing. If Alice and Bob encode their $n$-bit strings as particular quantum states called *quantum fingerprints*, then there exists a quantum protocol that communicates only $O(\log n)$ qubits [7] and succeeds with arbitrarily high constant probability.

This surprising result sparked significant interest from the perspective of computer science [8, 9] and information theory [10], as well as physics. Theoretically, it has been used to shed new light on the two-slit experiment [11] and detailed studies of fingerprinting schemes using few qubits have been undertaken [12, 13]. Proof-of-principle quantum fingerprinting experiments have been carried out with states of 1 qubit realized using linear optics [14] and nuclear magnetic resonance [15]. More recently, a variant of the quantum fingerprinting protocol based on coherent states [16] has also been implemented experimentally, surpassing the best known classical protocols [17] and even the classical theoretical limit [18].

However, equality is just one distance measure, and a very special one. Here we seek other measures of distance for which quantum information can achieve a similar exponential advantage. In addition to the inherent theoretical interest of this question in terms of giving insight into the expressive power of quantum states, quantum protocol for more general distance measures could find significantly broader applications.

One example where quantum fingerprinting has been generalised is an efficient quantum communication protocol of Kumar *et al.* based on coherent states [19], which can approximately compute the Euclidean distance between unit vectors up to low additive error. This protocol is directly based on the use of the swap test to approximate $\ell_2$ distances between quantum states [7]. There are many other distance measures of practical relevance where it is less clear whether similar techniques to quantum fingerprinting can be applied.

### A. Our results

Our main result is a quantum protocol for approximately computing another distance measure, the Ham-

* joao.doriguellodiniz@bristol.ac.uk
† ashley.montanaro@bristol.ac.uk

ming distance, up to low *relative* error. This notion of accuracy is important when one wishes to compare objects that are similar; for example, when one of the objects is produced by a small number of errors affecting the other [20]. Approximating the Hamming distance between two $n$-bit strings up to additive accuracy $\epsilon n$ (analogous to the accuracy achieved by the protocol of [19]) would give no useful information in this situation.

In the setting we consider, Alice and Bob are given $x, y \in \{0,1\}^n$, respectively. Their goal is to approximately calculate the Hamming distance $d(x,y)$ between $x$ and $y$, i.e., they must output $d_\epsilon(x,y)$ such that $(1-\epsilon)d(x,y) \leq d_\epsilon(x,y) \leq (1+\epsilon)d(x,y)$. Pang and El Gamal [21] proved a lower bound of $\Omega(n)$ for exactly calculating the Hamming distance in the multi-round two-party classical communication model. Here we describe a quantum protocol that approximately computes the Hamming distance in the SMP model by communicating $\mathrm{poly}(\log n)$ qubits.

**Theorem 1.** *There is a quantum protocol in the SMP model with private randomness which communicates $O((\log n)^2(\log\log n)/\epsilon^5)$ qubits and computes the Hamming distance between $n$-bit strings up to relative error $\epsilon$, for any $\epsilon = \Omega(1/\log n)$, with failure probability bounded above by an arbitrarily small constant.*

The protocol is based on a subroutine which determines whether, for some threshold $\delta$, $d(x,y) \leq \delta$ or $d(x,y) \geq (1+\epsilon)\delta$. This subroutine maps $x$ and $y$ to $N$-bit strings $Ax$, $Ay$ such that in the first case, $d(Ax,Ay)$ is low (less than $\alpha N$, for some constant $\alpha$), whereas in the second case, $d(Ax,Ay)$ is high (greater than $\beta N$, for some constant $\beta > \alpha$). Alice and Bob then encode the strings $Ax$ and $Ay$ as quantum superpositions, which the referee can distinguish between using the swap test [7].

Note that there exists a corresponding classical protocol in the SMP model with shared randomness, with a similar complexity. One way to see this is that the quantum protocol is ultimately based on the use of the swap test to approximately compute the inner product between unit vectors, for which there is an efficient classical protocol in this model [22].

We then generalise Theorem 1 to other distance measures: in particular, those which can be interpreted as distances in graphs. A graph $G = (V,E)$ is fixed in advance, and each of Alice and Bob is given a vertex of $G$ ($v$ and $w$, respectively). They aim to approximately compute $d_G(v,w)$, the length of a shortest path in $G$ between $v$ and $w$, up to relative error $\epsilon$.

We first observe that Theorem 1 can be applied to give an efficient protocol for this problem whenever there is a distance-preserving embedding of $G$ into the hypercube: the graph whose vertex set is $\{0,1\}^m$, for some $m$, and where two vertices are connected by an edge whenever their Hamming distance is 1. In fact, this can be generalised further, to graphs which are embeddable into the hypercube such that distances are preserved up to a constant factor $k$. Such graphs are known as $\ell_1$-graphs, because it turns out that this criterion is equivalent to the existence of a distance-preserving embedding of the graph in $\ell_1$ [23]. The class of $\ell_1$-graphs includes all trees, median graphs, Hamming graphs, and Johnson graphs [23]. (We include in the Appendix a characterization of $\ell_1$-graphs which we were not able to find in the literature.)

Distances in $\ell_1$-graphs are used in a variety of applications, a few of which we outline here. Partial cubes ($\ell_1$-graphs with embedding constant $k = 1$) were initially introduced by Graham and Pollak [24] as a model for interconnection networks in the Bell System, with distances between vertices corresponding to the number of hops between 'loops' in their network. Antimatroids (a specific subclass of $\ell_1$-graphs) are used as structures to represent the required steps to develop a student's knowledge in a certain topic, and the distance between two points that represent concepts in these structures corresponds to the length of a student's learning path [25]. The Barnes-Hut tree method in many-body physics [26] provides a systematic way of determining the degree of 'closeness' between two different particles. The distance between two nodes in the tree is linked to this 'closeness' property and can be used for various purposes, e.g., to calculate gravitational forces in star clusters and study galaxy evolution [27]. Tree structures are also used in biology, where phylogenetic trees classify organisms based on overall similarity, and the distance between vertices is related to genetic or mutation distance [28].

Our protocol is efficient for $\ell_1$-graphs $G$ whose diameter $\mathrm{diam}(G)$ is low, where the diameter is defined as $\mathrm{diam}(G) = \max_{v,w} d_G(v,w)$.

**Theorem 2.** *Let $G = (V,E)$ be an $\ell_1$-graph with $|V|$ vertices, and let $v, w \in V$. There is a quantum protocol in the SMP model with private randomness which communicates $O((\log \mathrm{diam}(G))(\log\log \mathrm{diam}(G))(\log\log|V|)/\epsilon^5)$ qubits and computes $d_G(v,w)$ up to relative error $\epsilon$, for any $\epsilon = \Omega(1/\log \mathrm{diam}(G))$, with failure probability bounded above by an arbitrarily small constant.*

For any graph $G$, even testing equality between vertices requires $\Omega(\sqrt{\log|V|})$ bits of classical communication in the SMP model without shared randomness [5], so this is an exponential separation for those $\ell_1$-graphs where, for example, $\mathrm{diam}(G) = O(\log|V|)$. $d_G(v,w)$ can be computed trivially using $O(\log|V|)$ bits of classical communication, by sending the labels of $v$ and $w$ to the referee. So for graphs $G$ where $\mathrm{diam}(G)$ is close to $|V|$, Theorem 2 gives little or no improvement on the classical complexity. One may wonder whether this is simply a limitation of our protocol, but we show that this is not the case.

**Theorem 3.** *Given a graph $G$ with diameter $\mathrm{diam}(G)$, any one-way quantum communication protocol that computes $d_G(v,w)$ up to relative error $\epsilon < 1/4$ with failure probability $1/3$ must transmit at least $\Omega(\log \mathrm{diam}(G))$ qubits.*

As every protocol in the SMP model implies a one-way

protocol, this shows that the complexity of our protocol is nearly optimal in terms of its dependence on $\mathrm{diam}(G)$.

Finally, we show that our protocol for approximately computing the Hamming distance can be used to give an efficient protocol for approximately computing the $\ell_1$ distance between vectors in $\mathbb{R}^n$.

**Theorem 4.** *Let $x, y \in [-1, 1]^n$ such that each entry of $x$ and $y$ is specified by a $k$-bit string, with $k = O(\log n)$. There is a quantum protocol in the SMP model which communicates $O((\log n)^2(\log\log n)/\epsilon^5)$ qubits and computes $\|x - y\|_1$ up to relative error $\epsilon$, for any $\epsilon = \Omega(1/\log n)$, with failure probability bounded above by an arbitrarily small constant.*

A natural special case of Theorem 4 is where $x$ and $y$ are probability distributions. Then our result enables Alice and Bob to determine the distance between two distributions, one of which is a small perturbation of the other.

Two interesting questions which remain open are whether one can find a similar result to Theorem 2 which holds for all graphs, without the restriction to $\ell_1$-graphs, and if the communication complexity dependence on $\epsilon$, currently at $1/\epsilon^5$, can be improved.

### B. Related work

The Hamming distance is a fundamental distance measure and has been studied in various forms. In the context of quantum communication complexity, Liu and Zhang [29] gave a quantum sketching protocol for the related "threshold" problem of determining whether the Hamming distance is larger than $d$, for some $d$. Their protocol uses $O(d \log n)$ communication, improving a previous $O(d \log^2 n)$ protocol of Gavinsky, Kempe and de Wolf [30]. Huang *et al.* [31] had previously proven an $\Omega(d)$ lower bound for even the two-way quantum communication complexity of the threshold Hamming distance problem, together with an $O(d \log d)$ upper bound in the classical SMP model with public randomness.

A key ingredient in the upper bound of Huang *et al.* is a protocol which communicates $O(1)$ bits and distinguishes between the case that the Hamming distance is at most $d$, and the case that the Hamming distance is at least $2d$, for arbitrary $d$. Their protocol can be seen as a variant of our Lemma 2 below with $N = 1$; similar analysis shows that it could be generalised to distinguish between Hamming distance $d$ and Hamming distance $(1+\epsilon)d$ with $O(1/\epsilon^2)$ bits of communication. Using a generic construction of Yao [8], improved by Gavinsky, Kempe, and de Wolf [9], this implies a quantum sketching protocol for the same task which communicates $2^{O(1/\epsilon^2)} \log n$ qubits. Using a similar approach to our work, this in turn implies a protocol which solves the approximate Hamming distance problem by transmitting $2^{O(1/\epsilon^2)} \operatorname{poly}\log n$ qubits. This is the same asymptotic complexity as our protocol for constant $\epsilon$, but in practice the $2^{O(1/\epsilon^2)}$ factor makes the protocol infeasible for even modest values of $\epsilon$.

Classically, there has also been substantial work on approximately computing the Hamming distance between a small "pattern" and a larger string, both locally and in a distributed context (see [32] and references therein).

More generally, the field of communication complexity studies the amount of communication needed between two or more parties to solve a particular problem [1, 33]. We now give a brief summary of this area. The simplest and most illustrative scenario is the one in which two parties, called Alice and Bob, each possesses some piece of information, often encoded into some string, so that Alice has $x \in X$ and Bob has $y \in Y$, and they want to compute some function $f(x, y)$. Since each does not know the piece of information the other has, they will need to communicate information in order to compute $f(x, y)$. The most straightforward way to solve the problem is to have Alice and Bob exchange their entire string, but sometimes more efficient protocols exist. This communication model was first introduced by Yao in 1979 [34].

An important variant of this usual general communication scenario is the model of quantum communication complexity, again introduced by Yao [35], where now Alice and Bob each has a quantum computer and they exchange qubits instead of bits and/or make use of shared entanglement. The question is whether Alice and Bob can now compute $f$ with less communication than in the classical case; in some cases, this is known to be possible [1].

The above communication scenarios can be narrowed down by imposing some restrictions on the communication process between Alice and Bob, and by restricting or allowing resources like randomness and entanglement. The three most common communication models are the one-way, the multi-round two-party and the simultaneous message passing (SMP) models. In the multi-round two-party model both Alice and Bob can communicate with the other. On the other hand, in the one-way model only one party can communicate with the other, e.g. Alice communicates with Bob. Finally, in the SMP model Alice and Bob are only allowed to send messages to a third party, called the referee, who then computes $f(x, y)$. The SMP model was also introduced by Yao (1979) [34] and is the weakest reasonable model of communication complexity. Considering the SMP model in particular, Buhrman *et al.* [7] proved that, if $f$ is the equality function, then a communication reduction from $\Theta(\sqrt{n})$ bits to $\Theta(\log n)$ qubits is possible.

Later, Yao showed that any classical SMP protocol with shared randomness that transmits $O(1)$ bits and computes a function on $n$ bits implies a quantum SMP protocol without shared randomness that transmits $O(\log n)$ qubits [8]. This result was generalised by Gavinsky *et al.* [9], who gave a quantum SMP protocol that simulates any 2-way quantum communication protocol with shared entanglement, at communication cost exponential in the cost of the original protocol. How-

ever, Gavinsky *et al.* also proved that for most functions, quantum fingerprinting protocols, which are a subclass of quantum SMP protocols, are exponentially worse than classical deterministic SMP protocols.

Recently, more exotic communication models based on indefinite causal structures were used to demonstrate exponential quantum advantage. Wei *et al.* [36] and Guérin *et al.* [37] showed such an exponential communication advantage by using the concept of a quantum switch (a device that controls the order in which two transformations are performed) to coherently superpose the one-way communication path of information in a tripartite setting, i.e., from Alice to Bob and then to the referee or from Bob to Alice and then to the referee.

## II. THE PROTOCOL

In this section we present our protocol for approximating the Hamming distance $d(x,y)$ between two strings $x,y \in \{0,1\}^n$ up to relative error $\epsilon$ in the SMP model. That is, Alice and Bob seek the referee to output $d_\epsilon(x,y)$ such that $(1-\epsilon)d(x,y) \leq d_\epsilon(x,y) \leq (1+\epsilon)d(x,y)$. Call this problem $\mathrm{HAM}_\epsilon$.

We first state a lemma that is going to be useful for our protocol and which encapsulates results on quantum fingerprinting by Yao [8].

**Definition 1.** *Given an $N$-bit string $x$, define the quantum state*

$$|h_x\rangle = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} |i\rangle|x_i\rangle, \tag{1}$$

*where $x_i$ is the $i$-th bit of $x$.*

**Lemma 1** (Yao [8])**.** *Given the $N$-bit strings $x$ and $y$, their Hamming distance $d(x,y)$ can be estimated up to additive accuracy $N\epsilon$ with failure probability $\delta$ using $O(\log(1/\delta)/\epsilon^4)$ copies of $|h_x\rangle$ and $|h_y\rangle$.*

*Proof.* Given the $N$-bit strings $x$ and $y$, we encode them with the states $|h_x\rangle$ and $|h_y\rangle$, respectively. Note that

$$\langle h_y|h_x\rangle = \frac{1}{N}\sum_{i=1}^{N}\langle y_i|x_i\rangle = 1 - \frac{d(x,y)}{N}. \tag{2}$$

The swap test [7] is a test which outputs either 0 or 1 on input $|h_x\rangle|h_y\rangle$, and outputs 1 with probability

$$\frac{1}{2}\left(1 - |\langle h_y|h_x\rangle|^2\right). \tag{3}$$

We apply the swap test to $k$ copies of $|h_x\rangle|h_y\rangle$, for some $k$ to be determined. Let $X_i$ correspond to the outcome of the $i$-th swap test. In [8] it is proven that

$$\Pr\left[\left|\tilde{\eta} - |\langle h_y|h_x\rangle|\right| \geq \epsilon\right] \leq 2e^{-k\epsilon^4/32}, \tag{4}$$

where $\tilde{\eta} = \sqrt{1 - \frac{2}{k}\sum_i X_i}$. We hence conclude that

$$\Pr\left[\left|\tilde{d} - d(x,y)\right| \geq N\epsilon\right] \leq 2e^{-k\epsilon^4/32}, \tag{5}$$

where $\tilde{d} = N\left(1 - \sqrt{1 - \frac{2}{k}\sum_i X_i}\right)$. Setting $\delta$ as the probability of error, we see that it is sufficient to use $k = O(\log(1/\delta)/\epsilon^4)$ copies of the states to estimate $d(x,y)$ up to additive accuracy $N\epsilon$ with failure probability $\delta$. □

(Given that we aim to approximately compute the inner product between $|h_x\rangle$ and $|h_y\rangle$ in Lemma 1, the reader may wonder why the Hadamard test [38] was not used instead, given that this test allows direct estimation of $\langle h_y|h_x\rangle$. The reason is that the Hadamard test requires the ability to produce the coherent superposition $\frac{1}{\sqrt{2}}(|0\rangle|h_x\rangle + |1\rangle|h_y\rangle)$, which is not available to the referee.)

In the following, we use the notation $|z|$ to mean the number of entries equal to 1 in a string $z \in \{0,1\}^n$.

**Lemma 2.** *Consider an $N \times n$ matrix $A$ over $\mathbb{F}_2$ whose entries are randomly chosen from $\{0,1\}$, and equal to 1 with independent probability $1/(2d)$ for some $d \geq 1$. Fix $\epsilon > 0$. Then there exist values $\delta_1 < \delta_2$ that do not depend on $N$ and $n$, such that $\delta_2 - \delta_1 = \Theta(\epsilon)$ and for any $\eta > 0$:*

- *for all $z \in \{0,1\}^n$ such that $|z| \leq d$, $\Pr_A\left[|Az| \geq N\delta_1 + N\eta\right] \leq e^{-2N\eta^2}$;*
- *for all $z \in \{0,1\}^n$ such that $|z| \geq (1+\epsilon)d$, $\Pr_A\left[|Az| \leq N\delta_2 - N\eta\right] \leq e^{-2N\eta^2}$.*

*Hence, for sufficiently large $N = \Theta(n/\epsilon^2)$, with high probability over the choice of $A$, it is sufficient to determine $|Az|$ up to additive accuracy $\Theta(N\epsilon)$ to distinguish between the cases $|z| \leq d$ and $|z| \geq (1+\epsilon)d$.*

*Proof.* It is shown in [39] that for any $z$, $\Pr[(Az)_i = 1] = \frac{1}{2}\left(1 - (1-1/(2d))^{|z|}\right)$ and that the probabilities of this event for $|z| \leq d$ and $|z| \geq (1+\epsilon)d$ are bounded by values $\delta_1, \delta_2$ that do not depend on $N$ and $n$ and are separated by $\Theta(1 - e^{-\epsilon/2}) = \Theta(\epsilon)$. That is,

$$\Pr_A\left[(Az)_i = 1\right] \leq \delta_1 = \frac{1}{2}\left(1 - \left(1 - \frac{1}{2d}\right)^{d}\right) \quad \text{if } |z| \leq d, \tag{6a}$$

$$\Pr_A\left[(Az)_i = 1\right] \geq \delta_2 = \frac{1}{2}\left(1 - \left(1 - \frac{1}{2d}\right)^{(1+\epsilon)d}\right) \quad \text{if } |z| \geq (1+\epsilon)d. \tag{6b}$$

The expected value of $|Az| = \sum_i (Az)_i$ then satisfies

$$\mathbb{E}[|Az|] \leq N\delta_1 \text{ if } |z| \leq d, \tag{7a}$$

$$\mathbb{E}[|Az|] \geq N\delta_2 \text{ if } |z| \geq (1+\epsilon)d. \tag{7b}$$

If $|z| \leq d$ so that $\mathbb{E}[|Az|] \leq N\delta_1$, by a Chernoff bound [40] we obtain

$$\Pr_A\big[|Az| \geq N\delta_1 + N\eta\big] \leq e^{-2N\eta^2}. \tag{8}$$

By the same token, if $|z| \geq (1+\epsilon)d$, so that $\mathbb{E}|Az|] \geq N\delta_2$, we obtain

$$\Pr_A\big[|Az| \leq N\delta_2 - N\eta\big] \leq e^{-2N\eta^2}. \tag{9}$$

Taking a union bound over all $z \in \{0,1\}^n$ in both cases, we have

$$\begin{aligned}\Pr_A\big[\exists z \text{ s.t. } |z| \leq d \text{ and } |Az| \geq N\delta_1 + N\eta\big] \\ \leq 2^n e^{-2N\eta^2} = e^{n\ln 2 - 2N\eta^2},\end{aligned}$$

$$\begin{aligned}\Pr_A\big[\exists z \text{ s.t. } |z| \geq (1+\epsilon)d \text{ and } |Az| \leq N\delta_2 - N\eta\big] \\ \leq 2^n e^{-2N\eta^2} = e^{n\ln 2 - 2N\eta^2},\end{aligned} \tag{10}$$

so that it is sufficient to choose $N = \Omega(n/\eta^2)$ to bound the probability that either case occurs by an arbitrarily small constant. Choosing $\eta = c\epsilon$ for a sufficiently small constant $c$, we have $|Az| \leq N(\delta_1 + c\epsilon)$ if $|z| \leq d$, and $|Az| \geq N(\delta_2 - c\epsilon)$ if $|z| \geq (1+\epsilon)d$. Therefore, it is sufficient to determine $|Az|$ up to additive accuracy $O(N\epsilon)$ to distinguish these two cases. □

The map $A$ in Lemma 2 can be interpreted as a linear code. Such codes are also used in quantum fingerprinting protocols [7, 17], but here, unlike previous protocols, we choose the matrix $A$ to be sparse and random. This enables us to control its behaviour when acting on strings $z$ such that $|z| \approx d$, when $d$ is small.

We now describe our protocol based on the two previous Lemmas. In this protocol, Alice and Bob have already agreed beforehand on the matrix $A$, guaranteed to exist by Lemma 2, to be used. We stress that this matrix is fixed in advance and does not need to be chosen using shared randomness.

**Protocol 1.** *Consider the following subroutine for arbitrary $d \in [1,n]$ and $\delta > 0$: Alice and Bob encode their $n$-bit strings $x$ and $y$ as $Ax$ and $Ay$, respectively, where $A$ is picked according to Lemma 2 and multiplication is over $\mathbb{F}_2$. They send $O((\log 1/\delta)/\epsilon^4)$ copies of the quantum states $|h_{Ax}\rangle$ and $|h_{Ay}\rangle$ to the referee, who performs swap tests and estimates the Hamming distance $d(Ax,Ay)$ up to accuracy $N\epsilon$ with failure probability $\delta$. By Lemma 2, this is sufficient to determine whether $d(x,y) \leq d$ or $d(x,y) \geq (1+\epsilon)d$ with failure probability $\delta$.*

*Alice and Bob then apply this subroutine to the sequence $S$ of values $d$*

$$0, 1, 1+\epsilon, (1+\epsilon)^2, \ldots \tag{11}$$

*where the last element in $S$ corresponds to the minimal $k$ such that $(1+\epsilon)^{k+1} > n$; there are $O(\log n/\log(1+\epsilon)) = O((\log n)/\epsilon)$ elements in the sequence. (In the case $d = 0$, they use the standard quantum fingerprinting protocol instead.) Given the $O((\log n)/\epsilon)$ results, the referee outputs the minimal $\tilde{d}$ such that the subroutine returned "$d(x,y) \leq \tilde{d}$".*

We first show that, if each use of the subroutine succeeds, the overall algorithm achieves the required level of accuracy. By the definition of $S$, there exist consecutive elements $d_0, d_1, d_2 \in S$ such that $d_0 \leq d(x,y)/(1+\epsilon)$, $d(x,y)/(1+\epsilon) \leq d_1 \leq d(x,y)$, $d(x,y) \leq d_2 \leq (1+\epsilon)d(x,y)$. Then on input $d_2$ the subroutine must return "$d(x,y) \leq d_2$", while for input $d_0$ it must return "$d(x,y) \geq (1+\epsilon)d_0$", so the output $\tilde{d}$ is either $d_1$ or $d_2$ and hence

$$(1-\epsilon)d(x,y) \leq \frac{d(x,y)}{1+\epsilon} \leq \tilde{d} \leq (1+\epsilon)d(x,y).$$

Setting $\delta = O(\epsilon/\log n)$ and using a union bound over the $O((\log n)/\epsilon)$ uses of the subroutine, the probability that any of the subroutines fails can be upper-bounded by an arbitrarily small positive constant.

The overall communication complexity is

$$\begin{aligned}&O\left(((\log n)/\epsilon)\cdot(\log(1/\delta)/\epsilon^4)\cdot(\log n + \log 1/\epsilon)\right) = \\ &O\left((\log n)^2(\log\log n)/\epsilon^5\right),\end{aligned} \tag{12}$$

assuming that $\epsilon \geq 1/\log n$. This completes the proof of Theorem 1.

## III. MEASURING DISTANCES IN GRAPHS

In the following, for an arbitrary graph $G$ and vertices $v$, $w$, let $d_G(v,w)$ denote the distance between $v$ and $w$ in $G$, i.e. the length of a shortest path between $v$ and $w$. Also, the hypercube graph $Q_n$ is defined as the graph with vertex set $\{0,1\}^n$, where distance between vertices is the Hamming distance.

The algorithm from last section for approximately measuring the Hamming distance between two strings in the SMP model can be slightly modified to approximately compute the distance between two vertices in specific graphs in the SMP model. That is, to solve the following problem: for some graph $G = (V,E)$, and given vertices $v$, $w$ as input, output $\tilde{d}$ such that $(1-\epsilon)d_G(v,w) \leq \tilde{d} \leq (1+\epsilon)d_G(v,w)$. Call this problem $\mathrm{DIS}_\epsilon[G]$. The idea is to embed a given graph $G$ into a hypercube graph such that all the distances between vertices are preserved or rescaled by a constant factor. Once this embedding is achieved, the hypercube structure allows the equivalence between vertex distance in the graph and Hamming distance, so that a binary string can be associated with each vertex and the algorithm can be applied to these binary strings.

The downside of the above approach is that it cannot be applied to any given graph, since most graphs are not isometrically embeddable into a hypercube. The graphs

which can be isometrically embedded into hypercubes are known as partial cubes [41, 42].

The identification of which graphs are partial cubes is an interesting question by itself. The class of partial cubes is relatively broad. The most important examples are hypercubes, trees [43] and median graphs [44]. It also includes other significant classes, e.g. tope graphs of oriented matroids (specially graphs of regions of hyperplane arrangements) [45, 46], antimatroids [46, 47], weak orderings [46], bipartite $(6, 3)$-graphs [23], tiled partial cubes [48] and netlike partial cubes [49].

Partial cubes can be fully characterized via *Djoković's Characterization* [50, 51], introduced by Djoković in 1973. It connects the property of isometric embedding to bipartiteness and convexity of some specific sub-graphs of the original graph. Here a set is said to be convex if it is closed under taking shortest paths, i.e., if the shortest paths between any two points from the set are also contained in the set. Djoković's Characterization states, more specifically, that a connected graph $G$ can be isometrically embedded into a hypercube if and only if $G$ is bipartite and $G(a|b)$ is convex for each edge $(a, b)$ of $G$, where $G(a|b) := \{x \in V(G)|\ d_G(x, a) < d_G(x, b)\}$ is the set of the vertices closer to $a$ than $b$. In other words, to check if a graph is a partial cube, one needs to check first if the graph is bipartite. Apart from that, one chooses an edge and constructs the set of all vertices that are closer to one vertex of the chosen edge than the other vertex. Then one needs to check if all shortest paths connecting any two vertices from this set only pass through the vertices of the set. If yes, the set is said to be convex and the same procedure is repeated for another edge of the original graph. If all sets constructed in this way are convex, then the graph is a partial cube.

Since the original protocol is unaffected if all distances are rescaled by a constant factor, the idea of partial cubes can be expanded by the following definitions.

**Definition 2** ([23, 52])**.** *Given two connected and unweighted graphs $G$ and $H$, we write $G \overset{k}{\hookrightarrow} H$ and say that $G$ is a scale $k$ embedding of $H$ if there exists a mapping $\sigma : V(G) \to V(H)$ such that $d_H(\sigma(a), \sigma(b)) = k \cdot d_G(a, b)$ for all nodes $a, b \in V(G)$.*

It is clear that partial cubes are just graphs which can be embedded in a hypercube with a scale 1 embedding. An example of a graph which is not a partial cube, but can be embedded in a hypercube with a scale $k$ embedding for $k > 1$, is a triangle, which embeds into $Q_3$ with $k = 2$.

**Definition 3** ([51])**.** *A graph $G$ is said to be an $\ell_1$-graph if its path metric $d_G$ is $\ell_1$-embeddable, i.e. there is a map $f$ between $V(G)$ and $\mathbb{R}^m$, for some $m$, such that $d_G(v, w) = \|f(v) - f(w)\|_1$.*

**Theorem 5** ([23])**.** *A graph $G$ is an $\ell_1$-graph if and only if it admits a scale embedding into a hypercube.*

This means that the graphs we are interested in are $\ell_1$-graphs. This class of $\ell_1$-graphs includes new graphs that are not partial cubes, e.g. Hamming graphs, half cubes and Johnson graphs are 2-embeddable into a hypercube [23]. In the Appendix we developed a similar characterization for $\ell_1$-graphs and the final result is the following theorem, which is Djokovic's characterization without the bipartite requirement.

**Theorem 6.** *A graph $G$ is an $\ell_1$-graph if and only if $G(a|b)$ is convex for each edge $(a, b)$ of $G$.*

By allowing the rescaling of all the distances by an even factor we can relax the bipartite requirement, but not the convexity of the $G(a|b)$ subgraphs. As an example of a direct consequence of the above result, it is known that graphs of the form $C_{2n}$ and $C_{2n}\square K_2$ for $n \geq 2$ are partial cubes, where $C_n$ is a cycle on $n$ vertices, $K_n$ is the complete graph with $n$ vertices, and $\square$ denotes the Cartesian product [41]; therefore all graphs of the form $C_n$ and $C_n\square K_2$, for $n \geq 2$, are $\ell_1$-graphs.

Before stating the communication protocol in the SMP model to approximately measure the distance between two vertices in an $\ell_1$-graph, we state the Johnson-Lindenstrauss lemma [9, 53, 54], which is going to be useful to reduce the protocol complexity. Note that we use Dirac notation for vectors which are not necessarily normalized.

**Lemma 3** (Johnson-Lindenstrauss lemma)**.** *Consider $0 < \epsilon < 1/2$ and a positive integer $n$. Then for any set $U$ of $k$ vectors in $\mathbb{R}^n$, there is a linear map $f : \mathbb{R}^n \to \mathbb{R}^{O((\log k)/\epsilon^2)}$ such that for all $|u\rangle, |v\rangle \in U$,*

$$(1-\epsilon)\||u\rangle - |v\rangle\|^2 \leq \|f|u\rangle - f|v\rangle\|^2 \leq (1+\epsilon)\||u\rangle - |v\rangle\|^2.$$

To find a map $f$ achieving the bounds of Lemma 3, one can choose it at random from an appropriate distribution. A number of different constructions of such random maps are known; one simple example is a suitably normalised projection onto a random subspace of $\mathbb{R}^n$.

As mentioned, e.g., in [9], if the set $U$ includes the 0-vector, then the map $f$ also approximately preserves the inner product between all the pairs of vectors in $U$. This implies the following Lemma.

**Lemma 4.** *Let $0 < \epsilon < 1/2$. Let $U$ be a set of unit vectors in $\mathbb{R}^n$ and let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map such that, for all $|u\rangle, |v\rangle \in U \cup \{\vec{0}\}$,*

$$(1-\epsilon)\||u\rangle - |v\rangle\|^2 \leq \|f|u\rangle - f|v\rangle\|^2 \leq (1+\epsilon)\||u\rangle - |v\rangle\|^2.$$

*Define the unit vectors $|\widetilde{u}\rangle = f|u\rangle/\|f|u\rangle\|$ for all $|u\rangle \in U$. Then*

$$\Big|\big|\langle\widetilde{u}|\widetilde{v}\rangle\big| - \big|\langle u|v\rangle\big|\Big| \leq 4\epsilon$$

*for all $|u\rangle, |v\rangle \in U$.*

*Proof.* For clear notation, define $|u'\rangle := f|u\rangle$. By the conditions on $f$, we have that

$$\begin{cases} 1-\epsilon \leq \langle u'|u'\rangle \leq 1+\epsilon, \\ (1-\epsilon)\||u\rangle - |v\rangle\|^2 \leq \||u'\rangle - |v'\rangle\|^2 \leq (1+\epsilon)\||u\rangle - |v\rangle\|^2 \end{cases}$$

for all $|u\rangle, |v\rangle \in U$, where the first line was obtained by taking the 0-vector as one of the vectors and using linearity of $f$. From the above inequalities it follows that

$$(1+\epsilon)\langle u|v\rangle - 2\epsilon \leq \langle u'|v'\rangle \leq (1-\epsilon)\langle u|v\rangle + 2\epsilon.$$

These new inequalities in turn lead to

$$\langle \tilde{u}|\tilde{v}\rangle \geq \frac{(1+\epsilon)\langle u|v\rangle - 2\epsilon}{1+\epsilon} \geq \langle u|v\rangle - 2\epsilon$$

and

$$\langle \tilde{u}|\tilde{v}\rangle \leq \frac{(1-\epsilon)\langle u|v\rangle + 2\epsilon}{1-\epsilon} \leq \langle u|v\rangle + 4\epsilon,$$

using that $0 < \epsilon < 1/2$. Therefore

$$\Big|\big|\langle \tilde{u}|\tilde{v}\rangle\big| - \big|\langle u|v\rangle\big|\Big| \leq \Big|\langle \tilde{u}|\tilde{v}\rangle - \langle u|v\rangle\Big| \leq 4\epsilon.$$

□

Consider applying Lemma 1 to the normalized quantum states $|\widetilde{h}_x\rangle$ and $|\widetilde{h}_y\rangle$ that are produced by using the Johnson-Lindenstrauss lemma, in the sense that the original states $|h_x\rangle$, $|h_y\rangle$ in Lemma 1 are replaced with the states $|\widetilde{h}_x\rangle$, $|\widetilde{h}_y\rangle$. We argue that this does not change the parameters of the lemma substantially. To see that, we note $\big|\tilde{\eta} - |\langle h_y|h_x\rangle|\big| + \big||\langle \widetilde{h}_y|\widetilde{h}_x\rangle| - |\langle h_y|h_x\rangle|\big| \geq \big|\tilde{\eta} - |\langle \widetilde{h}_y|\widetilde{h}_x\rangle|\big|$ and hence $\big|\tilde{\eta} - |\langle \widetilde{h}_y|\widetilde{h}_x\rangle|\big| \geq 5\epsilon \implies \big|\tilde{\eta} - |\langle h_y|h_x\rangle|\big| \geq \epsilon$, which means

$$\Pr\big[\big|\tilde{\eta} - |\langle \widetilde{h}_y|\widetilde{h}_x\rangle|\big| \geq 5\epsilon\big] \leq \Pr\big[\big|\tilde{\eta} - |\langle h_y|h_x\rangle|\big| \geq \epsilon\big], \quad (13)$$

where $\tilde{\eta}$ is as defined in Lemma 1.

With this in mind, and recalling that $\mathrm{diam}(G)$ is defined to be the diameter of the graph $G$, i.e., the greatest distance between any pair of vertices, we present the communication protocol.

**Protocol 2.** *Alice and Bob each hold vertices $u, v \in V(G)$, respectively, from a graph $G$ which admits a scale $k$ embedding into a hypercube $Q_n$, for some $n$. Their vertex images are the $n$-bit strings $x, y \in Q_n$, respectively. The communication protocol to measure $(1 \pm \epsilon)d_G(u, v)$ can be divided into three parts.*

*First, given $d \in [1, \mathrm{diam}(G)]$ and a matrix $A$ picked according to Lemma 2, Alice and Bob encode their $n$-bit strings $x$ and $y$ as $Ax$ and $Ay$, respectively, where multiplication is over $\mathbb{F}_2$. Differently from the original protocol, Alice and Bob apply the Johnson-Lindenstrauss lemma to their data $Ax$ and $Ay$, which are then encoded into the quantum states $|\widetilde{h}_{Ax}\rangle$ and $|\widetilde{h}_{Ay}\rangle$. There are $|V|$ possible vectors to encode, so the number of qubits to be used is reduced from $O(\log n + \log(1/\epsilon))$ to $O(\log\log|V| + \log(1/\epsilon))$.*

*Second, Alice and Bob send $O((\log 1/\delta)/\epsilon^4)$ copies of their quantum states $|\widetilde{h}_{Ax}\rangle$ and $|\widetilde{h}_{Ay}\rangle$ to the referee, who performs swap tests and estimates the Hamming distance $d(Ax, Ay)$ up to accuracy $N\epsilon$ with failure probability $\delta$, and from this decides if $d(x, y) \leq d$ or $d(x, y) \geq (1+\epsilon)d$.*

*The third and final part is to apply the first and second parts to the sequence $S$ of values $d$*

$$0, 1, 1+\epsilon, (1+\epsilon)^2, \ldots \quad (14)$$

*where the last element in $S$ corresponds to the minimal $k$ such that $(1+\epsilon)^{k+1} > \mathrm{diam}(G)$; there are $O((\log \mathrm{diam}(G))/\epsilon)$ elements in the sequence. Based on the results from the swap tests, the referee outputs $\tilde{d}$ such that $(1-\epsilon)d(x, y) \leq \tilde{d} \leq (1+\epsilon)d(x, y)$, in the same way as in Protocol 1.*

Setting $\delta = O(\epsilon/\log \mathrm{diam}(G))$, the overall communication complexity is then

$$O((\log \mathrm{diam}(G))(\log\log \mathrm{diam}(G))(\log\log|V|)/\epsilon^5), \quad (15)$$

assuming that $\epsilon \geq 1/(\log \mathrm{diam}(G))$.

The performance of the protocol is limited by the diameter of the graph. It is known that $\log_{\Delta-1}|V| - \frac{2}{\Delta} \leq \mathrm{diam}(G) < |V|$, where $\Delta$ is the maximum vertex degree [55]. If $\mathrm{diam}(G) = O(\log|V|)$, the overall complexity is polyloglog in $|V|$. On the other hand, if $\mathrm{diam}(G) = \Theta(|V|)$, the overall complexity is polylog in $|V|$, which is no better than the trivial protocol where Alice and Bob send their entire inputs to the referee.

### A. Lower bound

One can ask if there could exist other protocols substantially more efficient than ours. In order to answer this, we prove a lower bound on the quantum communication complexity for the problem of approximately calculating the graph distance between two vertices on a graph, which demonstrates that our protocol is essentially optimal in terms of the dependence of its complexity on the graph diameter. We do not know whether the 5th-power dependence on $\epsilon$ is optimal, and suspect that it may not be.

The idea behind our proof is to transform the approximate graph distance problem into the problem of approximating the modulus of the difference between two integers. We then show that two uses of a protocol for this approximate modulus problem can compute the greater than function in the one-way communication model. It was shown by Zhang [56] that the one-way quantum communication complexity of this problem is maximal, improving a previous lower bound of Klauck [57] by a logarithmic term. The bound of [56] is used to obtain the lower bound for the approximate modulus problem, and hence for the approximate graph distance problem.

The first step of our proof is to show that two uses of a protocol for the approximate modulus problem can solve the greater than function in the one-way communication model. Consider the greater than problem (GT) defined by the Boolean function $\mathrm{GT} : \{0,1\}^m \times \{0,1\}^m \to \{0,1\}$ as

$$\mathrm{GT}(x,y) = \begin{cases} 1 \text{ if } x \geq y, \\ 0 \text{ if } x < y, \end{cases}$$

where $x$ and $y$ are interpreted as $m$-bit integers. Given $0 \leq \epsilon < 1$, consider the approximate modulus problem where Alice and Bob are each given an integer $x$ and $y$ (respectively), each expressed as an $m$-bit string, and seek to output $\tilde{d}$ such that $(1-\epsilon)|x-y| \leq \tilde{d} \leq (1+\epsilon)|x-y|$. Call this problem $\mathrm{MOD}_\epsilon$. In the following we prove that two uses of this protocol suffice to solve the GT problem.

Let $\mathcal{P}$ be a quantum communication protocol in the one-way communication model which solves a problem $f$ with failure probability $\delta$. Denote by $Q^1(\mathcal{P})$ the communication cost of the protocol $\mathcal{P}$ (in qubits) and denote by $Q^1(f) = \min_{\mathcal{P}} Q^1(\mathcal{P})$ the minimum communication cost over all protocols $\mathcal{P}$ that solve $f$ with failure probability $1/3$.

**Lemma 5.** *For any $\epsilon < 1/4$, $Q^1(GT) = O(Q^1(MOD_\epsilon))$.*

*Proof.* Let $\mathcal{P}_{MOD}$ be a communication protocol for $\mathrm{MOD}_\epsilon$ in the one-way communication model with failure probability $1/6$. (We can obtain a protocol which achieves this failure probability and communicates $O(Q^1(\mathrm{MOD}_\epsilon))$ qubits using $O(1)$ repetitions of the protocol which achieves failure probability $1/3$ and communicates $Q^1(\mathrm{MOD}_\epsilon)$ qubits.)

Two uses of $\mathcal{P}_{MOD}$ suffice to obtain a communication protocol for GT in the one-way communication model with failure probability $1/3$ as follows: Alice and Bob apply the protocol $\mathcal{P}_{MOD}$ using $x$ and $y$ as inputs and Bob obtains $z_0 \in [(1-\epsilon)|x-y|, (1+\epsilon)|x-y|]$. They both apply the same protocol again, but now Bob inputs $y + z_0$ (Alice still inputs $x$). Bob obtains $z_1$. If $z_0 < z_1$, then $x < y$ and he outputs 0. Otherwise, $x \geq y$ and he outputs 1.

To see why this protocol works (assuming that each use of $\mathcal{P}_{MOD}$ succeeds), note that if $x < y$, then $(2-\epsilon)|x-y| \leq |x-y-z_0| \leq (2+\epsilon)|x-y|$, and hence

$$(2-\epsilon)(1-\epsilon)|x-y| \leq z_1 \leq (2+\epsilon)(1+\epsilon)|x-y|.$$

If $x \geq y$, then $0 \leq |x-y-z_0| \leq \epsilon|x-y|$, and hence

$$0 \leq z_1 \leq \epsilon(1+\epsilon)|x-y|.$$

For $x < y$ we want to have $z_0 < z_1$, i.e. $1+\epsilon < (2-\epsilon)(1-\epsilon)$, which holds if $\epsilon < 2-\sqrt{3}$. And for $x \geq y$ we need $z_0 \geq z_1$, i.e. $\epsilon(1+\epsilon) \leq 1-\epsilon$, which holds if $\epsilon \leq \sqrt{2}-1$. Therefore, by taking $\epsilon < 1/4$, for example, one can distinguish the cases $x < y$ and $x \geq y$ through a comparison between $z_0$ and $z_1$.

Given that every protocol for $\mathrm{MOD}_\epsilon$ in the one-way communication model implies a protocol for GT, we conclude that $Q^1(\mathrm{GT}) = O(Q^1(\mathrm{MOD}_\epsilon))$. □

The next step is to reduce the approximate modulus problem to the approximate graph distance problem, which we achieve as follows. Let $G$ be a graph with diameter $\mathrm{diam}(G)$. By the definition of diameter, there exists a path graph $P_n \subseteq G$ with $n = \mathrm{diam}(G)$. Therefore, a lower bound for the approximate graph distance problem on $P_n$ implies a lower bound for the same problem on $G$.

The vertices of $P_n$ can be listed in the order $v_1, v_2, ..., v_n$ such that the edges are $(v_i, v_{i+1})$, where $i = 1, 2, ..., n-1$. A given vertex $v_i$ can then be labeled by a binary string $x_i \in \{0,1\}^m$, with $m = \Theta(\log n)$, and hence, given $v_i, v_j \in G$, $d_G(v_i, v_j) = |x_i - x_j|$. Therefore, a communication protocol which outputs $\tilde{d}$ such that $(1-\epsilon)d_G(v_i,v_j) \leq \tilde{d} \leq (1+\epsilon)d_G(v_i,v_j)$ is equivalent to a communication protocol which solves $\mathrm{MOD}_\epsilon$ on inputs $x_i$, $x_j$. So computing an approximate modulus reduces to computing an approximate graph distance.

With this in mind, we can state our lower bound.

**Theorem 7.** *Given a graph $G$ with diameter $\mathrm{diam}(G)$, the quantum communication complexity for the problem $\mathrm{DIS}_\epsilon[G]$ in the one-way communication model with $\epsilon < 1/4$ and failure probability $1/3$ is $Q^1(\mathrm{DIS}_\epsilon[G]) = \Omega\big(\log \mathrm{diam}(G)\big)$.*

*Proof.* As mentioned before, the approximate graph distance problem on a path graph $P_n \subseteq G$ with $n = \mathrm{diam}(G)$ should be at least as hard as the same problem on $G$, i.e. $Q^1(\mathrm{DIS}_\epsilon[G]) \geq Q^1(\mathrm{DIS}_\epsilon[P_n])$. Moreover, $\mathrm{DIS}_\epsilon[P_n]$ is equivalent to $\mathrm{MOD}_\epsilon$ on inputs of size $m = \Theta\big(\log \mathrm{diam}(G)\big)$, hence $Q^1(\mathrm{DIS}_\epsilon[G]) \geq Q^1(\mathrm{MOD}_\epsilon)$. According to Lemma 5, $Q^1(\mathrm{MOD}_\epsilon) = \Omega(Q^1(\mathrm{GT}))$, but $Q^1(\mathrm{GT}) = \Theta(m)$ [56], therefore $Q^1(\mathrm{DIS}_\epsilon[G]) = \Omega\big(\log \mathrm{diam}(G)\big)$. □

The above result for the one-way communication model also holds for the SMP model. It then states that our communication protocol is optimal in terms of its dependence on $\mathrm{diam}(G)$.

## IV. MEASURING $\ell_1$ DISTANCES

As seen in the previous sections, our communication protocol for approximating the Hamming distance can be adapted to $\ell_1$-graphs. A graph $G$ is said to be an $\ell_1$-graph if there exist vectors $u_1, ..., u_n \in \mathbb{R}^m$ for some $m$, and with $n = |V(G)|$, such that $d_G(v_i, v_j) = \|u_i - u_j\|_1$ for all $v_i, v_j \in V(G)$. This connection between graphs and $\ell_1$-norm suggests an application of our approximate Hamming distance protocol to $\ell_1$ distances. More specifically, consider the following problem: Alice and Bob are each given a vector $x, y$ (respectively) from $[-1,1]^d$. Each entry of each vector is specified by $k$ bits, for some $k$ (1 bit to specify the sign, and $k-1$ bits $z_1, \ldots, z_{k-1}$ to

specify a binary fraction $z_1 2^{-1} + z_2 2^{-2} + \cdots + z_{k-1} 2^{1-k}$). Alice and Bob's task is to approximate the $\ell_1$ distance between $x$ and $y$ up to relative error $\epsilon$ in the SMP model.

A natural special case of this problem is where Alice and Bob are each given a probability distribution $x, y \in \mathbb{R}^d$, respectively, and are asked to approximately compute the $\ell_1$ distance between them (equivalently, the total variation distance, which is defined as half the $\ell_1$ distance). This corresponds to the special case where $x_i, y_i \geq 0$ for all $i$, and $\sum_i x_i = \sum_i y_i = 1$.

Alice and Bob can use our approximate Hamming distance protocol to approximately compute $\|x - y\|_1$: the idea is to map these vectors into a Hamming metric via a unary representation [58]. Each entry $z \in [-1, 1]$ of each vector is mapped to a $2^k$-bit string $s(z)$ such that the first $2^{k-1}(z + 1)$ bits of $s(z)$ are set to 1, and the remaining bits are set to 0. Then, for any $z$, $w$, $|z - w| = d(s(z), s(w))/2^{k-1}$. Letting $s(x)$ denote the result of applying this map to each entry of $x$ and concatenating the results, we have $\|x - y\|_1 = d(s(x), s(y))/2^{k-1}$ for bit strings $s(x)$, $s(y)$ of length $m = 2^k d$. So we can use our usual communication protocol (Protocol 1) to deliver an estimate of $\|x - y\|_1$ up to relative error $\epsilon$ using $O((\log^2 m)(\log \log m)/\epsilon^5)$ qubits of communication, which is $O((\log^2 d)(\log \log d)/\epsilon^5)$ when $k \leq \log d$.

The use of a unary representation may seem wasteful, but a straightforward binary representation would not preserve distances correctly for all inputs. There is also a lower bound that the communication complexity of this problem must have at least a linear dependence on $k$: by the lower bound on the complexity of the $\mathrm{MOD}_\epsilon$ problem that follows from Lemma 5, $\Omega(k)$ bits of communication are required to approximately compute $\|x - y\|_1$ even for $d = 1$. Finally, the protocol can easily be extended to the setting where $x, y \in [-M, M]^d$, for some $M \geq 1$, by rescaling the vectors appropriately.

## V. CONCLUSIONS

We developed an efficient quantum communication protocol to approximately compute the Hamming distance between two $n$-bit strings in the SMP model up to relative error $\epsilon$, which uses $\tilde{O}((\log n)^2/\epsilon^5)$ qubits of communication, whereas any classical communication protocol needs to transmit at least $\Omega(\sqrt{n})$ bits. We stress that the protocol approximates the Hamming distance up to a relative, and not additive, error, so that small Hamming distances are approximated accurately.

The Hamming distance protocol was modified to approximate the distance between any two vertices in a graph. This modification was based on embedding the graph into a subgraph of the Hamming cube such that all distances are preserved up to a constant factor. This requirement restricts the class of graphs to which the original Hamming distance protocol can be applied. Graphs with this property are known as $\ell_1$-graphs. The modified quantum communication protocol to approximate the vertex distance in $\ell_1$-graphs in the SMP model up to relative error $\epsilon$ transmits $\tilde{O}(\log(\mathrm{diam}(G))/\epsilon^5)$ qubits, where $\mathrm{diam}(G)$ is the diameter of the graph, so the protocol is only efficient for low diameter graphs. A lower bound on the number of communicated qubits needed to approximate the vertex distance shows that this limitation of our protocol is due to the problem itself. More specifically, we proved that any one-way quantum protocol to approximate the distance between any two vertices in a graph needs to communicate at least $\Omega(\log \mathrm{diam}(G))$ qubits. Finally, the original Hamming distance protocol was also modified to approximate the $\ell_1$-norm between two vectors $x, y \in [-1, 1]^d$ specified by $k \leq \log d$ bits.

Two interesting questions remain open: can one find a similar result to Theorem 2 which holds for all graphs, without the restriction to $\ell_1$-graphs? And can one improve the communication complexity dependence on $\epsilon$, currently at $1/\epsilon^5$?

No new data were created during this study.

### Acknowledgements

We would like to thank Raphaël Clifford for helpful discussions during the course of this work. We acknowledge support from the QuantERA ERA-NET Cofund in Quantum Technologies implemented within the European Union's Horizon 2020 Programme (QuantAlgo project). AM was supported by EPSRC Early Career Fellowship Grant No. EP/L021005/1. JFD was supported by the Bristol Quantum Engineering Centre for Doctoral Training, EPSRC Grant No. EP/L015730/1.

[1] H. Buhrman, R. Cleve, S. Massar, and R. de Wolf. Nonlocality and communication complexity. *Rev. Mod. Phys.*, 82(1):665–698, 2010.

[2] I. Mironov, M. Naor, and G. Segev. Sketching in adversarial environments. *SIAM Journal on Computing*, 40(6):1845–1870, 2011.

[3] J. Feigenbaum, Y. Ishai, T. Malkin, K. Nissim, M. Strauss, and R. Wright. Secure multiparty computation of approximations. *ACM Transactions on Algorithms*, 2:435–472, 2006.

[4] Z. Bar-Yossef, T. Jayram, R. Krauthgamer, and R. Kumar. Approximating edit distance efficiently. In *Proc. 45th Annual Symp. Foundations of Computer Science*, pages 550–559, 2004.

[5] A. Ambainis. Communication complexity in a 3-computer model. *Algorithmica*, 16(3):298–301, 1996.

[6] L. Babai and P. G. Kimmel. Randomized simultaneous messages: solution of a problem of yao in communication complexity. In *Proceedings of Computational Complexity. Twelfth Annual IEEE Conference*, pages 239–246, 1997.

[7] H. Buhrman, R. Cleve, J. Watrous, and R. de Wolf. Quantum fingerprinting. *Phys. Rev. Lett.*, 87:167902, 2001.

[8] A. Yao. On the power of quantum fingerprinting. In *Proceedings of the Thirty-fifth Annual ACM Symposium on Theory of Computing*, STOC '03, pages 77–81, New York, NY, USA, 2003. ACM.

[9] D. Gavinsky, J. Kempe, and R. de Wolf. Strengths and weaknesses of quantum fingerprinting. In *Proc. 21st Annual IEEE Conf. Computational Complexity*, pages 288–298. IEEE, 2006.

[10] A. Winter. Quantum and classical message identification via quantum channels. *Festschrift "A S Holevo 60" (O. Hirota, ed.)*, pages 171–188, 2004.

[11] S. Massar. Quantum fingerprinting with a single particle. *Phys. Rev. A*, 71(1):012310, 2005.

[12] J. N. de Beaudrap. One-qubit fingerprinting schemes. *Phys. Rev. A*, 69:022307, 2004.

[13] A. Scott, J. Walgate, and B. Sanders. Optimal fingerprinting strategies with one-sided error. *Quantum Information & Computation*, 7(3):243–264, 2007.

[14] R. T. Horn, S. A. Babichev, K.-P. Marzlin, A. I. Lvovsky, and B. C. Sanders. Single-qubit optical quantum fingerprinting. *Phys. Rev. Lett.*, 95(15):150502, 2005.

[15] J. Du, P. Zou, X. Peng, D. Oi, L. C. Kwek, C. H. Oh, and A. Ekert. Experimental quantum multimeter and one-qubit fingerprinting. *Phys. Rev. A*, 74(4):042319, 2006.

[16] J. Arrazola and N. Lütkenhaus. Quantum fingerprinting with coherent states and a constant mean number of photons. *Phys. Rev. A*, 89:062305, 2014.

[17] F. Xu, J. M. Arrazola, K. Wei, W. Wang, P. Palacios-Avila, C. Feng, S. Sajeed, N. Lütkenhaus, and H.-K. Lo. Experimental quantum fingerprinting with weak coherent pulses. *Nature Communications*, 6:8735, 2015.

[18] J.-Y. Guan, F. Xu, H.-L. Yin, Y. Li, W.-J. Zhang, S.-J. Chen, X.-Y. Yang, L. Li, L.-X. You, T.-Y. Chen, Z. Wang, Q. Zhang, and J.-W. Pan. Observation of quantum fingerprinting beating the classical limit. *Phys. Rev. Lett.*, 116(24):240502, 2016.

[19] N. Kumar, E. Diamanti, and I. Kerenidis. Efficient quantum communications with coherent state fingerprints over multiple channels. *Phys. Rev. A*, 95:032337, 2017.

[20] G. Cormode, M. Paterson, S. Sahinalp, and U. Vishkin. Communication complexity of document exchange. In *Proc. 11th ACM-SIAM Symp. Discrete Algorithms*, pages 197–206, 2000.

[21] K. Pang and A. El Gamal. Communication complexity of computing the hamming distance. *SIAM Journal on Computing*, 15(4):932–947, 1986.

[22] I. Kremer, N. Nisan, and D. Ron. On randomized one-round communication complexity. *Computational Complexity*, 8:21–49, 1999.

[23] H.-J. Bandelt and V. Chepoi. Metric graph theory and geometry: A survey. *Discrete & Computational Geometry - DCG*, 453, 01 2008.

[24] R. L. Graham and H. O. Pollak. On the addressing problem for loop switching. *The Bell System Technical Journal*, 50(8):2495–2519, 1971.

[25] J.-C. Falmagne, E. Cosyn, J.-P. Doignon, and N. Thiéry. The assessment of knowledge, in theory and in practice. In *Formal concept analysis*, pages 61–79. Springer, 2006.

[26] J. Barnes and P. Hut. A hierarchical $o(N \log N)$ force-calculation algorithm. *Nature*, 324(6096):446–449, 1986.

[27] S. Pfalzner and P. Gibbon. *Many-body tree methods in physics*. Cambridge University Press, 2005.

[28] W. M. Fitch and E. Margoliash. Construction of phylogenetic trees. *Science*, 155(3760):279–284, 1967.

[29] Y. Liu and S. Zhang. Quantum and randomized communication complexity of xor functions in the smp model. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 20, page 10, 2013.

[30] D. Gavinsky, J. Kempe, and R. de Wolf. Quantum communication cannot simulate a public coin, 2004. `quant-ph/0411051`.

[31] W. Huang, Y. Shi, S. Zhang, and Y. Zhu. The communication complexity of the Hamming distance problem. *Information Processing Letters*, 99(4):149–153, 2006.

[32] R. Clifford and T. Starikovskaya. Approximate hamming distance in a stream. In *43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016)*, pages 20:1–20:14, Dagstuhl, Germany, 2016. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

[33] E. Kushilevitz and N. Nisan. *Communication Complexity*. Cambridge University Press, New York, NY, USA, 1997.

[34] A. Yao. Some complexity questions related to distributive computing. In *Proceedings of the Eleventh Annual ACM Symposium on Theory of Computing*, STOC '79, pages 209–213, New York, NY, USA, 1979. ACM.

[35] A. Yao. Quantum circuit complexity. In *Proc. 34th Annual Symp. Foundations of Computer Science*, pages 352–361. IEEE, 1993.

[36] K. Wei, N. Tischler, S.-R. Zhao, Y.-H. Li, J. M. Arrazola, Y. Liu, W. Zhang, H. Li, L. You, Z. Wang, Y.-A. Chen, B. C. Sanders, Q. Zhang, G. J. Pryde, F. Xu, and J.-W. Pan. Experimental quantum switching for exponentially superior quantum communication complexity. *Phys. Rev. Lett.*, 122:120504, 2019.

[37] P. A. Guérin, A. Feix, M. Araújo, and Č. Brukner. Exponential communication complexity advantage from quantum superposition of the direction of communication. *Phys. Rev. Lett.*, 117:100502, 2016.

[38] D. Aharonov, V. Jones, and Z. Landau. A polynomial quantum algorithm for approximating the Jones polynomial. *Algorithmica*, 55(3):395–421, 2009.

[39] E. Kushilevitz, R. Ostrovsky, and Y. Rabani. Efficient search for approximate nearest neighbor in high dimensional spaces. *SIAM Journal on Computing*, 30(2):457–474, 2000.

[40] D. P Dubhashi and A. Panconesi. *Concentration of measure for the analysis of randomized algorithms*. Cambridge University Press, 2009.

[41] C. Bonnington, S. Klavzar, and A. Lipovec. On cubic and edge-critical isometric subgraphs of hypercubes. *Australasian J. Combinatorics*, 28:217–224, 2003.

[42] S. Ovchinnikov. Partial cubes: structures, characterizations, and constructions. *Discrete Mathematics*, 308(23):5597–5621, 2008.

[43] A. Y. Wu. Embedding of tree networks into hypercubes. *Journal of Parallel and Distributed Computing*, 2(3):238–249, 1985.

[44] S. Ovchinnikov. *Graphs and cubes*. Springer Science & Business Media, 2011.

[45] A. Björner, M. Las Vergnas, B. Sturmfels, N. White, and G. M. Ziegler. *Oriented matroids*, volume 46. Cambridge University Press, 1999.

[46] D. Eppstein, J.-C. Falmagne, and S. Ovchinnikov. *Media theory: interdisciplinary applied mathematics*. Springer Science & Business Media, 2007.

[47] Y. Kempner and V. E. Levit. Geometry of poset antimatroids. *Electronic Notes in Discrete Mathematics*, 40:169–173, 2013.

[48] B. Brešar, W. Imrich, S. Klavžar, H. M. Mulder, and R. Škrekovski. Tiled partial cubes. *Journal of Graph Theory*, 40(2):91–103, 2002.

[49] N. Polat. Netlike partial cubes I. General properties. *Discrete Mathematics*, 307(22):2704–2722, 2007.

[50] D. Ž. Djoković. Distance-preserving subgraphs of hypercubes. *Journal of Combinatorial Theory, Series B*, 14(3):263–267, 1973.

[51] M. M. Deza and M. Laurent. *Geometry of cuts and metrics*, volume 15. Springer, 2009.

[52] S. V. Shpectorov. On scale embeddings of graphs into hypercubes. *Eur. J. Comb.*, 14(2):117–130, 1993.

[53] W. Johnson and J. Lindenstrauss. Extensions of lipschitz mappings into a hilbert space. *Contemporary mathematics*, 26(189-206):1, 1984.

[54] S. Dasgupta and A. Gupta. An elementary proof of the Johnson-Lindenstrauss lemma. *International Computer Science Institute, Technical Report*, 22(1):1–5, 1999.

[55] F. Chung. Diameters of graphs: Old problems and new results. *Congressus Numerantium*, 60:295–317, 1987.

[56] S. Zhang. On the power of lower bound methods for one-way quantum communication complexity. In *Proc. 38th International Conference on Automata, Languages and Programming (ICALP'11)*, pages 49–60, 2011.

[57] H. Klauck. On quantum and probabilistic communication: Las Vegas and one-way protocols. In *Proc. 32nd Annual ACM Symp. Theory of Computing*, pages 644–651. ACM, 2000.

[58] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995.

## Appendix A: $\ell_1$-Graphs Characterization

In this appendix we shall prove Theorem 6. Remember that $G(a|b) := \{x \in V(G) |\ d_G(x,a) < d_G(x,b)\}$.

**Theorem 6.** *A graph $G$ is an $\ell_1$-graph if and only if $G(a|b)$ is convex for each edge $(a,b)$ of $G$.*

This theorem is a generalization of *Djoković's Characterization* [50, 51] for partial cubes by introducing the concept of scale $k$ embedding, which is linked to the concept of $\ell_1$-graphs. A partial cube is then just a special case of $\ell_1$-graphs.

While the idea of scale $k$ embedding and some of its properties related to partial cubes were already studied, we could not find a clear and direct characterization for $\ell_1$-graphs as it is stated in Theorem 6, similar to Djoković's. For example, in [52] it is proved that a graph is embeddable into a hypercube with an odd scale if and only if it is 1-embeddable into a hypercube, meaning that odd scale embeddings do not add anything new. This makes sense since an odd scale embedding cannot alter the bipartiteness requirement.

The proof of the theorem is sketched as follows. The direction $(i) \implies (ii)$ is a direct generalization of Djoković's proof (see [51]). On the other hand, the direction $(ii) \implies (i)$ does not follow Djoković's proof, but instead introduces the idea of a $k$-rescaling map which transforms a given connected and unweighted graph into a new graph by adding $k-1$ new vertices on each original edge. In this way, the original distances are rescaled by a factor of $k$. We show in Lemma 7 that if $k$ is even, then this new graph is bipartite. Also, we show in Lemma 9 that this map preserves the convexity of subgraphs. This means that, if the sets $G(a|b)$ are convex for each edge $(a,b)$, then the new rescaled graph will fulfill the requirements from Djoković's Characterization for $k$ even and is, therefore, a partial cube. Since the original vertices are a subset of the new ones, the original graph is an $\ell_1$-graph.

In all the following, let $G = (V, E)$ be a connected and unweighted graph. We start by proving $(i) \implies (ii)$.

**Lemma 6.** *If $G$ is an $\ell_1$-graph, then $G(a|b)$ is convex for each edge $(a,b)$ of $G$.*

*Proof.* Let $(a,b)$ be an edge of $G$, let $x, y \in G(a|b)$ and $z \in V(G)$ lying on a shortest path from $x$ to $y$. Consider a hypercube scale $k$ embedding $\sigma_k : V \to Q_n$ in which node $a$ is labeled by $\sigma_k(a) = 0^n$ (where $c^j = ccc\cdots c$ means $c$ repeated $j$ times), node $b$ is labeled by $\sigma_k(b) = 1^k 0^{n-k}$ and nodes $x, y, z$ are labeled by the strings $X, Y, Z$. Given a $n$-bit string $A$, we define its $i$-th bit as $A_i$.

We first prove that $v \in G(a|b)$ if and only if $[\sigma_k(v)]_i \neq 1$ for $i \in [1,k]$. Consider that $[\sigma_k(v)]_i \neq 1$ for $i \in [1,k]$. Therefore $d_{Q_n}(\sigma_k(v), \sigma_k(b)) = k + d_{Q_n}(\sigma_k(v), \sigma_k(a))$ and hence $v$ is closer to $a$ than $b$, i.e., $v \in G(a|b)$. Now consider that $v \in G(a|b)$. This means $d_{Q_n}(\sigma_k(v), \sigma_k(a)) = lk$ and $d_{Q_n}(\sigma_k(v), \sigma_k(b)) = (l+1)k$ for some $l \in \mathbb{N}$. Suppose that $[\sigma_k(v)]_i = 1$ for $m$ indices $i$ in $[1,k]$. Therefore $d_{Q_n}(\sigma_k(v), \sigma_k(b)) - k + m = d_{Q_n}(\sigma_k(v), \sigma_k(a)) - m$, which gives $(l+1)k - k + m = lk - m \implies m = 0$, i.e., $[\sigma_k(v)]_i \neq 1$ for $i \in [1,k]$.

Given this, then $X_i, Y_i \neq 1$ for $i \in [1,k]$, and $d_{Q_n}(X,Y) = d_{Q_n}(X,Z) + d_{Q_n}(Z,Y)$ since $d_G(x,y) = d_G(x,z) + d_G(z,y)$. This implies that $Z_i \neq 1$ for $i \in [1,k]$, i.e., $z \in G(a|b)$. This shows that the set $G(a|b)$ is convex. □

To prove $(ii) \implies (i)$, we first make the following definitions.

**Definition 4.** *Let $\mathbb{G}_k : G \to G^{(k)}$ be the $k$-rescaling map which adds $k-1$ new nodes on every edge $e \in E$. The resulting graph $G^{(k)} = (V^{(k)}, E^{(k)})$ is called the $k$-rescaled image of $G$. Also, $G^{(1)} = G$. It is straightforward that $|E^{(k)}| = k|E|$ and $|V^{(k)}| = |V| + (k-1)|E|$.*

**Definition 5.** *Let $v \in V$. We define $G \oplus (v, v')$ as the graph $G' = (V', E')$ obtained by connecting an extra node $v'$ to the node $v$, so that $V' = V \cup \{v'\}$ and $E' = E \cup (v, v')$. If $v' = v$, we define $G \oplus (v, v) = G$.*

**Definition 6.** *Let $\mathbb{G}_k : G \to G^{(k)}$. Given $(u,v) \in E$, we define the set $V^{(k)}(u,v) := \{w \in V^{(k)} |\ d_{G^{(k)}}(u,w) < k$ and $d_{G^{(k)}}(v,w) < k\}$.*

The set $V^{(k)}(u,v)$ is just the nodes added between the nodes $u, v \in V$. With the above definition, $V^{(k)} = V \cup \big(\bigcup_{e\in E} V^{(k)}(e)\big)$.

We now state the following auxiliary lemmas.

**Lemma 7.** *The $k$-rescaled image $G^{(k)}$ of $G$ is bipartite if $k$ is even.*

*Proof.* A graph $G$ is bipartite if and only if it does not contain an odd cycle. If $G$ does not have cycles, then neither does $G^{(k)}$, since the $k$-rescaling map $\mathbb{G}_k$ cannot create cycles. Therefore $G^{(k)}$ is bipartite. Now suppose $G$ has cycles. Given a cycle $S \subseteq V$, its $k$-rescaled image $S^{(k)} = \mathbb{G}_k(S)$ is such that $|S^{(k)}| = k|S|$. If $S$ is an even cycle, then so is $S^{(k)}$. If $S$ is an odd cycle, then $S^{(k)}$ is an even cycle if $k$ is even. Therefore $S^{(k)}$ cannot have odd cycles for $k$ even and hence is bipartite. □

**Lemma 8.** *Let $S \subseteq V$ and $v \in S$. Take a completely new vertex $v' \notin V$. Construct the new graph $G' = G \oplus (v,v')$ and consider the new subset $S' = S \cup \{v'\}$. If $S$ is convex, then so is $S'$.*

*Proof.* Let $x, y \in S'$ and $z \in V \cup \{v'\}$ be such that $d_{G'}(x,y) = d_{G'}(x,z) + d_{G'}(y,z)$. We have two cases: Either $x, y \in S$ or, without loss of generality, $x = v'$ and $y \in S$. If $x, y \in S$, then it is straightforward that $z \neq v'$, otherwise the node $v \in S$ would be traversed twice. Therefore $z \in V$ and since $S$ is convex, $z \in S \subset S'$ and $S'$ is convex. On the other hand, if $x = v'$ and $y \in S$, the fact that $v'$ is only connected to $v$ implies that $d_{G'}(v',y) = d_{G'}(v',z) + d_{G'}(y,z) \iff d_G(v,y) = d_G(v,z) + d_G(y,z)$, which, together with $S$ being convex, means that $z \in S \subset S'$. □

**Lemma 9.** *Let $G^{(k)} = (V^{(k)}, E^{(k)})$ be the $k$-rescaled image of $G$. Then $S \subseteq V$ is convex if and only if $S^{(k)} = \mathbb{G}_k(S) \subseteq V^{(k)}$ is convex.*

*Proof.* We start by proving $S$ convex $\implies S^{(k)}$ convex. Given $U \subseteq V$, we write $U^{(k)} = U \cup U'$, where $U' = \bigcup_{e\in E} U^{(k)}(e)$ is the set of added nodes. Let $x, y \in S^{(k)}$ and $z \in V^{(k)}$ be such that $d_{G^{(k)}}(x,y) = d_{G^{(k)}}(x,z) + d_{G^{(k)}}(z,y)$. We will show that $z \in S^{(k)}$. Let us define the sets $A = \{a \in V |\ d_{G^{(k)}}(x,y) = d_{G^{(k)}}(x,a) + d_{G^{(k)}}(a,y)\}$ and $A' = \{a \in V' |\ d_{G^{(k)}}(x,y) = d_{G^{(k)}}(x,a) + d_{G^{(k)}}(a,y)\}$, i.e., $A$ is the set of original nodes that lie in the shortest path between $x$ and $y$, while $A'$ is the set of added nodes that lie in the shortest path between $x$ and $y$. Note that $z \in A \cup A'$. Suppose $A = \emptyset$. This means that $x, y \in S^{(k)}(e)$ for some edge $e \in E$. Therefore we must have $z \in S^{(k)}(e) \subseteq S^{(k)}$.

Now suppose $A \neq \emptyset$. Let $a(x), a(y) \in A$ be the closest nodes to $x$ and $y$, respectively. We must have $a(x) \in S$ (and $a(y) \in S$) since either $x \in S$ and then $a(x) = x$, or $x \in S^{(k)}(e)$ for some edge $e$, and then $a(x)$ is an endnode of $e$. We can have two situations: either $a(x) = a(y)$ or $a(x) \neq a(y)$.

Suppose $a(x) = a(y)$. Since $x \neq y$, this means that $x \in S^{(k)}(a(x), v_1)$ and $y \in S^{(k)}(a(x), v_2)$, for some $v_1, v_2 \in V$ such that $v_1 \neq v_2$, i.e., they are added nodes to two different edges with the common node $a(x)$. Therefore either $z \in S^{(k)}(a(x), v_1)$ or $z \in S^{(k)}(a(x), v_2)$ or $z = a(x)$, which lead to $z \in S^{(k)}$.

Suppose then that $a(x) \neq a(y)$. Consider for now that $z \in A$. Since $S$ is convex and $d_{G^{(k)}}(a(x), a(y)) = d_{G^{(k)}}(a(x), z) + d_{G^{(k)}}(a(y), z) \iff d_G(a(x), a(y)) = d_G(a(x), z) + d_G(a(y), z)$, we conclude that $z \in S$, i.e. $A \subseteq S$. Now consider that $z \in A'$, so $z \in V^{(k)}(v_1, v_2)$ for some nodes $v_1, v_2 \in V$. We must have $v_1, v_2 \in S$. Indeed, if $x \in V^{(k)}(v_1, v_2)$ (or $y$), by construction $x \in S^{(k)}$ and $x \in V^{(k)}(v_1, v_2) \implies v_1, v_2 \in S$. And if $x, y \notin V^{(k)}(v_1, v_2)$, it implies that $v_1, v_2 \in A \subseteq S$. Hence $z \in S^{(k)}(v_1, v_2) \subseteq S^{(k)}$. Thus $z \in S^{(k)}$ in all cases, so $S^{(k)}$ is convex.

We now prove the other direction, $S^{(k)}$ convex $\implies$ $S$ convex. Let $x, y \in S$ and $z \in V$ be such that $d_G(x,y) = d_G(x,z) + d_G(z,y)$. Suppose $z \notin S$. Remembering the equivalence between $d_{G^{(k)}}$ and $d_G$, this implies that $\exists z \in V \subset V^{(k)}$ but $z \notin S^{(k)}$ such that $d_{G^{(k)}}(x,y) = d_{G^{(k)}}(x,z) + d_{G^{(k)}}(z,y)$ for $x, y \in S \subset S^{(k)}$, which is a contradiction since $S^{(k)}$ is convex. We conclude that $z \in S$ and $S$ is convex. □

The above lemmas lead to the following one.

**Lemma 10.** *Let $G^{(k)} = (V^{(k)}, E^{(k)})$ be the $k$-rescaled image of $G$. If $G(a|b)$ is convex for each $(a,b) \in E$, then $G^{(k)}(a'|b')$ is also convex for each $(a',b') \in E^{(k)}$.*

*Proof.* Consider the edge $(a',b') \in E^{(k)}$ such that $a', b' \in \{a,b\} \cup V^{(k)}(a,b)$ for $a, b \in V$, i.e., $(a,b) \in E$ is the original edge. We note that the subgraph $G^{(k)}(a'|b')$ is just $\mathbb{G}_k\big(G(a|b)\big) \oplus (a, w_1) \oplus (w_1, w_2) \oplus \cdots \oplus (w_j, a')$ for $j = d_{G^{(k)}}(a,a') - 1$ (if $d_{G^{(k)}}(a,a') = 1$, then $\mathbb{G}_k\big(G(a|b)\big) \oplus (a,a')$, and if $a = a'$, then just $\mathbb{G}_k\big(G(a|b)\big)$). Since $G(a|b)$ is convex, by Lemma 9 $\mathbb{G}_k\big(G(a|b)\big)$ is also convex, and by Lemma 8 $G^{(k)}(a'|b')$ is convex. □

Finally, with the above lemmas, we can prove $(ii) \implies (i)$ in Theorem 6.

**Lemma 11.** *If $G(a|b)$ is convex for each edge $(a,b)$ of $G$, then $G$ is an $\ell_1$-graph.*

*Proof.* Consider the $k$-rescaled graph $G^{(k)} = (V^{(k)}, E^{(k)})$ corresponding to $G$ for $k$ even. By Lemma 7, $G^{(k)}$ is bipartite. By Lemma 10, $G^{(k)}(a'|b')$ is convex for each $(a',b') \in E^{(k)}$. Therefore, by Djokovic's characterization $G^{(k)}$ can be isometrically embedded into a hypercube [50]. Since $V \subset V^{(k)}$, we conclude that $G$ can be $k$-embedded into the same hypercube, i.e., it is an $\ell_1$-graph. □