Algorithmica manuscript No.
(will be inserted by the editor)

Extended Learning Graphs for Triangle Finding

Titouan Carette · Mathieu Laurière · Frédéric Magniez

Received: date / Accepted: date

Abstract We present new quantum algorithms for Triangle Finding improving its best previously known quantum query complexities for both dense and sparse instances. For dense graphs on n vertices, we get a query complexity of $O(n^{5/4})$ without any of the extra logarithmic factors present in the previous algorithm of Le Gall [FOCS'14]. For sparse graphs with $m \geq n^{5/4}$ edges, we get a query complexity of $O(n^{11/12}m^{1/6}\sqrt{\log n})$, which is better than the one obtained by Le Gall and Nakajima [ISAAC'15] when $m \geq n^{3/2}$. We also obtain an algorithm with query complexity $O(n^{5/6}(m\log n)^{1/6} + d_2\sqrt{n})$ where d_2 is the quadratic mean of the degree distribution.

Our algorithms are designed and analyzed in a new model of learning graphs that we call extended learning graphs. In addition, we present a framework in order to easily combine and analyze them. As a consequence we get much simpler algorithms and analyses than previous algorithms of Le Gall *et al* based on the MNRS quantum walk framework [SICOMP'11].

Keywords Quantum query complexity · Quantum walk · Triangle finding · Learning graph

1 Introduction

Decision trees form a simple model for computing Boolean functions by successively reading the input bits until the value of the function can be determined. In this model, the *query complexity* is the number of input bits queried. This allows us to study the complexity of a function in terms of its structural properties. For instance, sorting an array of size n can be done using $O(n\log n)$ comparisons, and this is optimal for comparison-only algorithms.

In an extension of the deterministic model, one can also allow randomized and even quantum computations. Then the speed-up can be exponential for partial functions (*i.e.* problems with promise) when we compare deterministic with randomized computation, and randomized with quantum computation. The case of total functions is rather fascinating. For them, the best possible gap can only be polynomial between each models [27,5], which is still useful in practice for many problems. But surprisingly, the best possible gap is still an open question, even if it was improved for both models very recently [4,1]. In the context of quantum computing, query complexity captures the great algorithmic successes of quantum computing like the search algorithm of Grover [17] and the period finding subroutine of Shor's factoring algorithm [29], while at the same time it is simple enough that one can often show tight lower bounds.

This work has been partially supported by the European Commission project Quantum Algorithms (QALGO) and the French ANR Blanc project RDAM.

T. Carette
Ecole Normale Supérieure de Lyon, 69007 Lyon, France
E-mail: titouan.carette@ens-lyon.fr

M. Laurière
ORFE, Princeton University, Princeton, NJ 08540, USA
E-mail: lauriere@princeton.edu

F. Magniez
Université de Paris, IRIF, CNRS, F-75013 Paris, France
E-mail: magniez@irif.fr

Reichardt [28] showed that the general adversary bound, formerly just a lower bound technique for quantum query complexity [18], is also an upper bound. This characterization has opened an avenue for designing quantum query algorithms. However, even for simple functions it is challenging to find an optimal bound. Historically, studying the query complexity of specific functions led to amazing progresses in our understanding of quantum computation, by providing new algorithmic concepts and tools for analyzing them. Some of the most famous problems in that quest are Element Distinctness and Triangle Finding [12]. Element Distinctness consists in deciding if a function takes twice the same value on a domain of size n, whereas Triangle Finding consists in determining if an n-vertex graph has a triangle. Quantum walks were used to design algorithms with optimal query complexity for Element Distinctness. Later on, a general framework for designing quantum walk based algorithms was developed with various applications [25], including for Triangle Finding [26].

For seven years, no progress on Triangle Finding was done until Belovs developed his beautiful model of *learning graphs* [6]. Learning graphs can be viewed as the dual form of the general adversary bound with an additional structure imposed on the form of the solution. This additional structure makes learning graphs easier to reason about without any background on quantum computing. On the other hand, they may not always provide optimal algorithms. Learning graphs have an intuitive interpretation in terms of electrical networks [8]. Their complexity is directly connected to the total conductance of the underlying network and its effective resistance. Moreover this characterization leads to a generic quantum implementation which is time efficient and preserves query complexity.

Among other applications, learning graphs have been used to design an algorithm for Triangle Finding with query complexity $O(n^{35/27})$ [7], improving on the previously known bound $\tilde{O}(n^{1.3})$ obtained by a quantum walk based algorithm [26]. Then the former was improved by another learning graph using $O(n^{9/7})$ queries [23]. This learning graph has been proven optimal for the original class of learning graphs [10], known as *non-adaptive learning graphs*, for which the conductance of each edge is constant. Then, Le Gall showed that quantum walk based algorithms are indeed stronger than non-adaptive learning graphs for Triangle Finding by constructing a new quantum algorithm with query complexity $\tilde{O}(n^{5/4})$ [15]. His algorithm combines in a novel way combinatorial arguments on graphs with quantum walks. One of the key ingredient is the use of an algorithm due to Ambainis for implementing Grover Search in a model whose queries may have variable complexities [2]. Le Gall used this algorithm to average the complexity of different branches of its quantum walk in a quite involved way. In the specific case of sparse graphs, those ideas have also demonstrated their advantage for Triangle Finding on previously known algorithms [16].

The starting point of the present work is to investigate a deeper understanding of learning graphs and their extensions. Indeed, various variants have been considered without any unified and intuitive framework. For instance, the best known quantum algorithm for k-Element Distinctness (a variant of Element Distinctness where we are now checking if the function takes k times the same value) has been designed by several clever relaxations of the model of learning graphs [6]. Those relaxations led to algorithms more powerful than non-adaptive learning graphs, but at the price of a more complex and less intuitive analysis. In **Section 3**, we extract several of those concepts that we formalize in our new model of *extended learning graphs* (**Definition 3**). We prove that their complexity (**Definition 4**) is always an upper bound on the query complexity of the best quantum algorithm solving the same problem (**Theorem 2**). We also introduce the useful notion of *super edge* (**Definition 5**) for compressing some given portion of a learning graph. We use them to encode efficient learning graphs querying a part of the input on some given index set (**Lemmas 3 and 4**). In some sense, we transpose to the learning graph setting the strategy of finding all 1-bits of some given sparse input using Grover Search.

In **Section 4**, we provide several tools for composing our learning graphs. We should first remind the reader that, since extended learning graphs cover a restricted class of quantum algorithms, it is not possible to translate all quantum algorithms in that model. Nonetheless we succeed for two important algorithmic techniques: Grover Search with variable query complexities [2] (**Lemma 5**), and Johnson Walk based quantum algorithms [26,25] (**Theorem 3**). In the last case, we show how to incorporate the use of super edges for querying sparse inputs.

We validate the power and the ease of use of our framework on Triangle Finding in **Section 5**. First, denoting n is the number of vertices, we provide a simple adaptive learning graph with query complexity $O(n^{5/4})$, whose analysis is arguably much simpler than the algorithm of Le Gall, and whose complexity is cleared of logarithmic factors (**Theorem 4**). This also provides the first natural separation between non-adaptive and adaptive learning graphs. Then, we focus on sparse input graphs and develop extended learning graphs. All algorithms of [16] could be rephrased in our model. But more importantly, we show that one can design more efficient ones. For sparse graphs with $m \geq n^{5/4}$ edges, we get a learning graph with query complexity $O(n^{11/12}m^{1/6}\sqrt{\log n})$, which improves the results of [16] when $m \geq n^{3/2}$ (**Theorem 5**). We also construct another learning graph with query complexity $O(n^{5/6}(m\log n)^{1/6} + d_2\sqrt{n})$, where d_2 is the quadratic mean of the degree distribution (**Theorem 6**). To the best

of our knowledge, this is the first quantum algorithm for Triangle Finding whose complexity depends on this parameter d_2. The complexities obtained in Theorem 5 and Theorem 6 are displayed in Figure 1.

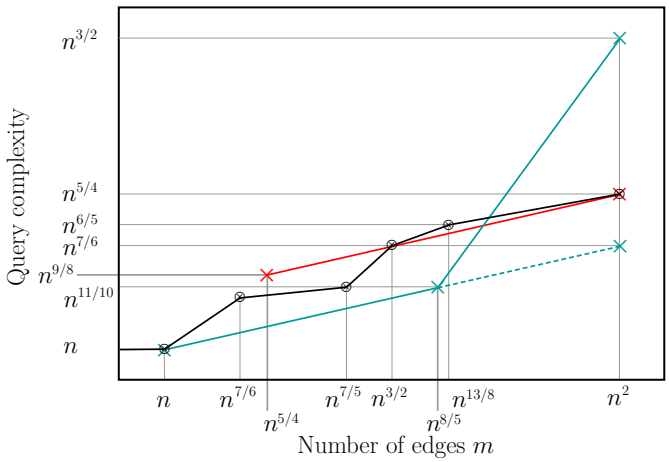


Fig. 1 Upper bounds on the query complexity as a function of the number of edges. Black: complexity given by Le Gall and Nakajima [16]. Red: $m \mapsto n^{11/12}m^{1/6}$, which corresponds (up to log factors) to the complexity given by Theorem 5. Green: $m \mapsto n^{5/6}m^{1/6} + m/\sqrt{n}$, which corresponds (up to log factors) to the complexity given by Theorem 6 under the assumption that d_2 is of order m/n (which is the case if the graph has uniform degree for instance). Dashed line: $m \mapsto n^{5/6}m^{1/6}$, which represents what one would obtain if we could get rid of the term involving d_2 in the query complexity given by Theorem 6.

2 Preliminaries

We will deal with Boolean functions of the form $f : Z \to \{0,1\}$, where $Z \subseteq \{0,1\}^N$. In the query model, given a function $f : Z \to \{0,1\}$, the goal is to evaluate $f(z)$ by making as few queries to the z as possible. A query is a question of the form 'What is the value of z in position $i \in [N]$?', to which is returned $z_i \in \{0,1\}$.

In this paper we will discuss functions taking as input the adjacency matrix of a graph. Then z will encode an undirected graph G on vertex set $[n]$, and $N = \binom{n}{2}$ in order to encode the possible edges of G. If $S \subseteq [N]$ is a subset of (indices of) edges, we encode into a partial assignment the corresponding assigned location, that is, $z_S = \{(i, z_i) : i \in S\}$. For $k_1, k_2 \in [n]$, it will be convenient to write $z_{k_1k_2}$ to denote the value of z at the index corresponding to the possible edge between vertices k_1 and k_2. So, with this notation, $z_{k_1k_2} = 1$ if and only if there is an edge between k_1 and k_2 in G.

In the quantum query model, these queries can be asked in superposition. We refer the reader to e.g. [20,3, 30,13] for precise definitions and background on the quantum query model. We denote by $Q(f)$ the number of queries needed by a quantum algorithm to evaluate f with error at most $1/3$. Surprisingly, the general adversary bound, that we define below, is a tight characterization of $Q(f)$.

Definition 1 Let $f : Z \to \{0,1\}$ be a function, with $Z \subseteq \{0,1\}^N$. The *general adversary bound* $\mathrm{Adv}^{\pm}(f)$ is defined as the optimal value of the following optimization problem:

$$\text{minimize: } \max_{z \in Z} \sum_{j \in [n]} X_j[z,z] \quad \text{subject to: } \sum_{j \in [n] : x_j \neq y_j} X_j[x,y] = 1, \text{ when } f(x) \neq f(y),$$
$$X_j \succeq 0, \ \forall j \in [N],$$

where the optimization is over positive semi-definite matrices X_j with rows and columns labeled by the elements of Z, and $X_j[x,y]$ is used to denote the (x,y)-entry of X_j.

Theorem 1 ([18,24,28]) $Q(f) = \Theta(\mathrm{Adv}^{\pm}(f))$.

3 Extended learning graphs

Consider some fixed Boolean function $f : Z \to \{0,1\}$, where $Z \subseteq \{0,1\}^N$. The set of positive inputs (or instances) will be usually denoted by $Y = f^{-1}(1)$. A 1-*certificate* for f on $y \in Y$ is a subset $I \subseteq [N]$ of indices such that $f(z) = 1$ for every $z \in Z$ with $z_I = y_I$.

3.1 Model and complexity

Intuitively, learning graphs are simply electric networks of a special type, see e.g. [8]: the network is embedded in a rooted directed acyclic graph, which has a few similarities with decision trees as we explain next. Indeed, in the learning graph model, vertices are labelled by subsets $S \subseteq [N]$ of indices (input positions) and edges are basically from any vertex labelled by, say, S to any other one labelled $S \cup \{j\}$, for some $j \notin S$. Such an edge can be interpreted as querying the input bit x_j, while x_S has been previously learnt. The weight on the edge is its conductance: the larger it is, the more flow will go through it. Sinks of the graph are labelled by potential 1-certificates of the function we wish to compute.

Thus a random walk on that network starting from the root (labelled by $\emptyset$), with probability transitions proportional to conductances, will hit a 1-certificate with average time proportional to the product of the total conductance by the effective resistance (the minimal possible energy of a flow) between the root of leaves having 1-certificates [8].

If weights are independent of the input, then the learning graph is called *non-adaptative*. When they depend on previously learned bits, it is called *adaptative*. In this case, the weight of an edge, say $(S, S \cup \{j\})$, is a function of z_S, where z is the input. Formally, adaptive learning graphs can be defined as follows.

Definition 2 (Adaptive learning graph) Let $Y \subseteq Z$ be finite sets. An *adaptive learning graph* $\mathscr{G}$ is a 5-tuple $(\mathscr{V}, \mathscr{E}, \mathscr{S}, \{w_z : z \in Z\}, \{p_y : y \in Y\})$ satisfying

- $(\mathscr{V}, \mathscr{E})$ is a directed acyclic graph rooted in some vertex $r \in \mathscr{V}$;
- $\mathscr{S}$ is a vertex labelling mapping each $v \in \mathscr{V}$ to $\mathscr{S}(v) \subseteq [N]$ such that $\mathscr{S}(r) = \emptyset$ and $\mathscr{S}(v) = \mathscr{S}(u) \cup \{j\}$ for every $(u,v) \in \mathscr{E}$ and some $j \notin \mathscr{S}(u)$;
- Values $w_z(u,v)$ are in $\mathbb{R}_{\geq 0}$ and depend on z only through $z_{\mathscr{S}(v)}$, for every $(u,v) \in \mathscr{E}$;
- $p_y : \mathscr{E} \to \mathbb{R}_{\geq 0}$ is a unit flow whose source is the root and such that $p_y(e) = 0$ when $w_y(e) = 0$, for every $y \in Y$.

We say that $\mathscr{G}$ is an *adaptive learning graph for* some function $f : Z \to \{0,1\}$, when $Y = f^{-1}(1)$ and each sink of p_y contains a 1-certificate for f on y, for all positive input $y \in f^{-1}(1)$.

When there is no ambiguity, we usually define $\mathscr{S}$ by stating the *label* of each vertex. We also say that an edge $e = (u,v)$ *loads* j when $\mathscr{S}(v) = \mathscr{S}(u) \cup \{j\}$.

In the extended model of learning graphs that we formalize below, the weights can also depend on both the value of the next queried bit and the value of the function itself through the XOR of these two bits. We call them *extended learning graphs*.

Formally, we generalize the original model of learning graphs by allowing two possible weights on each edge: one for positive instances and one for negative ones. Those weights are linked together as explained in the following definition.

Definition 3 (Extended learning graph) Let $Y \subseteq Z$ be finite sets. An *extended learning graph* $\mathscr{G}$ is a 5-tuple $(\mathscr{V}, \mathscr{E}, \mathscr{S}, \{w_z^b : z \in Z, b \in \{0,1\}\}, \{p_y : y \in Y\})$ satisfying

(i) $(\mathscr{V}, \mathscr{E})$ is a directed acyclic graph rooted in some vertex $r \in \mathscr{V}$;

(ii) $\mathscr{S}$ is a vertex labelling mapping each $v \in \mathscr{V}$ to $\mathscr{S}(v) \subseteq [N]$ such that $\mathscr{S}(r) = \emptyset$ and $\mathscr{S}(v) = \mathscr{S}(u) \cup \{j\}$ for every $(u,v) \in \mathscr{E}$ and some $j \notin \mathscr{S}(u)$;

(iii) Values $w_z^b(u,v)$ are in $\mathbb{R}_{\geq 0}$ and depend on z only through $z_{\mathscr{S}(v)}$, for every $(u,v) \in \mathscr{E}$;

(iv) $w_x^0(u,v) = w_y^1(u,v)$ for all $x \in Z \setminus Y, y \in Y$ and edges $(u,v) \in \mathscr{E}$ such that $x_{\mathscr{S}(u)} = y_{\mathscr{S}(u)}$ and $x_j \neq y_j$ with $\mathscr{S}(v) = \mathscr{S}(u) \cup \{j\}$.

(v) $p_y : \mathscr{E} \to \mathbb{R}_{\geq 0}$ is a unit flow whose source is the root and such that $p_y(e) = 0$ when $w_y^1(e) = 0$, for every $y \in Y$.

We say that $\mathscr{G}$ is a *learning graph for* some function $f : Z \to \{0,1\}$, when $Y = f^{-1}(1)$ and each sink of p_y contains a 1-certificate for f on y, for all positive input $y \in f^{-1}(1)$.

The combination of **(iii)** and **(iv)** yields that $w_z^b(u,v)$, with $\mathscr{S}(v) = \mathscr{S}(u) \cup \{j\}$, depends on z and b only through $z_{\mathscr{S}(u)}$ and the XOR of b and z_j. Notice that if $w_z^0 = w_z^1$ for all $z \in Z$, then this definition reduces to an adaptive learning graph (see Definition 2). In the sequel, unless otherwise specified, by *learning graph* we mean *extended learning graph*.

The complexity of extended learning graphs is inspired by the notion of complexity for learning graphs or adaptive learning graphs. It relies on the choice of the appropriate weight function for each complexity term.

Definition 4 (Extended learning graph complexity) Let $\mathscr{G}$ be an extended learning graph for a function $f : Z \to \{0,1\}$. Let $x \in Z \setminus Y$, $y \in Y$, and $\mathscr{F} \subseteq \mathscr{E}$. The *negative complexity* of $\mathscr{F}$ on x and the *positive complexity* of $\mathscr{F}$ on y (with respect to $\mathscr{G}$) are respectively defined by

$$C^0_{\mathscr{G}}(\mathscr{F},x) = \sum_{e\in\mathscr{F}} w^0_x(e) \quad \text{and} \quad C^1_{\mathscr{G}}(\mathscr{F},y) = \sum_{e\in\mathscr{F}} \frac{p_y(e)^2}{w^1_y(e)}. \tag{1}$$

Then the *negative and positive complexities* of $\mathscr{F}$ are $C^0_{\mathscr{G}}(\mathscr{F}) = \max_{x\in f^{-1}(0)} C^0_{\mathscr{G}}(\mathscr{F},x)$ and $C^1_{\mathscr{G}}(\mathscr{F}) = \max_{y\in f^{-1}(1)} C^1_{\mathscr{G}}(\mathscr{F},y)$. The *complexity* of $\mathscr{F}$ is $C_{\mathscr{G}}(\mathscr{F}) = \sqrt{C^0(\mathscr{F})C^1(\mathscr{F})}$ and the *complexity* of $\mathscr{G}$ is $C(\mathscr{G}) = C_{\mathscr{G}}(\mathscr{E})$. When the underlying learning graph $\mathscr{G}$ is clear from the context, we will not write explicitly the subscript and use the notations $C^0(\mathscr{F},x)$, $C^1(\mathscr{F},x)$, $C^0(\mathscr{F})$, $C^1(\mathscr{F})$, and $C(\mathscr{F})$. Last, the *extended learning graph complexity* of f, denoted $\mathscr{LG}^{\text{ext}}(f)$, is the minimum complexity of an extended learning graph for f.

Most often we will split a learning graph into *stages* $\mathscr{F}$, that is, when the flow through $\mathscr{F}$ has the same total amount 1 for every positive inputs. This allows us to analyze the learning graph separately on each stage.

As for adaptive learning graphs [7,9], the extended learning graph complexity upper bounds the standard query complexity.

Theorem 2 *For every function $f : Z \to \{0,1\}$, we have $Q(f) = O(\mathscr{LG}^{\text{ext}}(f))$.*

Proof We assume that f is not constant, otherwise the result holds readily. The proof follows the lines of the analysis of the learning graph for Graph collision in [6]. We already know that $Q(f) = O(\text{Adv}^{\pm}(f))$ by Theorem 1. Fix any extended learning graph $\mathscr{G}$ for f. Observe from Definition 1 that $\text{Adv}^{\pm}(f)$ is defined by a minimization problem. Therefore finding any feasible solution with objective value $C(\mathscr{G},f)$ would conclude the proof. Without loss of generality, assume that $C^0(\mathscr{G}) = C^1(\mathscr{G})$ (otherwise we can multiply all weights by $\sqrt{C^1(\mathscr{G})/C^0(\mathscr{G})}$). Then both complexities become $\sqrt{C^0(\mathscr{G})C^1(\mathscr{G})}$ and the total complexity remains $C(\mathscr{G})$.

For each edge $e = (u,v) \in \mathscr{E}$ with $\mathscr{S}(v) = \mathscr{S}(u) \cup \{j\}$, define a block-diagonal matrix $X^e_j = \sum_\alpha (Y^e_j)_\alpha$, where the sum is over all possible assignments α on $\mathscr{S}(u)$. Each $(Y^e_j)_\alpha$ is defined as $(\psi_0\psi_0^* + \psi_1\psi_1^*)$, where for each $z \in \{0,1\}^n$ and $b \in \{0,1\}$

$$\psi_b[z] = \begin{cases} p_e(z)/\sqrt{w^1_z(e)} & \text{if } z_{\mathscr{S}(u)} = \alpha,\ f(z) = 1 \text{ and } z_j = 1-b, \\ \sqrt{w^0_z(e)} & \text{if } z_{\mathscr{S}(u)} = \alpha,\ f(z) = 0 \text{ and } z_j = b, \\ 0 & \text{otherwise.} \end{cases}$$

Define now $X_j = \sum_e X^e_j$ where the sum is over all edges e loading j. Fix any $x \in f^{-1}(0)$ and $y \in f^{-1}(1)$. Then we have $X^e_j[x,x] = w^0_x(e)$ and $X^e_j[y,y] = (p_e(y))^2/w^1_y(e)$. So the objective value is

$$\begin{aligned} \max_{z\in\{0,1\}^n} \sum_{j\in[n]} X_j[z,z] &= \max\left\{ \max_{x\in f^{-1}(0)} \sum_j X_j[x,x],\ \max_{y\in f^{-1}(1)} \sum_j X_j[y,y] \right\} \\ &= \max\left\{C^0(\mathscr{G}), C^1(\mathscr{G})\right\} = C(\mathscr{G}). \end{aligned}$$

Consider the cut $\mathscr{F}$ over $\mathscr{G}$ of edges $(u,v) \in \mathscr{E}$ such that $\mathscr{S}(v) = \mathscr{S}(u) \cup \{j\}$ and $x_{\mathscr{S}(u)} = y_{\mathscr{S}(u)}$ but $x_j \neq y_j$. Then each edge $e \in \mathscr{F}$ loading j satisfies $w^0_x(e) = w^1_y(e)$ and therefore $X^e_j[x,y] = p_e(y)$. Thus, $\sum_{j:x_j\neq y_j} X_j[x,y] = \sum_{e\in\mathscr{F}} p_e(y) = 1$. Hence the constraints of Definition 1 are satisfied. □

3.2 Compression of learning graphs into super edges

We will simplify the presentation of our learning graphs by introducing a new type of edge encoding specific learning graphs as sub-procedures. Since an edge has a single 'exit', we can only encode learning graphs whose flows have unique sinks.

Definition 5 (Super edge) A *super edge* is an extended learning graph such that each possible flow has the same unique sink. If $\widetilde{\mathscr{G}}$ is a super edge, by analogy with edges, we will sometimes denote its *positive* and *negative edge-complexities* on input $x \in Z \setminus Y$ and $y \in Y$ by $c^0(\widetilde{\mathscr{G}},x) = C^0_{\widetilde{\mathscr{G}}}(\widetilde{\mathscr{G}},x)$ and $c^1(\widetilde{\mathscr{G}},y) = C^1_{\widetilde{\mathscr{G}}}(\widetilde{\mathscr{G}},y)$ respectively.

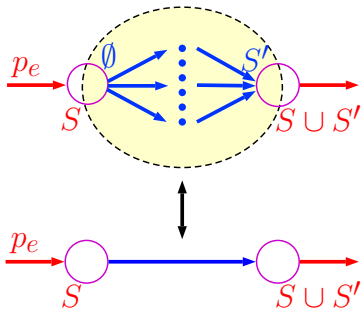


Fig. 2 Expansion and contraction of a super edge (in blue). The incoming flow is p_e. The set S has been loaded before the super edge, and the super edge loads S'.

In particular, an edge can be viewed as a super edge in the following way. Let $\mathscr{G} = (\mathscr{V}, \mathscr{E}, \mathscr{S}, \{w_z : z \in Z\}, \{p_y : y \in Y\})$ be a learning graph, and let $e = (u,v) \in \mathscr{E}$ be one of its edges. Let us define $\mathscr{G}_e = (\{u,v\}, \{e\}, \mathscr{S}, \{w_z : z \in Z\}, \{p_y \equiv 1 : y \in Y\})$ where the flow is constant (the labelling mapping and the weights are restricted respectively to $\{u,v\}$ and $\{e\}$). We say that $\mathscr{G}_e$ is the super edge associated to the edge e. To alleviate further notations, we will also use the notion of positive and negative edge-complexities for edges. So we will write $c^0_{\mathscr{G}}(e,x) = C^0(\mathscr{G}_e,x)$, $c^1_{\mathscr{G}}(e,y) = C^1(\mathscr{G}_e,y)$, $c^0_{\mathscr{G}}(e) = C^0(\mathscr{G}_e)$, and $c^1_{\mathscr{G}}(e) = C^1(\mathscr{G}_e)$. Observe that in particular, using the notation (1) with $\mathscr{F} = \{e\}$, $C^0_{\mathscr{G}}(\{e\},x) = c^0_{\mathscr{G}}(e,x)$ and $C^1_{\mathscr{G}}(\{e\},y) = p_y(e)^2 \times c^1_{\mathscr{G}}(e,y)$, where p_y is the flow in $\mathscr{G}$. If the underlying graph $\mathscr{G}$ is clear from the context, we shall drop the subscript $\mathscr{G}$. We now use this idea in order to define the complexity of learning graphs with super edges.

Indeed, one can now consider learning graphs with super edges. They are equivalent to learning graphs without super edges by doing recursively the following replacement for each super edge, say e: (1) replace it by its underlying learning graph, say $\mathscr{G}_e$, plugging the root to all incoming edges and the unique flow sink to all outgoing edges; (2) root the incoming flow according to the plugged learning graph. Let us call this learning graph the *expansion* of the original one with super edges. In Figure 2, we provide an graphical representation of the expansion and the contraction of a super-edge. Then, a direct inspection leads to the following result that we will use in order to compute complexities directly on our original learning graphs.

Lemma 1 *Let $\mathscr{G}$ be a learning graph with super edges for some function f. Then the expansion of $\mathscr{G}$ is also a learning graph for f. Moreover, let $\exp(\mathscr{F})$ be the expansion of $\mathscr{F} \subseteq \mathscr{E}$. Then $\exp(\mathscr{F})$ has positive and negative complexities*

$$C^0(\exp(\mathscr{F}),x) = \sum_{e \in \mathscr{F}} c^0(e,x) \quad \text{and} \quad C^1(\exp(\mathscr{F}),y) = \sum_{e \in \mathscr{F}} p_y(e)^2 \times c^1(e,y)$$

where the sums are over edges or super edges $e \in \mathscr{F}$.

Fix some stage $\mathscr{F} \subseteq \mathscr{E}$ of $\mathscr{G}$, that is such that the flow through $\mathscr{F}$ has the same total amount 1 for every positive input. We will use the following lemma, that we adapt from non-adaptive learning graphs, to assume that a learning graph has positive complexity at most 1 on $\mathscr{F}$.

Lemma 2 (Speciality [6]) *Let $\mathscr{G} = (\mathscr{V}, \mathscr{E}, \mathscr{S}, \{w_z : z \in Z\}, \{p_y : y \in Y\})$ be a learning graph for a function $f : Z \to \{0,1\}$. Let $\mathscr{F} \subseteq \mathscr{E}$ be a stage of $\mathscr{G}$ whose flow always uses the same ratio $1/T$ of transitions and is uniform on them. Then there is a learning graph $\widetilde{\mathscr{G}} = (\mathscr{V}, \mathscr{E}, \mathscr{S}, \{\tilde{w}_z : z \in Z\}, \{p_y : y \in Y\})$ with the same structure as $\mathscr{G}$ but weights $\tilde{w}$ which can be different for edges in $\mathscr{F}$ (that is, $\tilde{w}_z(e) = w_z(e)$ for all $z \in Z$ and all $e \notin \mathscr{F}$) such that for every $x \in f^{-1}(0)$ and $y \in f^{-1}(1)$*

$$C^0_{\widetilde{\mathscr{G}}}(\mathscr{F},x) \leq T \mathop{\mathbb{E}}_{e \in \mathscr{F}} \left[c^0_{\mathscr{G}}(e,x) c^1_{\mathscr{G}}(e)\right] \quad \text{and} \quad C^1_{\widetilde{\mathscr{G}}}(\mathscr{F},y) \leq 1.$$

The parameter T involved in the above result is usually called the *speciality* of $\mathscr{F}$.

Proof Let n_{total} be the number of transitions in $\mathscr{F}$ and n_{used} the number of them used by each flow (i.e. with positive flow). Therefore $T = n_{\text{total}}/n_{\text{used}}$. By assumption, the flow on each edge is either 0 or $1/n_{\text{used}}$. For each edge e in $\mathscr{F}$, let $\lambda_e = c^1(e)/n_{\text{used}}$, and for every edge $e \notin \mathscr{F}$, let $\lambda_e = 1$. For every input z, we let $\tilde{w}^b_z(e) = \lambda_e w^b_z(e)$. We name $\widetilde{\mathscr{G}}$ the associated learning graph with the new weights.

Then for any $x \in f^{-1}(0)$, $C^0_{\widetilde{\mathscr{G}}}(\mathscr{F},x) = \sum_{e \in \mathscr{F}} \lambda_e c^0_{\mathscr{G}}(e,x) = T\, \mathbb{E}_{e \in \mathscr{F}} \left[c^0_{\mathscr{G}}(e,x) c^1_{\mathscr{G}}(e)\right]$. Similarly, for any $y \in f^{-1}(1)$, $C^1_{\widetilde{\mathscr{G}}}(\mathscr{F},y) = \sum_{e \in \mathscr{F}} p_y(e)^2 c^1_{\mathscr{G}}(e,y)/\lambda_e \leq 1$, since each term in the sum is positive only for edges with positive flow. □

3.3 Loading sparse inputs

We study a particular type of super edge, that we will use repeatedly in the sequel. To construct a learning graph for a given function, one often needs to load a subset S of the labels. This can be done by a path of length $|S|$ with negative and positive complexities $|S|$, which, after some rebalancing, leads directly to the following lemma.

Lemma 3 *For any set S, there exists a super edge denoted $\mathsf{DenseLoad}_S$ loading S with the following complexities for any input $z \in \{0,1\}^N$:*

$$c^0(\mathsf{DenseLoad}_S, z) = |S|^2 \quad \textit{and} \quad c^1(\mathsf{DenseLoad}_S, z) = 1.$$

Proof Let us assume for simplicity that $N = |S|$ and denote $S = \{1, \ldots, N\}$. We define the learning graph $\mathsf{DenseLoad}_S$ as the sequence of edges $e_1 = (\emptyset, \{1\}), e_2 = (\{1\}, \{1,2\}), e_N = ([N-1], |S|)$. The weights are defined as:

$$w^0_{e_k} = N \quad \text{and} \quad w^1_{e_k} = N.$$

They satisfy the requirements of Definition 3. We obtain the values of the complexities by using their definition. □

When the input is sparse one can do significantly better as we describe now, where $|z_S|$ denotes the Hamming weight of z_S.

Lemma 4 *For any set S, there exists a super edge denoted $\mathsf{SparseLoad}_S$ loading S with the following complexities for any input $z \in \{0,1\}^N$:*

$$c^0(\mathsf{SparseLoad}_S, z) \leq 6|S|(|z_S|+1)\log(|S|+1) \quad \textit{and} \quad c^1(\mathsf{SparseLoad}_S, z) \leq 1.$$

Proof Let us assume for simplicity that $N = |S|$ and $S = \{1, \ldots, N\}$. We define the learning graph $\mathsf{SparseLoad}_S$ as the path through edges $e_1 = (\emptyset, \{1\})$, $e_2 = (\{1\}, \{1,2\})$, …, $e_N = (\{1, \ldots, N-1\}, S)$. The weights are defined as, for $b \in \{0,1\}$ and $z \in Z$,

$$w^b_{e_k}(z) = \begin{cases} 3 \cdot (|z_{[j-1]}|+1) \cdot \log(N+1) & \text{if } z_j = b, \\ 3N \cdot \log(N+1) & \text{if } z_j = 1-b, \end{cases}$$

When $|z| > 0$, let us denote $i_0 = 0$, $i_{|z|+1} = N+1$ and $(i_k)_{k=1,\ldots,|z|}$ the increasing sequence of indices j such that $z_j = 1$. Then, for $k = 1, \ldots, |z|+1$, we define m_k as the number of indices $j \in (i_{k-1}, i_k)$ such that $z_j = 0$. More precisely, $m_k = i_k - i_{k-1} - 1$ for $1 \leq k \leq |z|$ and $m_{|z|+1} = N - i_{|z|}$. So $\sum_{k=1}^{|z|+1} m_k = N - |z|$. Then, for any input z,

$$C^0(\mathsf{SparseLoad}_S, z) = \begin{cases} 3N \cdot \log(N+1) & \text{if } |z| = 0, \\ 3 \cdot \left(|z|N + \sum_{i=1}^{|z|+1} i \times m_i \right) \cdot \log(N+1) & \text{otherwise,} \end{cases}$$

which is bounded above by $6N \cdot (|z|+1) \cdot \log(N+1)$. Moreover, using $\sum_{i=1}^{|z|+1} \frac{1}{i} \leq \log(|z|+1)+1$, we get

$$C^1(\mathsf{SparseLoad}_S, z) = \frac{1}{3 \cdot \log(N+1)} \left((N-|z|)\frac{1}{N} + \sum_{i=1}^{|z|+1} \frac{1}{i} \right) \leq 1.$$

□

4 Composition of learning graphs

To simplify our presentation, we will use the term *empty transition* for an edge between two vertices representing the same set. They carry zero flow and weight, and they do not contribute to any complexity.

4.1 Learning graph for OR

Consider n Boolean functions $f_1,\dots,f_n$ with respective learning graphs $\mathscr{G}_1,\dots,\mathscr{G}_n$. The following lemma explains how to design a learning graph $\mathscr{G}_{\mathrm{OR}}$ for $f=\bigvee_{i\in[n]} f_i$ whose complexity is the squared mean of former ones. We will represent $\mathscr{G}_{\mathrm{OR}}$ graphically as

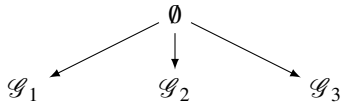
$$\emptyset \xrightarrow{\ i\ } \mathscr{G}_i$$

This result is similar to the one of [2], where a search procedure is designed for the case of variable query costs, or equivalently for a search problem divided into subproblems with variable complexities.

Lemma 5 *Let $\mathscr{G}_1,\dots,\mathscr{G}_n$ be learning graphs for Boolean functions $f_1,\dots,f_n$ over Z. Assume further that for every x such that $f(x)=1$, there are at least k functions f_i such that $f_i(x)=1$. Then there is a learning graph $\mathscr{G}$ for $f=\bigvee_{i\in[n]} f_i$ such that for every $z\in Z$*

$$\begin{cases} C^0(\mathscr{G},z) \le \dfrac{n}{k}\times \underset{i\in[n]}{\mathbb{E}}\left(C^0(\mathscr{G}_i,z)C^1(\mathscr{G}_i)\right) & \text{when } f(z)=0,\\ C^1(\mathscr{G},z)\le 1 & \text{when } f(z)=1. \end{cases}$$

Proof We define the new learning graph $\mathscr{G}$ by considering a new root $\emptyset$ that we link to the roots of each $\mathscr{G}_i$. In particular, each $\mathscr{G}_i$ lies in a different connected component. For $n=3$, the graph is displayed below:



Then, we rescale the original weights of edges in each component $\mathscr{G}_i$ by $\lambda_i = C^1(\mathscr{G}_i)/k$.

The complexity $C^0(\mathscr{G},x)$ for a negative instance x is

$$C^0(\mathscr{G},x)=\sum_{i=1}^{n}\lambda_i C^0(\mathscr{G}_i,x)=\frac{n}{k}\times\underset{i}{\mathbb{E}}\left(C^0(\mathscr{G}_i,x)C^1(\mathscr{G}_i)\right).$$

Consider now a positive instance y. Then y is also a positive instance for at least k functions f_i. Without loss of generality, assume further that these k functions are $f_1,f_2,\dots,f_k$. We define the flow of $\mathscr{G}$ (for y) as a flow uniformly directed from $\emptyset$ to $\mathscr{G}_i$ for $i=1,2,\dots,k$. In each component $\mathscr{G}_i$, the flow is then routed as in $\mathscr{G}_i$. Therefore we have

$$C^1(\mathscr{G},y)=\sum_{i=1}^{k}\frac{1}{k^2}\times\frac{C^1(\mathscr{G}_i,y)}{\lambda_i}\le 1.$$

Finally, observe that by construction the flow is directed to sinks having 1-certificates, thus $\mathscr{G}_{\mathrm{OR}}$ indeed computes $f=\bigvee_{i\in[n]} f_i$. □

4.2 Learning graph for Johnson walks

In [26,25], a new method was proposed for designing quantum search algorithms to find a "marked" element in the state space of a classical Markov chain. The complexity analysis of the quantum algorithms is based on the following three costs: the set-up cost for sampling a state; the update cost for simulating a transition along of the Markov chain; and the checking cost for checking if the current state is marked. Then, intuitively, a quantum algorithm can simulate the amplitude amplification algorithm [17,11,19] by first generating the quantum analogue of the stationary distribution of states according to the Markov chain, and then iterating a rotation made of the composition of a reflection through the marked states and of a reflection through the starting state. Creating the starting state and implementing the first reflection requires the two basic operations, namely set-up and checking. A spectral analysis of the Markov chain reveals that the last reflection can be approximated using Phase Estimation [21, 22,14] on the update operator.

Although this approach mostly requires a worst-case definition of the three corresponding costs, we can reprove a similar result using expected costs for the case of a Markov chain on Johnson graphs. In that case, states are k-subsets of $[n]$, and one transition consists in replacing one element of the current state by another one. Using the Learning Graph framework, the set-up corresponds to loading all elements but k of them, whereas the update

corresponds to loading the last k elements, which are marked on the flow. The checking part is represented by an extra learning graph.

To formalize these ideas, let us introduce some notations. Recall that we encode into a partial assignment the corresponding assigned location, that is, $z_S = \{(i, z_i) : i \in S\}$. Fix some parameters $r \leq k \leq n$. We would like to define a learning graph $\mathscr{G}_{\text{Johnson}}$ for $f = \bigvee_A f_A$, where A ranges over k-subsets of $[n]$ and f_A are Boolean functions over Z, but differently than in Lemma 5. For this, we are going to use a learning graph for f_A when the input has been already partially loaded, that is, loaded on $I(A)$ for some subset $I(A) \subseteq [N]$ depending on A only. Namely, we assume we are given, for every partial assignment λ, a learning graph $\mathscr{G}_{A,\lambda}$ defined over inputs $Z_\lambda = \{z \in Z : z(I(A)) = \lambda\}$ for f_A restricted to Z_λ.

Then, instead of the learning graph of Lemma 5, our learning graph $\mathscr{G}_{\text{Johnson}}$ factorizes the load of input z over $I(A)$ for $|A| = k$ and then uses $\mathscr{G}_{A,z_{I(A)}}$. This approach is more efficient when, for every positive instance y, there is a 1-certificate $I(T_y)$ for some r-subset T_y, and $A \mapsto I(A)$ is monotone. This is indeed the analogue of a walk on the Johnson Graph.

We will represent the resulting learning graph $\mathscr{G}_{\text{Johnson}}$ graphically using $r+1$ arrows: one for the first load of $(k-r)$ elements, and r smaller ones for each of the last r loads of a single element. For example, when $r = 2$ we draw:

$$\emptyset \xrightarrow{\quad A \quad} \cdot \rightarrow \cdot \rightarrow \mathscr{G}_{A,x_{I(A)}}$$

In the following, Load_S denotes any super edge loading the elements of S, such as DenseLoad or SparseLoad that we have defined in Lemmas 3 and 4.

Theorem 3 *For every subset $S \subseteq [N]$, let Load_S be any super edge loading S with $c^1(\mathsf{Load}_S) \leq 1$. Let $r \leq k \leq n$ and let $f = \bigvee_A f_A$, where A ranges over k-subsets of $[n]$ and f_A are Boolean functions over Z.*

Let I be a monotone mapping from subsets of $[n]$ to subsets of $[N]$ with the property that, for every $y \in f^{-1}(1)$, there is an r-subset $T_y \subseteq [n]$ whose image $I(T_y)$ is a 1-certificate for y.

Let $\mathbf{S}, \mathbf{U} > 0$ be such that every $x \in f^{-1}(0)$ satisfies

$$\mathop{\mathbb{E}}_{A' \subset [n] : |A'| = k-r} \left(C^0(\mathsf{Load}_{I(A')}, z)\right) \leq \mathbf{S}^2; \tag{2}$$

$$\mathop{\mathbb{E}}_{A' \subset A'' \subseteq [n] : |A'| = |A''| - 1 = i} \left(C^0(\mathsf{Load}_{I(A'') \setminus I(A')}, z)\right) \leq \mathbf{U}^2, \quad \text{for } k-r \leq i < k. \tag{3}$$

Let $\mathscr{G}_{A,\lambda}$ be learning graphs for functions f_A on Z restricted to inputs $Z_\lambda = \{z \in Z : z(I(A)) = \lambda\}$, for all k-subsets A of $[n]$ and all possible assignments λ over $I(A)$. Let finally $\mathbf{C} > 0$ be such that every $x \in f^{-1}(0)$ satisfies

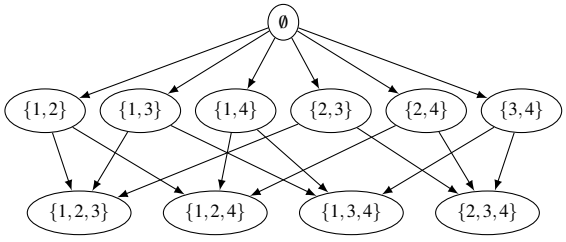
$$\mathop{\mathbb{E}}_{A \subseteq [n] : |A| = k} \left(C^0(\mathscr{G}_{A,x_{I(A)}}, x) C^1(\mathscr{G}_{A,x_{I(A)}}, f)\right) \leq \mathbf{C}^2. \tag{4}$$

Then there is a learning graph $\mathscr{G}_{\text{Johnson}}$ for f such that for every $z \in Z$

$$\begin{cases} C^0(\mathscr{G}_{\text{Johnson}}, z) = O\left(\mathbf{S}^2 + \left(\frac{n}{k}\right)^r \left(k \times \mathbf{U}^2 + \mathbf{C}^2\right)\right) & \text{when } f(z) = 0, \\ C^1(\mathscr{G}_{\text{Johnson}}, z) = 1 & \text{when } f(z) = 1. \end{cases}$$

Proof **Construction.** We define $\mathscr{G}_{\text{Johnson}}$ by emulating a walk on the Johnson graph $J(n,k)$ for searching a k-subset A having an r-subset T_y such that $I(T_y)$ is a 1-certificate for y. In that case, by monotonicity of I, the set $I(A)$ will be also a 1-certificate for y.

Our learning graph $\mathscr{G}_{\text{Johnson}}$ is composed of $(r+2)$ stages (that is, layers whose total incoming flow is 1), that we call Stage ℓ, for $\ell = 0, 1, \ldots, r+1$. An example of such a learning graph for $n = 4$, $k = 3$ and $r = 1$ is represented below:



Stage 0 of $\mathscr{G}_{\text{Johnson}}$ consists in $\binom{n}{(k-r)}$ disjoint paths, all of same weights, leading to vertices labelled by some $(k-r)$-subset A' and loading $I(A')$. They can be implemented by the super edges $\mathsf{Load}_{I(A')}$. For positive instances y, the flow goes from $\emptyset$ to subsets $I(A')$ such that $I(A') \cap T_y = \emptyset$.

For $\ell = 1, \ldots, r$, Stage ℓ consists in $(n-(k-r)-\ell+1)$ outgoing edges to each node labeled by a $(k-r+\ell-1)$-subset A'. Those edges are labelled by (A', j) where $j \notin A'$ and load $I(A' \cup \{j\}) \setminus I(A')$. They can be implemented by the super edges $\mathsf{Load}_{I(A' \cup \{j\}) \setminus I(A')}$. For positive instances y, for each vertex A' getting some positive flow, the flow goes out only to the edge (A', j_ℓ), with the convention $T_y = \{j_1, \ldots, j_r\}$.

The final Stage $(r+1)$ consists in plugging in nodes A the corresponding learning graph $\mathscr{G}_{A,x_{I(A)}}$, for each k-subset A. We take a similar approach than in the construction of $\mathscr{G}_{\text{OR}}$ above. The weights of the edges in each component $\mathscr{G}_{A,z_{I(A)}}$ are rescaled by a factor $\lambda_A = C^1(\mathscr{G}_{A,x_{I(A)}}) / \binom{n-r}{k-r}$. For a positive instance y, the flow is directed uniformly to each $\mathscr{G}_{A,y_{I(A)}}$ such that $T(y) \subseteq A$, and then according to $\mathscr{G}_{A,y_{I(A)}}$.

Observe that by construction, on positive inputs the flow reaches only 1-certificates of f. Therefore $\mathscr{G}_{\text{Johnson}}$ indeed computes f.

Analysis. Remember that the positive edge-complexity of our super edge Load is at most 1.

At Stage 0, the $\binom{n}{(k-r)}$ disjoint paths are all of same weights. The flow satisfies the hypotheses of Lemma 2 with a speciality of $O(1)$. Therefore, using inequality (2), the complexity of this stage is $O(\mathbf{S}^2)$ when $f(x) = 0$, and at most 1 otherwise.

For $\ell = 1, \ldots, r$, at Stage ℓ consists of $(n-(k-r)-\ell+1)$ outgoing edges to each node labeled by a $(k-r+\ell-1)$-subset. Take a positive instance y. Recall that, for each vertex A' getting some positive flow, the flow goes out only to the edge (A', j_ℓ). By induction on ℓ, the incoming flow is uniform when positive. Therefore, the flow on each edge with positive flow is also uniform, and the speciality of the stage is $O((\frac{n}{k})^\ell \cdot k)$. Hence, by Lemma 2 and using inequality 3, the cost of each such stage is $O((\frac{n}{k})^\ell \cdot k \cdot \mathbf{U}^2)$. The dominating term is thus $O\left((\frac{n}{k})^r \cdot k \cdot \mathbf{U}^2\right)$.

The analysis of the final stage (Stage $(r+1)$) is similar to the proof of Lemma 5. For a negative instance x, the complexity of this stage is:

$$\begin{aligned}\sum_A \lambda_A C^0(\mathscr{G}_{A,x_{I(A)}}, x) &= \frac{\binom{n}{k}}{\binom{n-r}{k-r}} \mathop{\mathbb{E}}_A \left(C^0(\mathscr{G}_{A,x_{I(A)}}, x) C^1(\mathscr{G}_{A,x_{I(A)}}) \right) \\ &= O\left(\left(\frac{n}{k}\right)^r \times \mathop{\mathbb{E}}_A \left(C^0(\mathscr{G}_{A,x_{I(A)}}, x) C^1(\mathscr{G}_{A,x_{I(A)}}) \right) \right).\end{aligned}$$

Similarly, when $f(y) = 1$, we get a complexity at most 1. □

5 Application to Triangle Finding

5.1 An adaptive Learning graph for dense case

We start by reviewing the main ideas of Le Gall's algorithm in order to find a triangle in an input graph G with n vertices. More precisely, we decompose the problem into similar subproblems, and we build up our adaptive learning graph on top of it. Doing so, we get rid of most of the technical difficulties that arise in the resolution of the underlying problems using quantum walk based algorithms.

Let V be the vertex set of G. For a vertex u, let N_u be the neighborhood of u, and for two vertices u, v, let $N_{u,v} = N_u \cap N_v$. Figure 3 and the algorithm described in Figure 4 illustrate the following strategy for finding a potential triangle in some given graph $G = (V, E)$. Here the sizes of the sets X, A, and B, namely x, a, and b respectively, are integers playing the role of parameters whose values will be determined later (see Theorem 4 below).

First, fix an x-subset X of vertices, that is, a subset of X of size x. Then, either G has a triangle with one vertex in X or each (potential) triangle vertex is outside X. The first case is quite easy to deal with, so we ignore it for now and we only focus on the second case. Thus there is no need to query any possible edge between two vertices u, v connected to the same vertex in X. Indeed, if such an edge exists, the first case will detect a triangle. Therefore one only needs to look for a triangle edge in $\Delta(X) = \{(u, v) \in V^2 : N_{u,v} \cap X = \emptyset\}$.

Second, search for an a-subset A with two triangle vertices in it. For this, construct the set $\Delta(X, A) = A^2 \cap \Delta(X)$ of potential triangle edges in A^2. The set $\Delta(X, A)$ can be easily set once all edges between X and A are known.

Third, in order to decide if $\Delta(X, A)$ has a triangle edge, search for a vertex w making a triangle with an edge of $\Delta(X, A)$.

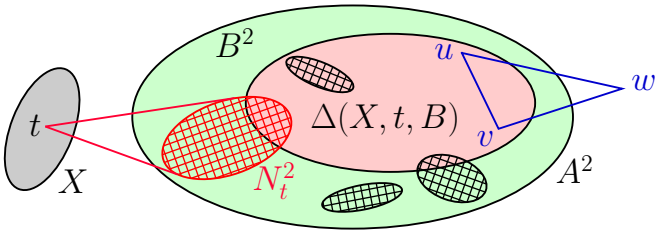


Fig. 3 Sets involved in Le Gall's algorithm.

Algorithm (simplified version of Le Gall's algorithm)

Search for $X \subseteq V$, $|X| = x$, such that (all possible subsets X of size x are valid or none)

(1) Either X has a triangle vertex, in which case
- **(a)** Search for $t \in X$
- **(b)** Search for an edge of G connected to t

(2) Or there is a triangle with no vertex in X, in which case
- **(a)** Perform a Johnson walk on $A \subseteq V$, $|A| = a$, with $I(A) = X \times A$ (see Theorem 3)
- **(b)** Search for t
- **(c)** Perform a Johnson walk on $B \subseteq V$, $|B| = b$, with $J(B) = \{t\} \times B$
- **(d)** Look for an edge in $\Delta(X,t,B)$, pairs of vertices in $J(B)$ not connected to the same vertex in X

Fig. 4 Simplified version of Le Gall's algorithm for finding a triangle in $G = (V,E)$.

Otherwise, search for a b-subset B of A such that w makes a triangle with two vertices of B. For this last step, we construct the set $\Delta(X,B,w) = (N_w)^2 \cap \Delta(X,B)$ of pairs of vertices connected to w. If any of such pair is an actual edge, then we have found a triangle.

We will use learning graphs of type $\mathscr{G}_{\text{OR}}$ for the first step, for finding an appropriate vertex w, and for deciding weither $\Delta(X,B,w)$ has an edge; and learning graphs of type $\mathscr{G}_{\text{Johnson}}$ for finding subsets A and B.

More formally now, let Triangle be the Boolean function such that $\text{Triangle}(G) = 1$ iff graph input G has a triangle. We do the following decomposition. First, observe that $\text{Triangle} = \bigvee_{X:|X|=x}(h_X \vee f_X)$ with $h_X(G) = 1$ (resp. $f_X(G) = 1$) iff G has a triangle with a vertex in X (resp. with no vertex in X). Then, we pursue the decomposition for $f_X(G)$ as $f_X(G) = \bigvee_{A:|A|=a} f_{X,A}(G)$ and $f_{X,A}(G) = \bigvee_{w \in V} f_{X,A,w}(G)$, for $A \subseteq V$ and $w \in V$, where

- $f_{X,A}(G) = 1$ iff G has a triangle between two vertices in $A \setminus X$ and a third one outside X;
- $f_{X,A,w}(G) = 1$ iff $w \notin X$ and G has a triangle between w and two vertices in $A \setminus X$.

Last, we can write $f_{X,A,w}(G) = \bigvee_{B \subset A,\, |B|=b} f_{X,B,w}(G)$.

With our notations introduced in Section 4, our adaptative learning graph $\mathscr{G}$ for Triangle Finding can be represented as in Figure 5.

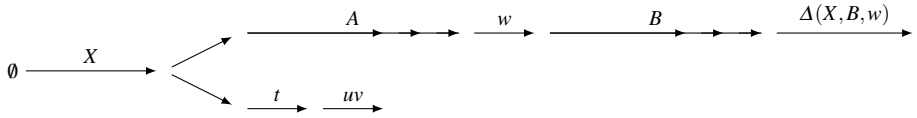


Fig. 5 Learning graph for Triangle Finding with complexity $O(n^{5/4})$.

Using adaptive learning graphs instead of the framework of quantum walk based algorithms from [25] simplifies the implementation of the above strategy because one can consider all the possible subsets X instead of choosing just a random one. Then one only needs to estimate the average complexity over all possible X. Such an average analysis was not considered in the framework of [25]. In addition, we do not need to estimate the size of $\Delta(X,A,w)$ at any moment of our algorithm. As a consequence, our framework greatly simplifies the combinatorial analysis of our algorithm as compared to the one of Le Gall, and lets us shave off some logarithmic factors.

Theorem 4 *The adaptive learning graph of Figure 5 with $|X| = x$, $|A| = a$, $|B| = b$, and using* $\mathsf{Load} = \mathsf{DenseLoad}$*, has complexity*

$$O\left(\sqrt{xn^2 + (ax)^2 + \left(\frac{n}{a}\right)^2\left(a\cdot x^2 + n\left(b^2 + \left(\frac{a}{b}\right)^2\left(b + \frac{b^2}{x}\right)\right)\right)}\right).$$

In particular, taking $a = n^{3/4}$ and $b = x = \sqrt{n}$ leads to $Q(\mathsf{Triangle}) = O(n^{5/4})$.

Proof From now on, fix some input graph $G = (V,E)$ (with or without a triangle). We compute the complexity $C(\mathscr{G}, G)$ of $\mathscr{G}$ on G using Lemma 5 and Theorem 3. From the decomposition of Triangle one can already check that the resulting learning graph computes the function Triangle.

Also all complexities for positive instances will be at most 1. Therefore, we only compute the complexity of negative instances, and drop multiplicative factors corresponding to the complexity of a learning graph on positive instances.

We decompose the analysis in stages as in Figure 6, and we compute their respective negative complexities on some given graph G. In the sequel, x, a and b are positive integers playing the role of parameters whose values are to be determined later.

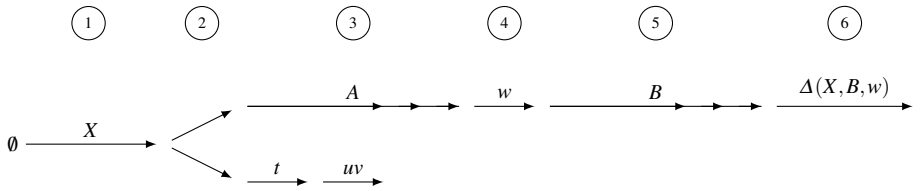


Fig. 6 Adaptive learning graph for Triangle Finding with its corresponding stages.

Stage 1. $\mathscr{G}$ consists in the combination of learning graphs $\mathscr{G}_X$, where X is a x-subset of $[n]$, as in Lemma 5. The particularity of the learning graphs $\mathscr{G}_X$ is that they all compute Triangle.
Stage 2. Each $\mathscr{G}_X$ is again a combination of two learning graphs $\mathscr{F}_X$ and $\mathscr{H}_X$ as in Lemma 5. The learning graph $\mathscr{F}_X$, described in the remaining stages, computes f_X, whereas $\mathscr{H}_X$ computes h_X. Observe that $\mathscr{H}_X$ consists in a very simple non-adaptative learning graph with negative complexity xn^2. Therefore we can already deduce that

$$C^0(\mathscr{G}, G) \leq \mathop{\mathbb{E}}_X(C^0(\mathscr{F}_X, G)) + xn^2.$$

In the sequel we focus on the analysis of $\mathscr{F}_X$.
Stage 3. $\mathscr{F}_X$ is decomposed using Theorem 3 with $\mathsf{Load}_S = \mathsf{DenseLoad}_S$ and parameters $n = |V|$, $k = a$, $r = 2$. Therefore

$$C^0(\mathscr{E}_X, G) = O\left((\mathbf{S}_{G,X})^2 + \left(\frac{n}{a}\right)^2\left(a\cdot(\mathbf{U}_{G,X})^2 + (\mathbf{C}_{G,X})^2\right)\right)$$

where we take

$$\begin{aligned}
(\mathbf{S}_{G,X})^2 &= \mathop{\mathbb{E}}_{A'\subset V\,:\,|A'|=a-2}\left(|I_X(A')|^2\right),\\
(\mathbf{U}_{G,X})^2 &= \max_{a-2\leq i<a}\left(\mathop{\mathbb{E}}_{A'\subset A''\subseteq V\,:\,|A'|=|A''|-1=i}\left(|I_X(A'')\backslash I_X(A')|^2\right)\right),\\
(\mathbf{C}_{G,X})^2 &= \mathop{\mathbb{E}}_{A\subseteq V\,:\,|A|=a}\left(C^0(\mathscr{F}_{X,A}, G)\right),
\end{aligned}$$

with $I_X(A') = X \times A'$, and $\mathscr{F}_{X,A}$ is the learning graph for $f_{X,A}$ that we describe in the remaining stages.
Stage 4. We use Lemma 5 and w stands for the third triangle vertex. Therefore,

$$C^0(\mathscr{F}_{X,A}, G) \leq n \times \mathop{\mathbb{E}}_w(C^0(\mathscr{F}_{X,A,w}, G)),$$

where $w \in [n]$ and $\mathscr{F}_{X,A,w}$ is the learning graph for $f_{X,A,w}$ described below.

Stage 5. Next, we use Theorem 3 with $\mathsf{Load}_S = \mathsf{DenseLoad}_S$ and parameters $n' = a$, $k' = b$, $r' = 2$. Therefore

$$C^0(\mathscr{F}_{X,A,w}, G) = O\left((\mathbf{S}_{G,X,A,w})^2 + \left(\frac{a}{b}\right)^2 \left(b \cdot (\mathbf{U}_{G,X,A,w})^2 + (\mathbf{C}_{G,X,A,w})^2\right)\right)$$

where we take

$$\begin{aligned}
(\mathbf{S}_{G,X,A,w})^2 &= \mathop{\mathbb{E}}_{B' \subset A \,:\, |B'| = b-2} \left(|I_w(B')|^2\right), \\
(\mathbf{U}_{G,X,A,w})^2 &= \max_{b-2 \le i < b} \left(\mathop{\mathbb{E}}_{B' \subset B'' \subseteq V \,:\, |B'| = |B''| - 1 = i} \left(|I_w(B'') \setminus I_w(B')|^2\right) \right), \\
(\mathbf{C}_{G,X,A,w})^2 &= \mathop{\mathbb{E}}_{B \subseteq A \,:\, |B| = b} \left(C^0(\mathscr{F}_{X,A,w,B}, G)\right),
\end{aligned}$$

with $I_w(B') = \{w\} \times B'$, and $\mathscr{F}_{X,A,w,B}$ is the learning graph for $f_{X,A,w,B}$ described in the last stage.

Stage 6. The last stage consists in the learning graph obtained by Lemma 5, with negative complexity of order $|\Delta(X,B,w)|$, for searching a potential edge in $\Delta(X,B,w)$.

In order to conclude, we observe that for any $w \in V$ and any set of vertices $V_1 \subseteq V$, we have $|I_X(V_1)| = x|V_1|$ and $|I_w(V_1)| = |V_1|$. Applying this for $V_1 = A$ and $V_1 = B$ we obtain

$$\mathbf{S}_{G,X} \le ax, \qquad \mathbf{U}_{G,X} \le x, \qquad \mathbf{S}_{G,X,A,w} \le b, \qquad \mathbf{U}_{G,X,A,w} \le 1.$$

We therefore get that $C^0(\mathscr{G}, G)$ has order

$$xn^2 + (ax)^2 + \left(\frac{n}{a}\right)^2 \left(a \cdot x^2 + n\left(b^2 + \left(\frac{a}{b}\right)^2 \left(b + \mathop{\mathbb{E}}_{X,w,B}\left[|\Delta(X,B,w)|\right]\right)\right)\right).$$

We now conclude using Lemma 6 with $V_1 = V$, and $C(\mathscr{G}) = \sqrt{C^0(\mathscr{G})}$ since $C^1(\mathscr{G}) \le 1$. □

Lemma 6 *Let x, b be positive integers. Let G be a graph on a vertex set V of size n and let $B \subseteq V$ be a b-subset. Then*

$$\mathop{\mathbb{E}}_{X,w}\left[|\Delta(X,B,w)|\right] \le \frac{b^2}{x},$$

where the expectation is taken over x-subsets $X \subseteq V$ and vertices $w \in V$.

Proof Let $\Delta(X)$ be the set of pairs of vertices which are not both neighbors of any vertex in X. Let $B \subseteq V$ of size b, the expectation on X and w is:

$$\mathop{\mathbb{E}}_{X,w}\left[|\Delta(X,B,w)|\right] = \sum_{(u,v) \in B^2} \Pr_{X,w}((u,v) \in \Delta(X,B,w)). \tag{5}$$

In order to bound the probabilities of the events on the right hand side, fix $(u,v) \in B^2$ and let $N_{u,v}$ be the intersection of the neighborhoods of u and v in V. Then

$$\Pr_{X,w}((u,v) \in \Delta(X,B,w)) = \Pr_{X,w}(w \in N_{u,v} \text{ and } (u,v) \in \Delta(X)).$$

The two events of the right hand side are independent, therefore with $t = |N_{u,v}|$ and $n = |V|$ we get

$$\Pr_{X,w}((u,v) \in \Delta(X,B,w)) = \frac{t}{n}\left(1 - \frac{t}{n}\right)^x.$$

Renaming $\alpha = \frac{tx}{n}$ leads to

$$\Pr_{X,w}((u,v) \in \Delta(X,B,w)) = \frac{\alpha}{x}\left(1 - \frac{\alpha}{x}\right)^x \le \frac{\alpha e^{-\alpha}}{x} \le \frac{1}{x}.$$

Finally, combining the above bound with equation (5) gives the result. □

Remark 1 Notice that the complexity of $\mathscr{H}_X$ (see **Stage 2** in the proof), namely xn^2, is negligible compared to the other terms. Hence the lower branch of the learning graph presented in Figure 6, could be removed by directly bounding the probability that X contains a vertex which belongs to a triangle.

5.2 Sparse graphs

In the sparse case we now show how to use extended learning graphs in order to get a better complexity than the one of Theorem 4.

First, the same learning graph of Theorem 4 has a much smaller complexity for sparse graphs when SparseLoad is used instead of DenseLoad.

Theorem 5 *The learning graph of Figure 5, using* Load = SparseLoad*, has complexity over graphs with m edges*

$$O\left(\sqrt{\left(xm+(ax)^2\cdot\frac{m}{n^2}+\left(\frac{n}{a}\right)^2\left(a\cdot x^2\cdot\frac{m}{n^2}+n\left(b^2\cdot\frac{m}{n^2}+\left(\frac{a}{b}\right)^2\left(b+\frac{b^2}{x}\right)\right)\right)\right)\log n}\right).$$

In particular, taking $a=n^{3/4}$ *and* $b=x=\sqrt{n}/(m/n^2)^{1/3}$ *leads to a complexity of* $O(n^{11/12}m^{1/6}\sqrt{\log n})$ *when* $m\geq n^{5/4}$.

Proof We reuse the notations introduced in the proof of Theorem 4. In addition we let $d=2m/n$ be the average degree of the input graph.

Here $\mathscr{H}_X$ has complexity $O(xdn\log n+x(d_2)^2)=O(xdn\log n)$ on any negative instance, where $d_2=\sqrt{\mathbb{E}_v(|N_v|^2)}$. Indeed, fix any negative instance. For each $v\in X$, we learn N_v by loading $\{v\}\times V$ with negative complexity $|N_v|(n+1)\log(n+1)$. Then we load $N_v\times N_v$ simply using DenseLoad with negative complexity $|N_v|^2$. So summing those complexities for every $v\in X$ and taking the expectation on x-subsets $X\subseteq V$, the average negative complexity becomes

$$\begin{aligned}&\mathop{\mathbb{E}}_{X\subseteq V,|X|=x}\left[\sum_{v\in X}\left(|N_v|(n+1)\log(n+1)+|N_v|^2\right)\right]\\&=x\times\mathop{\mathbb{E}}_{v\in V}\left(|N_v|(n+1)\log(n+1)+|N_v|^2\right)\\&=xd(n+1)\log(n+1)+x(d_2)^2.\end{aligned}$$

For the first step of the Johnson walk in $\mathscr{F}_X$, we now get using SparseLoad

$$\begin{aligned}\mathop{\mathbb{E}}_X\left((\mathbf{S}_{G,X})^2\right)&=\mathop{\mathbb{E}}_X\mathop{\mathbb{E}}_{A'\subseteq V:|A'|=a-2}\left(|I_X(A')|\cdot\log(|I_X(A')|+1)\cdot(|G_{I_X(A')}|+1)\right)\\&\leq ax\log(ax+1)\mathop{\mathbb{E}}_X\mathop{\mathbb{E}}_{A'\subseteq V:|A'|=a-2}(|E(X,A')|+1)\\&=O\left(\left(\frac{ax}{n}\right)^2m\log(ax+1)\right),\end{aligned}$$

where for the second step we used that $|I_X(A')|=|X\times A'|\leq ax$, and for the last one Lemma 8 below with X and $Y=A$. Similarly, using the fact that $|I_X(A'')\backslash I_X(A')|=|X\times(A''\backslash A')|=x$, we obtain

$$\begin{aligned}\mathop{\mathbb{E}}_X\left((\mathbf{U}_{G,X})^2\right)&=\mathop{\mathbb{E}}_X\left(x\cdot\log(x+1)\cdot\max_{a-2\leq i<a}\left(\mathop{\mathbb{E}}_{A'\subset V:|A'|=i}\left(\mathop{\mathbb{E}}_{v\in V\backslash A'}(|E(X,v)|+1)\right)\right)\right)\\&=O\left(\mathop{\mathbb{E}}_X\left(x\cdot\log(x+1)\cdot\left(\mathop{\mathbb{E}}_{v\in V}(|E(X,v)|+1)\right)\right)\right)\\&=O\left(\frac{x^2m}{n^2}\cdot\log(x+1)\right),\end{aligned}$$

where the second equality holds by Lemma 8 below with X and $Y=\{v\}$.

For the second step of the walk, since $|I_w(B')|=|\{w\}\times B'|\leq b$, we have by Lemma 7, with $x=b-2$, $V_1=A$ and $N=N_w\cap A$:

$$\begin{aligned}(\mathbf{S}_{G,X,A,w})^2&\leq b\log(b+1)\mathop{\mathbb{E}}_{B'\subset A:|B'|=b-2}\left(|E(B',w)|+1\right)\\&=O\left(\frac{b^2|N_w\cap A|}{a}\log(b+1)\right).\end{aligned}$$

Moreover, again by Lemma 7 below but this time with $x = a$, $V_1 = V$ and $N = N_w$, we get

$$\mathop{\mathbb{E}}_{w\in V}\mathop{\mathbb{E}}_{A\subseteq V\,:\,|A|=a}(|N_w\cap A|) = \frac{a}{n}\mathop{\mathbb{E}}_{w\in V}|N_w\cap V| = \frac{ad}{n}.$$

So,

$$\mathop{\mathbb{E}}_{A\subseteq V\,:\,|A|=a}\left((\mathbf{S}_{G,X,A,w})^2\right) = O\left(b^2\cdot\frac{d}{n}\log(b+1)\right).$$

Last, since $|I_w(B'')\backslash I_w(B')| = |\{w\}\times(B''\backslash B')| = 1$, we directly obtain:

$$\mathbf{U}^2_{G,X,A,w} = O(1).$$

Thus, the total negative complexity is of order

$$\underbrace{\left[xdn + (ax)^2\cdot\frac{d}{n} + \left(\frac{n}{a}\right)^2\left(a\cdot x^2\cdot\frac{d}{n} + n\left(b^2\cdot\frac{d}{n} + \left(\frac{a}{b}\right)^2\left(b+\frac{b^2}{x}\right)\right)\right)\right]}_{K_n}\times\log(n). \tag{6}$$

Denoting $t = \frac{d}{n} \leq 1$, we have:

$$K_n = t\left(xn^2 + a^2x^2 + \frac{n^2}{a^2}\left(ax^2+nb^2\right)\right) + \frac{n^3}{b}\left(1+\frac{b}{x}\right).$$

If $x = b \leq a \leq n$, we have $xn^2 \geq x^2n^2/a$ and $\frac{n^3}{b} \geq n^2$, hence :

$$K_n = O\left(t\left(xn^2 + a^2x^2 + \frac{n^3b^2}{a^2}\right) + \frac{n^3}{b}\right).$$

Taking $a = n^{3/4}$ and $b = x = \sqrt{n}/t^{1/3}$, leads to:

$$K_n = O\left(tb^2n^{3/2} + \frac{n^3}{b}\right) = O\left(n^{5/2}t^{1/3}\right).$$

Going back to (6), this yields a negative complexity of order:

$$n^{5/2}t^{1/3}\times\log(n),$$

and thus a total complexity of order:

$$n^{5/4}t^{1/6}\cdot\log(n)^{1/2} = n^{11/12}m^{1/6}\log(n)^{1/2}.$$

This complexity is valid until $\sqrt{n}/t^{1/3} \leq n^{3/4}$, that is when $t \geq 1/n^{3/4}$, *i.e.* $d \geq n^{1/4}$. This concludes the proof of the theorem. □

Lemma 7 *Let $1 \leq x \leq |V|$ and $N \subseteq V_1 \subseteq V$. Then*

$$\mathop{\mathbb{E}}_{X\subseteq V_1,\,|X|=x}|N\cap X| = \frac{x|N|}{|V_1|}.$$

Proof Let $\mathbf{1}_X$ be the indicator function of X. Then observe that the left hand side can be rewritten as

$$\mathop{\mathbb{E}}_{X}|N\cap X| = \mathop{\mathbb{E}}_{X}\left(\sum_{u\in N}\mathbf{1}_X(u)\right) = \sum_{u\in N}\mathop{\mathbb{E}}_{X}(\mathbf{1}_X(u)).$$

Then we conclude by observing that each term of the sum on the right hand side satisfies $\mathbb{E}_X(\mathbf{1}_X(u)) = \frac{x}{|V_1|}$, independently of $u \in V_1$. □

Lemma 8 *Let $1 \leq x,y \leq |V|$. Let $E(X,Y)$ denote the set of edges between X and Y. Then*

$$\mathop{\mathbb{E}}_{X,Y\subseteq V,\,|X|=x,\,|Y|=y}|E(X,Y)| = \frac{2xym}{n^2},$$

Proof For any $v \in V$ we denote $N_v \subseteq V$ its neighbors. We prove the equality by decomposition of the expectation term:

$$\begin{aligned}
\mathop{\mathbb{E}}_{X,Y} |E(X,Y)| &= \mathop{\mathbb{E}}_{X,Y} \sum_{v \in Y} |E(X,\{v\})| \\
&= \mathop{\mathbb{E}}_{X} \sum_{v \in V} |E(X,\{v\})| \times \Pr_{Y}(v \in Y) \\
&= \frac{y}{n} \times \sum_{v \in V} \mathop{\mathbb{E}}_{X} |N_v \cap X| \\
&= \frac{xy}{n^2} \times \sum_{v \in V} |N_v| \quad \text{by Lemma 7 with } k=1,\ V_1 = V \text{ and } N = N_v \\
&= \frac{2xym}{n^2}.
\end{aligned}$$

□

We now end with an even simpler learning graph whose complexity depends on its average of squared degrees. It consists in searching for a triangle vertex w. In order to check if w is such a vertex, we search for a b-subset B with an edge connected to w. For this purpose, we first connect w to B, and then check if there is an edge in $(N_w \cap B)^2$. See Figure 7 for the illustration.

Formally, we do the decomposition $\mathsf{Triangle} = \bigvee_{w \in V} f_w$, with $f_w(G) = 1$ iff w is a triangle vertex in G. Then, we pursue the decomposition with $f_w(G) = \bigvee_{B \subseteq V\,:\,|B|=b} f_{w,B}(G)$ where $f_{w,B}(G) = 1$ iff G has a triangle formed by w and two vertices of B. Using our notations, the resulting learning graph is represented by the diagram in Figure 8.

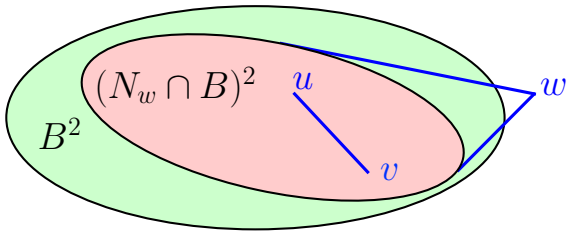


Fig. 7 Sets involved in the sparse decomposition used in Theorem 6.

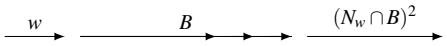


Fig. 8 Learning graph for Triangle Finding with complexity $\widetilde{O}((n^{5/6}m^{1/6} + d_2\sqrt{n})\log n)$.

We prove the following theorem, where $d_2 = \sqrt{\mathbb{E}_v\left[|N_v|^2\right]}$ denotes the quadratic mean of the degrees.

Theorem 6 *Let $b \geq n^2/m$. The learning graph of Figure 8, using* SparseLoad *for the first stage of $\mathscr{G}_{\text{Johnson}}$ and* DenseLoad *otherwise, has complexity over graphs with m edges*

$$O\left(\sqrt{n\left(b^2\frac{m}{n^2}\log n + \frac{n^2}{b^2}\left(b + \frac{b^2(d_2)^2}{n^2}\right)\right)}\right).$$

Taking $b = n^{4/3}/(m\log n)^{1/3}$ leads to a complexity of $O(n^{5/6}(m\log n)^{1/6} + d_2\sqrt{n})$.

Proof Let us denote $\mathscr{G}$ the learning graph of Figure 8. It can be seen as a special case of the one of Figure 5 with $X = \emptyset$ and $A = V$ (*i.e.* $x = 0$ and $a = n$). That is we start at Stage 5, and in our case $\Delta(X,B,w) = (N_w \cap B)^2$. Moreover we are going to use DenseLoad everywhere except for the first part, where we use SparseLoad in order to minimize the term $(\mathbf{S}_{G,X,A,w})^2$.

Therefore we can duplicate the analysis in the proof of Theorem 5 starting from Stage 4 and replacing $\Delta(X,B,w)$ by $(N_w \cap B)$. Then we get that the negative complexity for any graph G satisfies

$$C^0(\mathscr{G},G) = O\left(n\left(\frac{b^2 d}{n}\cdot\log(b+1) + \left(\frac{n}{b}\right)^2\left(b + \mathop{\mathbb{E}}_{w,B}\left(|N_w\cap B|^2\right)\right)\right)\right).$$

Then, the last piece of the proof is provided by Lemma 9 below which gives, with $x = b, V_1 = V$, and $N = N_w$,

$$\mathop{\mathbb{E}}_{w\in V, B\subseteq V\,:\,|B|=b}\left(|N_w\cap B|^2\right) \le 2\left(\mathop{\mathbb{E}}_{w}\left(\frac{b^2|N_w|^2}{n^2}\right)\right) \le 2\left(\frac{b^2(d_2)^2}{n^2}\right),$$

where $d_2 = \sqrt{\mathbb{E}_v\left[|N_v|^2\right]}$.

This concludes the proof of the theorem. □

Lemma 9 *Let $1 \le x \le |V|$ and $N \subseteq V_1 \subseteq V$ be such that $x|N| \ge |V_1|$. Then*

$$\mathop{\mathbb{E}}_{X\subseteq V_1,\,|X|=x}\left(|N\cap X|^2\right) \le 2\left(\frac{x|N|}{|V_1|}\right)^2.$$

Proof Similarly to the proof of Lemma 7, let $\mathbf{1}_X$ be the indicator function of X. Then

$$\mathop{\mathbb{E}}_{X}\left(|N\cap X|^2\right) = \mathop{\mathbb{E}}_{X}\left(\left(\sum_{u\in N}\mathbf{1}_X(u)\right)^2\right) = \sum_{u,v\in N}\mathop{\mathbb{E}}_{X}\left(\mathbf{1}_X(u)\mathbf{1}_X(v)\right).$$

Observe that $\mathbf{1}_X(u)$ and $\mathbf{1}_X(v)$ are independent for $u \ne v$, and that $(\mathbf{1}_X(u))^2 = \mathbf{1}_X(u)$. Therefore

$$\mathop{\mathbb{E}}_{X}\left(|N\cap X|^2\right) = \sum_{u,v\in N,\,u\ne v}\left(\mathop{\mathbb{E}}_{X}\mathbf{1}_X(u)\right)\left(\mathop{\mathbb{E}}_{X}\mathbf{1}_X(v)\right) + \sum_{u\in N}\mathop{\mathbb{E}}_{X}\mathbf{1}_X(u).$$

Remember that for all $u \in V_1$, $\mathbb{E}_X\,\mathbf{1}_X(u) = \frac{x}{|V_1|}$. Thus

$$\mathop{\mathbb{E}}_{X}\left(|N\cap X|^2\right) = |N|(|N|-1)\left(\frac{x}{|V_1|}\right)^2 + |N|\frac{x}{|V_1|}.$$

Using $x|N| \ge |V_1|$, we finally get

$$\mathop{\mathbb{E}}_{X}\left(|N\cap X|^2\right) \le |N|(|N|-1)\left(\frac{x}{|V_1|}\right)^2 + \left(|N|\frac{x}{|V_1|}\right)^2 \le 2\left(|N|\frac{x}{|V_1|}\right)^2.$$

□

Acknowledgements Frédéric Magniez was partially supported by the ERA-NET Cofund in Quantum Technologies project QuantAlgo and the French ANR Blanc project QuData.

References

1. Aaronson, S., Ben-David, S., Kothari, R.: Separations in query complexity using cheat sheets. In: Proceedings of 48th ACM Symposium on Theory of Computing, pp. 863–876 (2016)
2. Ambainis, A.: Quantum search with variable times. Theory of Computing Systems **47**(3), 786–807 (2010)
3. Ambainis, A.: Understanding quantum algorithms via query complexity (2017). ArXiv:1712.06349
4. Ambainis, A., Balodis, K., Belovs, A., Lee, T., Santha, M., Smotrovs, J.: Separations in query complexity based on pointer functions. In: Proceedings of 48th ACM Symposium on Theory of Computing, pp. 800–813 (2016)
5. Beals, R., Buhrman, H., Cleve, R., Mosca, M., Wolf, R.: Quantum lower bounds by polynomials. Journal of the ACM **48**(4), 778–797 (2001)
6. Belovs, A.: Learning-graph-based quantum algorithm for k-distinctness. In: Prooceedings of 53rd IEEE Symposium on Foundations of Computer Science, pp. 207–216 (2012)
7. Belovs, A.: Span programs for functions with constant-sized 1-certificates. In: Proceedings of 44th Symposium on Theory of Computing Conference, pp. 77–84 (2012)
8. Belovs, A., Childs, A., Jeffery, S., Kothari, R., Magniez, F.: Time-efficient quantum walks for 3-distinctness. In: Proceedings of 40th International Colloquium on Automata, Languages and Programming, pp. 105–122 (2013)

9. Belovs, A., Lee, T.: Quantum algorithm for k-distinctness with prior knowledge on the input. Tech. Rep. arXiv:1108.3022, arXiv (2011)
10. Belovs, A., Rosmanis, A.: On the power of non-adaptive learning graphs. In: Proceedings of 28th IEEE Conference on Computational Complexity, pp. 44–55 (2013)
11. Brassard, G., Høyer, P., Mosca, M., Tapp, A.: Quantum amplitude amplification and estimation. In: Quantum computation and information (Washington, DC, 2000), *Contemp. Math.*, vol. 305, pp. 53–74. Amer. Math. Soc., Providence, RI (2002). DOI 10.1090/conm/305/05215. URL https://doi.org/10.1090/conm/305/05215
12. Buhrman, H., Dürr, C., Heiligman, M., Høyer, P., Magniez, F., Santha, M., Wolf, R.: Quantum algorithms for element distinctness. SIAM Journal on Computing **34**(6), 1324–1330 (2005)
13. Childs, A.M.: Lecture notes on quantum algorithms (2017). Lecture notes
14. Cleve, R., Ekert, A., Macchiavello, C., Mosca, M.: Quantum algorithms revisited. R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci. **454**(1969), 339–354 (1998). DOI 10.1098/rspa.1998.0164. URL https://doi.org/10.1098/rspa.1998.0164. Quantum coherence and decoherence (Santa Barbara, CA, 1996)
15. Gall, F.L.: Improved quantum algorithm for triangle finding via combinatorial arguments. In: Proceedings of 55th IEEE Foundations of Computer Science, pp. 216–225 (2014)
16. Gall, F.L., Nakajima, S.: Quantum algorithm for triangle finding in sparse graphs. In: Proc. of 26th International Symposium Algorithms and Computation, pp. 590–600 (2015)
17. Grover, L.: A fast quantum mechanical algorithm for database search. In: Proceedings of 28th ACM Symposium on the Theory of Computing, pp. 212–219 (1996)
18. Høyer, P., Lee, T., Špalek, R.: Negative weights make adversaries stronger. In: Proceedings of 39th ACM Symposium on Theory of Computing, pp. 526–535 (2007)
19. Høyer, P., Mosca, M., de Wolf, R.: Quantum search on bounded-error imputs. In: Automata, languages and programming, *Lecture Notes in Comput. Sci.*, vol. 2719, pp. 291–299. Springer, Berlin (2003). DOI 10.1007/3-540-45061-0_25. URL https://doi.org/10.1007/3-540-45061-0_25
20. Høyer, P., Špalek, R.: Lower bounds on quantum query complexity. Bulletin of the European Association for Theoretical Computer Science **87** (2005)
21. Kitaev, A.: Quantum measurements and the Abelian stabilizer problem. Tech. Rep. quant-ph/9511026, arXiv (1995)
22. Kitaev, A.Y., Shen, A.H., Vyalyi, M.N.: Classical and quantum computation, *Graduate Studies in Mathematics*, vol. 47. American Mathematical Society, Providence, RI (2002). DOI 10.1090/gsm/047. URL https://doi.org/10.1090/gsm/047. Translated from the 1999 Russian original by Lester J. Senechal
23. Lee, T., Magniez, F., Santha, M.: Improved quantum query algorithms for triangle finding and associativity testing. Algorithmica (2015). To appear
24. Lee, T., Mittal, R., Reichardt, B., Špalek, R., Szegedy, M.: Quantum query complexity of state conversion. In: Proceedings of 52nd IEEE Symposium on Foundations of Computer Science, pp. 344–353 (2011)
25. Magniez, F., Nayak, A., Roland, J., Santha, M.: Search via quantum walk. SIAM Journal on Computing **40**(1), 142–164 (2011)
26. Magniez, F., Santha, M., Szegedy, M.: Quantum algorithms for the triangle problem. SIAM Journal on Computing **37**(2), 413–424 (2007)
27. Nisan, N.: Crew prams and decision trees. SIAM Journal on Computing **20**(6), 999–1007 (1991)
28. Reichardt, B.: Reflections for quantum query algorithms. In: Proceedings of 22nd ACM-SIAM Symposium on Discrete Algorithms, pp. 560–569 (2011)
29. Shor, P.: Algorithms for quantum computation: Discrete logarithm and factoring. SIAM Journal on Computing **26**(5), 1484–1509 (1997)
30. Špalek, R.: Quantum algorithms, lower bounds, and time-space tradeoffs (2006). PhD thesis