

Pre-Reduction Graph Products: Hardnesses of Properly Learning DFAs and Approximating EDP on DAGs

Parinya Chalermsook* Bundit Laekhanukit† Danupon Nanongkai‡

Abstract

The study of graph products is a major research topic and typically concerns the term $f(G * H)$, e.g., to show that $f(G * H) = f(G)f(H)$. In this paper, we study graph products in a non-standard form $f(R[G * H])$ where R is a “reduction”, a transformation of any graph into an instance of an intended optimization problem. We resolve some open problems as applications.

The first problem is *minimum consistent deterministic finite automaton (DFA)*. We show a tight $n^{1-\epsilon}$ -approximation hardness, improving the $n^{1/14-\epsilon}$ hardness of [Pitt and Warmuth, STOC 1989 and JACM 1993], where n is the sample size. (In fact, we also give improved hardnesses for the case of *acyclic* DFA and NFA.) Due to Board and Pitt [Theoretical Computer Science 1992], this implies the *hardness of properly learning DFAs* assuming $\text{NP} \neq \text{RP}$ (the weakest possible assumption). This affirmatively answers an open problem raised 25 years ago in the paper of Pitt and Warmuth and the survey of Pitt [All 1989]. Prior to our results, this hardness only follows from the stronger hardness of *improperly* learning DFAs, which requires stronger assumptions, i.e., either a cryptographic or an average case complexity assumption [Kearns and Valiant STOC 1989 and J. ACM 1994; Daniely et al. STOC 2014]. The second problem is *edge-disjoint paths* (EDP) on *directed acyclic graphs* (DAGs). This problem admits an $O(\sqrt{n})$ -approximation algorithm [Chekuri, Khanna, and Shepherd, Theory of Computing 2006] and a matching $\Omega(\sqrt{n})$ integrality gap, but so far only an $n^{1/26-\epsilon}$ hardness factor is known [Chuzhoy et al., STOC 2007]. (n denotes the number of vertices.) Our techniques give a tight $n^{1/2-\epsilon}$ hardness for EDP on DAGs, thus resolving its approximability status.

As by-products of our techniques: (i) We give a tight hardness of packing vertex-disjoint k -cycles for large k , complimenting [Guruswami and Lee, ECCC 2014] and matching [Krivelevich et al., SODA 2005 and ACM Transactions on Algorithms 2007]. (ii) We give an alternative (and perhaps simpler) proof for the hardness of properly learning DNF, CNF and intersection of halfspaces [Alekhovich et al., FOCS 2004 and J. Comput.Syst. Sci. 2008]. Our new concept reduces the task of proving hardnesses to merely analyzing graph product inequalities, which are often as simple as textbook exercises. This concept was inspired by, and can be viewed as a generalization of, the *graph product subadditivity* technique we previously introduced in SODA 2013. This more general concept might be useful in proving other hardness results as well.

*Max-Planck-Institut für Informatik, Germany. Work partially done while at IDSIA, Switzerland. Supported by the Swiss National Science Foundation project 200020_144491/1

†McGill University, Canada. Supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) grant no. 28833 and by Dr&Mrs M.Leong fellowship. Istituto Dalle Molle di Studi sull’Intelligenza Artificiale (IDSIA). Supported by ERC starting grant 279352 (NEWNET).

‡University of Vienna, Austria. Work partially done while at Nanyang Technological University (NTU), Singapore, and ICERM, Brown University, USA.

Contents

1	Introduction	1
1.1	The Concept of Pre-Reduction Graph Product	1
1.2	Problems and Our Results	2
2	Overview	6
2.1	Example of Reduction R : Vertex-Disjoint Paths	6
2.2	The Use of Pre-Reduction Products	7
2.3	Toward A Tight Hardness of the Minimum Consistent DFA Problem	8
2.4	Related Concept	9
2.5	Organization	10
3	Preliminaries	10
3.1	Terms	10
3.2	Problems	10
4	Meta Theorems	12
4.1	Proof of Theorem 4.1 (Meta Theorem for Maximization Problems)	13
4.2	Proof of Theorem 4.2 (Meta Theorem for Minimization Problems)	14
4.3	Overview of Applications	14
5	Hardnesses of Finite Automata Problems: Minimum Consistency and Proper PAC Learning	15
5.1	Hardness of MINCON(ADFA, NFA) via Graph Products	15
5.2	Tight Hardness for MINCON(ADFA, DFA)	18
5.3	Hardness of Proper PAC-Learning	22
6	Hardness of EDP on DAGs	23
6.1	Reduction R	23
6.2	α -Projection Property	25
6.3	Geometry of Paths: Regions, switching boxes, and configurations	25
6.4	Proof	26
7	Other Problems	29
7.1	Hardness of k -Cycle Packing for Large k	29
7.2	Learning CNF Formula	30
8	Conclusion and Open Problems	31
A	List of Bad Examples	35
A.1	Edge Disjoint Paths	35

1 Introduction

1.1 The Concept of Pre-Reduction Graph Product

Background: Graph Product and Hardness of Approximation. Graph product is a fundamental tool with rich applications in both graph theory and theoretical computer science. It is, roughly speaking, a way to combine two graphs, say G and H , into a new graph denoted by $G * H$. For example, the following *lexicographic product*, denoted by $G \cdot H$, will be particularly useful in this paper.

$$\begin{aligned} \text{(Lexicographic Product)} \quad V(G \cdot H) &= V(G) \times V(H) = \{(u, v) : u \in V(G) \text{ and } v \in V(H)\}. \\ E(G \cdot H) &= \{(u, a)(v, b) : uv \in E(G) \text{ or } (u = v \text{ and } ab \in E(H))\}. \end{aligned} \quad (1)$$

A common study of graph product aims at understanding how $f(G * H)$ behaves for some function f on graphs denoting a graph property. For example, if we let $\alpha(G)$ be the *independence number* of G (i.e., the cardinality of the maximum independent set), then $\alpha(G \cdot H) = \alpha(G)\alpha(H)$.

Graph products have been extremely useful in *boosting* the hardness of approximation. One textbook example is proving the hardness of n^ϵ for approximating the maximum independent set problem (i.e., approximating $\alpha(G)$ of an input graph G): Berman and Schnitger [BS92] showed that we can reduce from Max 2SAT to get a constant approximation hardness $c > 1$ for the maximum independent set problem, and then use a graph product to boost the resulting hardness to n^ϵ for some (small) constant ϵ . To illustrate how graph products amplify hardness, suppose we have a (1.001)-gap reduction $R[I]$ that transforms an instance I of SAT into a graph G . Since $\alpha(\cdot)$ is multiplicative, if we take a product $R[I]^k$ for any integer k , the hardness gap immediately becomes $(1.0001)^k = 2^{\Omega(k)}$. Choosing k to be large enough gives $2^{\log^{1-\epsilon} n}$ hardness. Therefore, once we can rule out the PTAS, graph products can be used to boost the hardness to almost polynomial. This idea is also used in many other problems, e.g., in proving the hardness of the longest path problem [KMR97].

Our Concept: Pre-Reduction Graph Product. This paper studies a reversed way to apply graph products: instead of the commonly used form of $(R[I])^k = (R[I] * R[I] * \dots)$ to boost the hardness of approximation, we will use $R[I^k] = R[I * I * \dots]$; here, I is a graph which is an instance of a hard graph problem such as maximum independent set or minimum coloring. We refer to this approach as *pre-reduction graph product* to contrast the previous approach in which graph product is performed *after* a reduction (which will be referred to as *post-reduction graph product*). The main conceptual contribution of this paper is the demonstration to the power and versatility of this approach in proving approximation hardnesses. We show our results in Section 1.2 and will come back to explain this concept in more detail in Section 2.

We note one conceptual difference here between the previous post-reduction and our pre-reduction approaches: While the previous approach starts from a reduction R that already gives some hardness result, our approach usually starts from a reduction that does *not immediately* provide any hardness result; in other words, such reduction alone cannot be used to even prove NP-hardness. (See Section 2 for an illustration.) Moreover, in contrast to the previous use of $(R[I])^k$ which requires $R[I]$ to be a graph, our approach allows us to prove hardnesses of problems whose input instances are not graphs. Also note that our approach gives rise to a study of graph products in a new form: in contrast to the usual study of $f(G * H)$, our hardness results crucially

Problems	Upper Bounds	Prev. Hardness	New Hardness
MINCON(DFA, DFA)	$O(n)$	$n^{1/14-\epsilon}$ [PW93]	$n^{1-\epsilon}$
EDP on DAGs	$\tilde{O}(n^{1/2})$ [CKS06]	$n^{1/26-\epsilon}$ [CGKT07]	$n^{1/2-\epsilon}$
k -cycle packing	$O(\min(k, n^{1/2}))$	$\Omega(k)$ [GL14]	$O(\min(k, n^{1/2-\epsilon}))$
MinCon(CNF, CNF), MinCon(DNF, DNF), MinCon(Halfspace, Halfspace)	$O(n)$	$n^{1-\epsilon}$	$n^{1-\epsilon}$ (Alternative proof)

Table 1: Summary of our hardness results.

rely on understanding the behavior of $f(R[G * H])$ for some function f , reduction R , and graph product $*$ (which happens to always be the lexicographic product in this paper). Another feature of this approach is that it usually leads to simple proofs that do not require heavy machineries (such as the PCP-based construction) – some of our hardness proofs are arguably simplifications of the previous ones; in fact, most of our hardness results follow from the meta-theorem (see Section 4) which shows that a bounds of $f(R[G * H])$ in a certain form will immediately lead to hardness results. We list some bounds of $f(R[G * H])$ in Theorem 2.1.

1.2 Problems and Our Results

1.2.1 Minimum Consistent DFA and Proper PAC-Learning DFAs

In the *minimum consistent deterministic finite automaton* (DFA) problem, denoted by MINCON(DFA, DFA), we are given two sets \mathcal{P} and \mathcal{N} of *positive* and *negative sample* strings in $\{0, 1\}^*$. We let the *sample size*, denoted by n , be the total number of bits in all sample strings. Our goal is to construct a DFA M (see Section 3 for a definition) of *minimum size* that is *consistent* with all strings in $\mathcal{P} \cup \mathcal{N}$. That is, M accepts all positive strings $x \in \mathcal{P}$ and rejects all negative strings $y \in \mathcal{N}$.

This problem can be easily approximated within $O(n)$. Due to its connections to PAC-learning automata and grammars (e.g. [DIH10, Pit89]), the problem has received a lot of attention from the late 70s to the early 90s. The NP-hardness of this problem was proved by Gold [Gol78] and Angluin [Ang78]. Li and Vazirani [LV88] later provided the first hardness of approximation result of $(9/8 - \epsilon)$. This was greatly improved to $n^{1/14-\epsilon}$ by Pitt and Warmuth [PW93]. Our first result is a tight $n^{1-\epsilon}$ hardness for this problem, improving [PW93]. In fact, our hardness result holds even when we allow an algorithm to compare its result to the optimal *acyclic* DFA (ADFA), which is larger than the optimal DFA. This problem is called MINCON(ADFA, DFA); see Section 3 for detailed definitions.

Theorem 1.1. *Given a pair of positive and negative samples $(\mathcal{P}, \mathcal{N})$ of size n where each sample has length $O(\log n)$, for any constant $\epsilon > 0$, it is NP-hard to distinguish between the following two cases of MinCon(ADFA, DFA):*

- YES-INSTANCE: *There is an ADFA of size n^ϵ consistent with $(\mathcal{P}, \mathcal{N})$.*
- NO-INSTANCE: *Any DFA that is consistent with $(\mathcal{P}, \mathcal{N})$ has size at least $n^{1-\epsilon}$.*

In particular, it is NP-hard to approximate the minimum consistent DFA problem to within a factor of $n^{1-\epsilon}$.

The main motivation of this problem is its connection to the notion of *properly PAC-learning* DFAs. It is one of the most basic problems in the area of proper PAC-learning [DIH10, Pit89, PW93]. Roughly speaking, the problem is to learn an unknown DFA M from given random samples, where a learner is asked to output (based on such random samples) a DFA M' that closely approximates M (see, e.g., [Fel08] for details). The main question is whether DFA is properly PAC-learnable.

This question was the main motivation behind [PW93]; however, the $n^{1/14-\epsilon}$ hardness in [PW93] was not strong enough to prove this. Kearns and Valiant [KV94] showed that a proper PAC-learning of DFAs is not possible if we assume a cryptographic assumption stronger than $P \neq NP$. In fact, their result implies that even *improperly* PAC-learning DFAs (i.e., the output does not have to be a DFA) is impossible. Very recently, Daniely et al. [DLSS14] obtained a similar result by assuming a (fairly strong) average-case complexity assumption generalizing Feige’s assumption [Fei02].

The question whether the cryptographic assumption could be replaced by the $RP \neq NP$ assumption (which would be the weakest assumption possible) was asked 25 years ago in [Pit89, PW93]. In particular, the following is the first open problem in [Pit89]: (i) *Can it be shown that DFAs are not properly PAC-learnable based only on the assumption that $RP \neq NP$?* (ii) *Stronger still, can the improper learnability result of [KV94] be strengthened by replacing the cryptographic assumptions with only the assumption that $RP \neq NP$?*

Applebaum, Barak and Xiao [ABX08] showed that proving lower bounds for improper learning using many standard ways of reductions from NP-hard problems will not work unless the polynomial hierarchy collapses, suggesting that an answer to the second question is likely to be negative. For the first question, some hardnesses of proper PAC-learning assuming $RP \neq NP$ were already known at the time (e.g. [PV88]) and there are many more recent results (see, e.g., [Fel08] and references therein). Despite this, the basic problem of learning DFAs (originally asked in the above question) remained open. Theorem 1.1 together with a result of Board and Pitt [BP92] immediately resolve this problem.

Corollary 1.2. *Unless $NP = RP$, the class of DFAs is not properly PAC-learnable.*

We also note an amusing connection between this type of result and Chomsky’s “Poverty of the Stimulus Argument”, as noted by Aaronson [Aar08]: “Let’s say I give you a list of n -bit strings, and I tell you that there’s some nondeterministic finite automaton M , with much fewer than n states, such that each string was produced by following a path in M . Given that information, can you reconstruct M (probably and approximately)? It’s been proven that if you can, then you can also break RSA!” Our Corollary 1.2 implies that for the case of deterministic finite automaton, being able to reconstruct M will imply not only that one can break RSA but also solve, for instance, the traveling salesman problem (TSP) probabilistically.

1.2.2 Edge-Disjoint Paths on DAGs

In the edge-disjoint paths problem (EDP), we are given a graph $G = (V, E)$ (which could be directed or undirected) and k source-sink pairs $s_1t_1, s_2t_2, \dots, s_kt_k$ (a pair can occur multiple times). The objective is to connect as many pairs as possible via edge-disjoint paths. Throughout, we let n and m be the number of vertices and edges in G , respectively. Approximating EDP has been extensively studied. It is one of the major challenges in the field of approximation algorithms. The problem has received significant attention from many groups of researchers, attacking the problem from many angles and considering a few variants and special cases (see, e.g., [RS95, Chu12, CL12, CKS09, CKS05, Kle05, KT98, KK10] and references therein).

Cases	Upper Bounds	Integrality Gap	Prev. Hardness
Undirected	$O(n^{1/2})$ [CKS06]	$\Omega(n^{1/2})$	$\log^{1/2-\epsilon} n$ [ACG+10]
DAGs	$\tilde{O}(n^{1/2})$ [CKS06]	$\Omega(n^{1/2})$	$n^{1/26-\epsilon}$ [CGKT07]
Directed	$O(\min(m^{1/2}, n^{2/3}))$ [Kle96, CK07, VV04]	$\Omega(n^{1/2})$	$n^{1/2-\epsilon}$ [GKR+03]

Table 2: The current status of EDP.

In directed graphs, EDP can be approximated within a factor of $O(\min(m^{1/2}, n^{2/3}))$ [Kle96, CK07, VV04]. The $O(m^{1/2})$ factor is *tight* on sparse graphs since directed EDP is NP-hard to approximate within a factor of $n^{1/2-\epsilon}$, for any $\epsilon > 0$ [GKR+03]. In contrast to the directed case, undirected EDP is much less understood: The approximation factor for this case is $O(n^{1/2})$ [CKS06] with a matching integrality gap of $\Omega(n^{1/2})$ for its natural LP relaxation, suggesting an $n^{1/2-\epsilon}$ hardness. Despite these facts, we only know a $\log^{1/2-\epsilon} n$ hardness of approximation assuming $\text{NP} \not\subseteq \text{ZPTIME}(n^{\text{polylog}(n)})$. Even in special cases such as planar graphs (or, even simpler, brick-wall graphs, a very structured subclass of planar graphs), it is still open whether undirected EDP admits an $o(n^{1/2})$ approximation algorithm. This obscure state of the art made undirected EDP one of the most important, intriguing open problems in graph routing. (Table 2 summarizes the current status of EDP.)

One problem that may help in understanding undirected EDP is perhaps EDP on *directed acyclic graphs* (DAGs). This case is interesting because (i) its complexity seems to lie somewhere between the directed and undirected cases, (ii) it shares some similar statuses and structures with undirected EDP, and (iii) it has close connections to directed cycle packing [KNS+07] (i.e., hard instances for EDP on DAGs are used as gadgets in constructing the hard instance for directed cycle packing). In particular, on the upper bound side, the technique in [CKS06] gives an $O(n^{1/2} \text{poly log } n)$ upper bound not only to undirected EDP but also to EDP on DAGs. Moreover, the integrality gap of $\Omega(n^{1/2})$ applies to both cases, suggesting a hardness of $n^{1/2-\epsilon}$ for them. However, previous hardness techniques for the case of general directed graphs [GKR+03] completely fail to give a lower bound on both DAGs and undirected graphs¹. On the other hand, subsequent techniques that were invented in [AZ06, ACG+10] to deal with undirected EDP can be strengthened to prove the currently best hardness for EDP on DAGs [CGKT07]², which is $n^{1/26-\epsilon}$. These results suggest that the complexity of EDP on DAGs lies between undirected and directed graphs. In this paper, we show that our techniques give a hardness of $n^{1/2-\epsilon}$ for this case, thus completely settling its approximability status. Our result is formally stated in the following theorem.

Theorem 1.3. *Given an instance of EDP on DAGs, consisting of a graph $G = (V, E)$ on n vertices and a source-sink pairs $(s_1, t_1), \dots, (s_k, t_k)$, for any $\epsilon > 0$, it is NP-hard to distinguish between the following two cases:*

- YES-INSTANCE: *There is a collection of edge disjoint paths in G that connects $1/n^\epsilon$ fraction of the source-sink pairs.*
- NO-INSTANCE: *Any collection of edge disjoint paths in G connects at most $1/n^{1/2-\epsilon}$ fraction of the source-sink pairs.*

¹The result in [GKR+03] crucially relies on the fact that EDP with 2 terminal pairs is hard on directed graphs. This is not true if the graph is a DAG or undirected.

²Their result is in fact proved in a more general setting of EDP with congestion c for any $c \geq 1$

In particular, it is NP-hard to approximate EDP on DAGs to within a factor of $n^{1/2-\epsilon}$.

1.2.3 Other Results

Minimum Consistent NFA. Our techniques also allow us to prove a hardness result for the *minimum consistent NFA* problem as stated formally in the following theorem.

Theorem 1.4. *Given a pair of positive and negative samples $(\mathcal{P}, \mathcal{N})$ of size n where each sample has length $O(\log n)$, for any constant $\epsilon > 0$, it is NP-hard to distinguish between the following two cases of $\text{MinCon}(\text{ADFA}, \text{NFA})$:*

- YES-INSTANCE: *There is an ADFA of size n^ϵ consistent with $(\mathcal{P}, \mathcal{N})$.*
- NO-INSTANCE: *Any NFA that is consistent with $(\mathcal{P}, \mathcal{N})$ has size at least $n^{1/2-\epsilon}$.*

In particular, it is NP-hard to approximate the minimum consistent NFA problem to within a factor of $n^{1/2-\epsilon}$.

This improves upon the $n^{1/4-\epsilon}$ hardness of Pitt and Warmuth [PW93]. We note that this hardness result is not strong enough to imply a PAC-learning lower bound for NFAs. Such hardness was already known based on some cryptographic or average-case complexity assumptions [KV94, DLSS14]. We think it is an interesting open problem to remove these assumptions as we did for the case of learning DFAs.

k -Cycle Packing. Our reduction for EDP can be slightly modified to obtain hardness results for k -Cycle Packing, when k is large. In the k -cycle packing problem, given an input graph G , one wants to pack as many disjoint cycles as possible into the graph while we are only interested in cycles of length at most k . An $O(\min(k, n^{1/2}))$ -approximation algorithm for this problem can be easily obtained by modifying the algorithm of Krivelevich et al. [KNS⁺07]. Very recently, Guruswami and Lee [GL14] obtained a hardness of $\Omega(k)$, assuming the Unique Game Conjecture, when k is a constant. This matches the upper bound of Krivelevich et al. for small k . In this paper, we compliment the result of Guruswami and Lee by showing a hardness of $n^{1/2-\epsilon}$ for some $k \geq n^{1/2}$, matching the upper bound of Krivelevich et al. for the case of large k .

Theorem 1.5. *Given a directed graph G , for any $\epsilon > 0$ and some $k \geq |V(G)|^{1/2}$, it is NP-hard to distinguish between the following cases:*

- *There are at least $|V(G)|^{1/2-\epsilon}$ disjoint cycles of length k in G .*
- *There are at most $|V(G)|^\epsilon$ disjoint cycles of length at most $2k - 1$ in G .*

In particular, for some $k \geq n^{1/2}$, the k -cycle packing problem on n -vertex graphs is hard to approximate to within a factor of $n^{1/2-\epsilon}$.

Alternative Hardness Proof for Minimum Consistent CNF, DNF, and Intersections of Halfspaces. Our techniques for proving the DFA hardness result can be used to give an alternative proof for the hardness of the minimum consistent DNF, CNF, and intersections of thresholded halfspaces problems. In the minimum consistent CNF problem, we are given a collection of samples of size n , and our goal is to output a small CNF formula that is consistent with all such

samples. Alekhovich et al. [ABF⁺08] previously showed tight hardnesses for these problems, which imply that the classes of CNFs, DNFs, and the intersections of halfspaces are not properly PAC-learnable. Our techniques give an alternative proof (which might be simpler) for these results. More specifically, we give an alternative proof for the following theorem and corollary (stated in terms of CNF, but the same holds for DNF and intersection of halfspaces³).

Theorem 1.6. *Let $\epsilon > 0$ be any constant. Given a pair of positive and negative samples $(\mathcal{P}, \mathcal{N})$ of size n where each sample has length at most n^ϵ , it is NP-hard to distinguish between the following two cases:*

- YES-INSTANCE: *There is a CNF formula of size n^ϵ consistent with $(\mathcal{P}, \mathcal{N})$.*
- NO-INSTANCE: *Any CNF consistent with $(\mathcal{P}, \mathcal{N})$ must have size at least $n^{1-\epsilon}$.*

In particular, it is NP-hard to approximate the minimum consistent CNF problem to within a factor of $n^{1-\epsilon}$.

Corollary 1.7. *Unless $\text{NP} = \text{RP}$, the class of CNF is not properly PAC-learnable.*

2 Overview

2.1 Example of Reduction R : Vertex-Disjoint Paths

To illustrate the pre-reduction graph product concept, consider the *vertex-disjoint path* (VDP) problem. The objective of VDP is the same as that of EDP except that we want paths to be vertex-disjoint instead of edge-disjoint. The approximability statuses of EDP and VDP on DAGs and undirected graphs are the same, and we choose to present VDP due to its simpler gadget construction. Our hardness of VDP can be easily turned into a hardness of EDP.

Our goal is to show that this problem has an approximation hardness of $n^{1/2-\epsilon}$, where n is the number of vertices. We will use the following reduction⁴ R which transforms a graph G (supposedly an input instance of the maximum independent set problem) into an instance $R[G]$ of the vertex-disjoint paths problem with $\Theta(|V(G)|^2)$ vertices. We start with an instance $R[G]$ as in Figure 1a where there are k source-sink pairs (Figure 1a shows an example where $k = 6$) and edges are oriented from left to right and from top to bottom. Let us name vertices in G by $1, 2, \dots, k$. For any pair of vertices i and j , where $i < j$, such that edge ij does *not* present in G , we remove a vertex v_{ij} from $R[G]$, as shown in Figure 1b (this means that two edges that point to v_{ij} will continue on their directions without intersecting each other). See Section 6 for the full description of R in the context of EDP.

To see an intuition of this reduction, define a *canonical path* be a path that starts at some source s_i , goes all the way right, and then goes all the way down to t_i (e.g., a thick (green) path in Figure 1b). It can be easily seen that any set of vertex-disjoint paths in $R[G]$ that consists only of canonical paths can be converted to a solution for the maximum independent set problem. Conversely any independent set S in G can be converted to a set of $|S|$ vertex-disjoint paths. For example, canonical paths between the pairs (s_1, t_1) and (s_2, t_2) in $R[G]$ in Figure 1b can be converted

³It is noted in [ABF⁺08] that one only needs to prove the hardness of CNF, since this problem is a special case of the intersection of thresholded halfspaces problem, and the proof for DNF would work similarly.

⁴We thank Julia Chuzhoy who suggested this reduction to us (private communication).

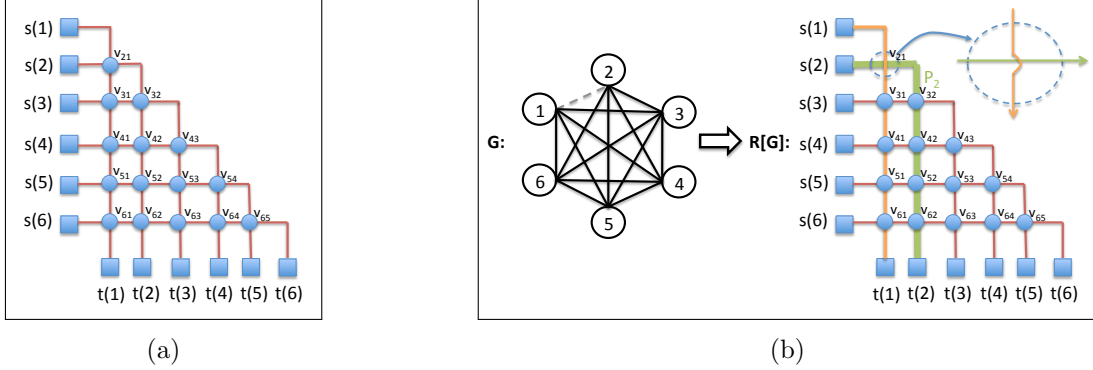


Figure 1: The reduction R for the vertex-disjoint paths problem. The thick (green) path in Figure 1b shows an example of a canonical path.

to an independent set $\{1, 2\}$ in G and vice versa. In other words, if we can *force* the VDP solution to consist only of canonical paths, then we can potentially use the $|V|^{1-\epsilon}$ hardness of maximum independent set to prove a tight $|V|^{1-\epsilon} = |V(R[G])|^{1/2-\epsilon}$ hardness of VDP. This intuition, however, cannot be easily turned into a hardness result since the VDP solution can use non-canonical paths, and it is possible that $\text{VDP}(R[G])$ is much larger than $\alpha(G)$; see Appendix A.1 for an example where $\alpha(G) = O(1)$ and $\text{VDP}(R[G]) = \Omega(|V(G)|)$. Thus, the reduction R by itself cannot be used even to prove that VDP is NP-hard!

2.2 The Use of Pre-Reduction Products

The above situation is very common in attempts to prove hardnesses for various problems. A usual way to obtain hardness results is to modify R into some reduction R' . This modification, however, often blows up the size of the reduction, thus affecting its tightness. For example, VDP and EDP on DAGs are only known to be $n^{1/26-\epsilon}$ -hard, as opposed to being potentially $n^{1/2-\epsilon}$ -hard, as suggested by the integrality gap. Moreover, the reduction R' is usually much more complicated than R . In this paper, we show that for many problems the above difficulties can be avoided by simply picking an appropriate graph product $*$ and understanding the structure of $R[G * H * \dots]$. To this end, it is sometimes easier to study $f(R[G * H])$ for any graphs G and H , although we eventually need only the case where $G = H$. This gives rise to the study of the behavior of $f(R[G * H])$ which is a non-standard form of graph product in comparison with the standard study of $f(G * H)$. In fact, most results in this paper follow merely from bounding $f(R[G * H])$ in the form

$$g(G * H) \leq f(R[G * H]) \leq g(G)f(H) + \text{poly}(|V(G)|), \quad (2)$$

where g is an objective function of a problem whose hardness is already known (in this paper, g is either maximum independent set or minimum coloring), and f is an objective function of a problem that we intend to prove hardness. Our bounds for functions f corresponding to problems that we want to solve, e.g. the minimum consistent DFA (function dfa) and maximum edge-disjoint paths (function edp), are listed in the theorem below. (Recall that $G \cdot H$ denotes the lexicographic product as defined in Equation (1).)

Theorem 2.1 (Bounds of graph products; informal). *There is a reduction R_1 (respectively R_2) that transforms a graph G into an instance of the minimum consistent DFA problem of size $\tilde{\Theta}(|V(G)|^2)$ (respectively the maximum edge-disjoint paths problem of size $\Theta(|V(G)|^2)$) such that, for any graphs G and H ,*

$$\chi(G \cdot H) \leq \text{dfa}(R_1[G \cdot H]) \leq \chi(G)\text{dfa}(R_1[H]) + O(|V(G)|^2) \quad (3)$$

$$\alpha(G \cdot H) \leq \text{edp}(R_2[G \cdot H]) \leq \alpha(G)\text{edp}(R_2[H]) + O(|V(G)|^2) \quad (4)$$

See Section 5.1 (especially, Corollary 5.3 and Lemma 5.4) and Section 6 (especially, Lemma 6.3) for the details and proofs of Equations (3) and (4), respectively. It only requires a systematic, simple calculation to show that these inequalities imply hardnesses of approximation; we formulate this implication as a “meta theorem” (see Section 4) which roughly states that for large enough k ,

$$f(R[G^k]) \approx g(G^k) \quad (5)$$

where G^k is $G * G * G * \dots$ (k times). (For an intuition, observe that when k is large enough, the term $\text{poly}(|V(G)|)$ in Equation (2) will be negligible and an inductive argument can be used to show that $g(G^k) \leq f(R[G^k]) \leq g(G)^{O(k)}$ (recall that, in our case, g is multiplicative)). This means that the hardness of f^5 is at least the same as the hardness of g on graph product instances G^k . For the case of DFA and EDP, $R_1(G)$ and $R_2(G)$ increase the size of input size to $|V(G)|^2$ while α and χ have the hardness of $|V(G)|^{1-\epsilon}$. Thus, we get a hardness of $n^{1/2-\epsilon}$ where n is the input size of DFA and EDP. This immediately implies a tight hardness for EDP and an improved hardness of DFA. How this translates to a hardness of f depends on how much instance blowup the reduction $R[G^k]$ causes. For our problems of DFA and EDP, it is a well known result that the hardness of α and χ stays roughly the same under the lexicographic product, i.e., α and χ on G^k have a hardness of $|V(G^k)|^{1-\epsilon}$. The meta theorem and Theorem 2.1 say that this hardness also holds for DFA and EDP. Since $R_1[G^k]$ and $R_2[G^k]$ increase the size of input instances by a quadratic factor — from $|V(G^k)|$ to $n = |V(G^k)|^2$ — we get a hardness of $n^{1/2-\epsilon}$ where n is the input size of DFA and EDP. This immediately implies a tight hardness for EDP and an improved hardness for DFA.

2.3 Toward A Tight Hardness of the Minimum Consistent DFA Problem

To get the tight $n^{1-\epsilon}$ hardness for DFA, we have to adjust R_1 in Theorem 2.1 to avoid the quadratic blowup. We will exploit the fact that, to get a result similar to Equation (5), we only need a reduction R defined on the k -fold graph product G^k instead of on an arbitrary graph G as in the case of R_1 . We modify reduction R_1 to $R_{1,k}$ that works only on an input graph in the form G^k and produces an instance $R_{1,k}[G^k]$ of size almost linear in $|V(G^k)|$ while inequalities as in Theorem 2.1 still hold, and obtain the following.

Lemma 2.2. *For any k , there is a reduction $R_{1,k}$ that reduces a graph $G^k = G \cdot G \cdot \dots$ into an instance of the minimum consistent DFA problem of size $O(k \cdot |V(G^k)| \cdot |V(G)|^2)$ such that*

$$\chi(G)^k \leq \text{dfa}(R_{1,k}[G^k]) \leq \chi(G)^{2k} |V(G)|^4 \quad (6)$$

The description of reduction $R_{1,k}$ and the proof of Lemma 2.2 can be found in Section 5.2. Observe that the size $O(k \cdot |V(G^k)| \cdot |V(G)|^2)$ of $R_{1,k}(G^k)$ is almost linear (almost $O(|V(G^k)|)$)

⁵For conciseness, we will use g and f to refer to problems and their objective functions interchangeably.

as the extra $O(k|V(G)|^2)$ is negligible when k is sufficiently large. Similarly, the term $|V(G)|^4$ in Equation (6) is negligible and thus the value of $\text{dfa}(R_{1,k}[G^k])$ is sandwiched by $\chi(G)^k$ and $\chi(G)^{2k}$. This means that if $\chi(G)$ is small (i.e., $\chi(G) \leq |V(G)|^\epsilon$), then $\text{dfa}(R_{1,k}[G^k])$ will be small (i.e., $\text{dfa}(R_{1,k}[G^k]) \leq |V(G^k)|^{2\epsilon}$), and if $\chi(G)$ is large (i.e., $\chi(G) \geq |V(G)|^{1-\epsilon}$), then $\text{dfa}(R_{1,k}[G^k])$ will be also large (i.e., $\text{dfa}(R_{1,k}[G^k]) \geq |V(G^k)|^{1-\epsilon}$). The hardness of $n^{1-\epsilon}$ for DFA thus follows.

We note that in Theorem 2.1, we can replace DFA by NFA, a function corresponds to the minimum consistent NFA problem, thus getting a hardness of $n^{1/2-\epsilon}$ for this problem as well. This is, however, not yet tight. We would get a tight hardness if we can replace DFA by NFA in Lemma 2.2, which is not the case. We also note that the proof for the tight hardness for the minimum consistent CNF problem follows from the same type of inequalities: We show that there exists a near-linear-size reduction $R_{3,k}$ from the minimum coloring problem to the minimum consistent CNF problem (with function cnf) such that

$$\chi(G)^k \leq \text{cnf}(R_{3,k}[G^k]) \leq \chi(G)^k |V(G)|^{O(1)}. \quad (7)$$

The proofs of the bounds of graph products (Equations (3), (4), (6) and (7)) are fairly short and elementary; in fact, we believe that they can be given as textbook exercises. These proofs can be found in Sections 5 to 7.

2.4 Related Concept

Our pre-reduction graph product concept was inspired by the *graph product subadditivity* concept we previously introduced in [CLN13a] (some of these ideas were later used in [CLN13b, CLN14]). There, we prove a hardness of approximation using the following framework. As before, let f be an objective function of a problem that we intend to prove hardness and g be an objective function of a problem whose hardness is already known. We show that there are graph products \oplus , $*_e$, and $*$ such that

- We can “decompose” $f(G *_e J)$: $g(G) \leq f(G *_e J) \leq g(G) + f(G * J)$, and
- $f((G \oplus H) * J)$ is “subadditive”: $f((G \oplus H) * J) \leq f(G * J) + f(H * J)$.

We then use the above inequalities to show that if we let $G^k = G \oplus G \oplus \dots$ (k times), then

$$g(G^k) \leq f(G^k *_e J) \leq g(G^k) + kf(G * J).$$

For large enough k , the term $kf(G * J)$ is negligible and thus $f(G^k *_e J) \approx g(G^k)$. We use this fact to show that the approximation harness of f is roughly the same as the hardness of g . Observe that if we let $R[G] = G *_e J$, the above inequalities can then be used to show that

$$g(G \oplus H) \leq f(R[G \oplus H]) \leq g(G \oplus H) + f(R[G]) + f(R[H]).$$

In the problems considered in [CLN13a], one can easily bound $f(R[G])$ and $f(R[H])$ by $|V(G)|$ and $|V(H)|$, respectively. So, our meta theorem will imply that $f(G^k *_e J) \approx g(G^k)$, which leads to the approximation hardness of f . This means that the previous concept in [CLN13a] can be viewed as a special case of our new concept where we restrict the reduction R to be a graph product $R[G] = G *_e J$. The way we use the reduction R in this paper goes beyond this. For example, our reduction R_2 for EDP as illustrated in Figure 1 cannot be viewed as a natural graph product.

Moreover, our reduction R_1 reduces a graph G to an instance of DFA which has *nothing* to do with graphs. (This is possible only when we abandon viewing reduction R as a graph product.) Our meta theorem also shows that bounds of graph products in a much more general form can imply hardness results. Finally, the way we exploit graph products using the reduction $R_{1,k}$ has never appeared in [CLN13a].

2.5 Organization

After giving necessary definitions in Section 3, we prove meta theorems in Section 4. These theorems show that bounding $f(R[G * H])$ in a certain way will immediately imply a hardness result. They allow us to focus on proving appropriate bounds in later sections. In Section 5, we prove such bounds for the consistency problems and their implications to the hardness of proper PAC-learning. In Section 6, we prove such bounds of the edge-disjoint paths problem on DAGs. Bounds for other problems can be found in Section 7.

3 Preliminaries

3.1 Terms

Given two graph G and H , the *lexicographic product* of G and H , denoted by $G \cdot H$, is defined as

$$\begin{aligned} V(G \cdot H) &= V(G) \times V(H) = \{(u, v) : u \in V(G) \text{ and } v \in V(H)\}. \\ E(G \cdot H) &= \{(u, a)(v, b) : uv \in E(G) \text{ or } (u = v \text{ and } ab \in E(H))\}. \end{aligned}$$

Since the lexicographic product is the only graph product concerned in this paper, later on, we will simply use the term *graph product* to mean the lexicographic product. We define the k -fold graph product of G , denoted by G^k , as

$$G^k = G \cdot G^{k-1} \text{ for any integer } k > 1 \text{ and } G^1 = G$$

The properties of the lexicographic product that makes it becomes an import tools in proving hardness of approximation is that it multiplicatively increases the independent and chromatic numbers of graphs, without creating an overly dense resulting graph (the OR product also satisfies multiplicativity of independent and chromatic numbers, but it does not serve our purpose).

Theorem 3.1. *Let G and H be any graphs. The followings hold on $G \cdot H$.*

- $\alpha(G \cdot H) = \alpha(G)\alpha(H)$.
- $\frac{\chi(G)\chi(H)}{\log |V(G)|} \leq \chi(G \cdot H) \leq \chi(G) \cdot \chi(H)$.

In particular, for any $k \geq 1$, $\alpha(G^k) = \alpha(G)^k$ and $\chi(G)^{k-o(1)} \leq \chi(G^k) \leq \chi(G)^k$.

A *deterministic finite automaton* (DFA) is defined as a 5-tuple $(Q, \Sigma, \delta, q_0, F)$ where Q is the set of states, Σ is the set of alphabets, $\delta : Q \times \Sigma \rightarrow Q$ is a transition function, q_0 is initial state, and $F \subseteq Q$ is the set of accepting states. One can naturally extend the transition function δ into $\delta^* : Q \times \Sigma^* \rightarrow Q$ by inductively defining $\delta^*(q, x_1, \dots, x_\ell)$ as $\delta^*(\delta(q, x_1), x_2, \dots, x_\ell)$ and $\delta^*(q, null) = q$. We say that M *accepts* x if and only if $\delta^*(q_0, x) \in F$. The size of DFA M is

measured by the number of states of M , i.e., $|Q|$. We say that a DFA is *acyclic* if there is no state $q \in Q$ and string x such that $\delta^*(q, x) = q$. For NFA, the transition is defined by $\delta : Q \times \Sigma \rightarrow 2^Q$ instead, i.e., each transition possibly maps to several states. An NFA M accepts a string $x \in \Sigma^*$ if and only if the transition $\delta^*(q_0, x)$ contains an accepting state, i.e. $\delta^*(q_0, x) \cap F \neq \emptyset$.

3.2 Problems

In this section, we list all problems considered in this paper.

Minimum Consistency: In the MINIMUM CONSISTENCY problem, denoted by $\text{MINCON}(\mathcal{H}, \mathcal{F})$, we are given collections \mathcal{P} and \mathcal{N} of positive and negative sample strings in $\{0, 1\}^*$, for which we are guaranteed that there is a hypothesis $h \in \mathcal{H}$ that is consistent with all samples in $\mathcal{P} \cup \mathcal{N}$, i.e., $h(x) = 1$ for all $x \in \mathcal{P}$ and $h(x) = 0$ for all $x \in \mathcal{N}$. Our goal is to output a function $f \in \mathcal{F}$ that is consistent with all these samples, while minimizing $|f|$. In other words, \mathcal{H} and \mathcal{F} are the classes of the real hypothesis that we want to learn and those that our algorithm outputs respectively. This notion of learning allows our algorithm to output the hypothesis that is outside of the hypothesis class we want to learn.

Now we need a slightly modified notion of approximation factor. For any instance $(\mathcal{P}, \mathcal{N})$, we denote by $\text{OPT}_{\mathcal{H}}(\mathcal{P}, \mathcal{N})$ the size of the smallest hypothesis $h \in \mathcal{H}$ consistent with $(\mathcal{P}, \mathcal{N})$. Let \mathcal{A} be any algorithm for $\text{MINCON}(\mathcal{H}, \mathcal{F})$, i.e., \mathcal{A} always outputs the hypothesis in \mathcal{F} . The approximation guaranteed provided by \mathcal{A} is:

$$\sup_{\mathcal{P}, \mathcal{N}} \frac{|\mathcal{A}(\mathcal{P}, \mathcal{N})|}{\text{OPT}_{\mathcal{H}}(\mathcal{P}, \mathcal{N})}$$

With this terminology, the problem of learning DFA can be abbreviated as $\text{MINCON}(\text{DFA}, \text{DFA})$.

Edge Disjoint Paths: In the *edge-disjoint paths* (EDP) problem, given a graph $G = (V, E)$ and a set of source-sink pairs $\{(s_1, t_1), \dots, (s_k, t_k)\}$, our goal is to find a collection of paths $\mathcal{P} = \{P_{i_1}, P_{i_2}, \dots, P_{i_\ell} : i_j \in [k], P_{i_j} \text{ connects } s_{i_j} \text{ to } t_{i_j}\}$ that are edge disjoint while maximizing $|\mathcal{P}|$. That is, we want to connect as many source-sink pairs as possible using a collection of edge-disjoint paths.

Our focus is on the special case of EDP where G is a *directed acycle graph* (DAG).

Bounded-Length Edge-Disjoint Cycles: Given a graph $G = (V, E)$, the *cycle packing number* of G , denoted by $\nu(G)$, is the maximum integer ℓ such that there exist cycles C_1, \dots, C_ℓ which are pairwise edge-disjoint in G . The *edge-disjoint cycle* problem (EDC) asks to compute the value of $\nu(G)$. If we are additionally given an integer k , the *k-cycle packing number* of G , denoted by $\nu_k(G)$, is the maximum integer ℓ for which there exist pairwise edge-disjoint cycles C_1, \dots, C_ℓ where each cycle C_j contains at most k vertices. In the *k-edge-disjoint cycle* problem (*k*-EDC), we are asked to compute $\nu_k(G)$ given an input (G, k) .

Maximum Independent Set: Given a graph $G = (V, E)$, a subset of vertices $S \subseteq V$ is *independent* in G if and only if G has no edge joining any two vertices in S . The *independence number* of G , denoted by $\alpha(G)$, is the size of a largest independent set in G . In the *maximum independent set* problem, we are asked to compute an independent set S in G with maximum size.

The following is the hardness results of the maximum independent set problem by Håstad⁶, which will be used to obtain the hardness of EDP on DAGs.

Theorem 3.2 ([Hås96]+[Zuc07]). *Let $\epsilon > 0$ be any constant. Given graph $G = (V, E)$, it is NP-hard to distinguish between the following two cases:*

- (YES-INSTANCE:) $\alpha(G) \leq |V(G)|^\epsilon$
- (NO-INSTANCE:) $\alpha(G) \geq |V(G)|^{1-\epsilon}$

Chromatic Number: Given a graph $G = (V, E)$, a *proper coloring* $\sigma : V(G) \rightarrow [c]$ is a function that assigns colors to vertices of G so that any two adjacent vertices receive different colors assigned by σ (i.e., $uv \in E \implies \sigma(u) \neq \sigma(v)$). The *chromatic number* of G , denoted by $\chi(G)$, is the minimum integer c such that a proper coloring $\sigma : V(G) \rightarrow [c]$ exists, i.e., G can be properly colored by c colors. In the *graph coloring* problem, we are asked to compute a proper coloring $\sigma : V(G) \rightarrow [c]$ while minimizing c . We will be using the following hardness of approximation result by Feige and Kilian [FK98]⁶.

Theorem 3.3 ([FK98]+[Zuc07]). *Let $\epsilon > 0$ be any constant. Given graph $G = (V, E)$, it is NP-hard to distinguish between the following two cases:*

- (YES-INSTANCE:) $\chi(G) \leq |V(G)|^\epsilon$
- (NO-INSTANCE:) $\chi(G) \geq |V(G)|^{1-\epsilon}$

4 Meta Theorems

In this section, we prove general theorems that will be used in proving most hardness results in this paper. These theorems give *abstractions* of the (graph product) properties one needs to prove in order to obtain hardness of approximation results. Our techniques can be used to derive hardnesses for both minimization and maximization problems. For the former, the reduction is from minimum coloring, while the latter is obtained via a reduction from maximum independent set.

Let us start with maximization problems. Suppose we have an optimization problem Π such that any instance $I \in \Pi$ is associated with an optimal function $\text{OPT}_\Pi(I)$. We consider a transformation R that maps any graph G into an instance $R[G]$ of the problem Π . We say that a transformation R satisfies a *low α -projection property* with respect to a *maximization problem* Π if and only if the following two conditions hold:

- (I) For any graph $G = (V, E)$, $\text{OPT}_\Pi(R[G]) \geq \alpha(G)$.
- (II) There are universal constants $c_1, c_2 > 0$ (independent of the choices of graphs) such that, for any two graphs G and H ,

$$\text{OPT}_\Pi(R[G \cdot H]) \leq |V(G)|^{c_1} + \alpha(G)^{c_2} \text{OPT}_\Pi(R[H]).$$

⁶ The hardness results of the maximum independent set problem [Hås96] and the graph coloring problem [FK98] hold under the assumption $\text{NP} \neq \text{ZPP}$. The results were later derandomized by Zuckerman in [Zuc07] and thus hold under the assumption $\text{P} \neq \text{NP}$.

- (III) There is a universal constant $c_0 > 0$ such that

$$\text{OPT}_{\Pi}(R[G]) \leq c_0 |R[G]|.$$

Intuitively, the transformation R with the low α -projection property tells us that there are relationships between the optimal solution of the problem Π on $R[G]$ and the independence number of G . Instead of looking for a sophisticated construction of R , we focus on a “simple” transformation R that establishes a connection on one side, i.e., $\text{OPT}_{\Pi}(R[G]) \geq \alpha(G)$, and the “growth” of OPT_{Π} is “slow” with respect to graph products. Property (III) of the low α -projection property says that the optimal is at most linear in the size of the instance, which is the case for almost every natural combinatorial optimization problem.

Next, we turn our focus to a minimization problem. In this case, we relate the optimal solution to the chromatic number of an input graph. Specifically, one can define the *low χ -projection property* with respect to a *minimization problem* Π as follows.

- (I) For any graph $G = (V, E)$, $\text{OPT}_{\Pi}(R[G]) \geq \chi(G)$.
- (II) There are universal constants $c_1, c_2 > 0$ (independent of the choices of graphs) such that, for any two graphs G and H , we have

$$\text{OPT}_{\Pi}(R[G \cdot H]) \leq |V(G)|^{c_1} + \chi(G)^{c_2} \text{OPT}_{\Pi}(R[H]).$$

- (III) There is a universal constant $c_0 > 0$ such that

$$\text{OPT}_{\Pi}(R[G]) \leq c_0 |R[G]|.$$

We observe that the existence of such reductions is sufficient for establishing hardness of approximation results, and the hardness factors achievable from the theorems depend on the size of the reduction.

Theorem 4.1 (Meta-Theorem for Maximization Problems). *Let Π be a maximization problem for which there is a reduction R for Π that satisfies low α -projection property with $|R[G]| = O(|V(G)|^d)$. Then for any $\epsilon > 0$, given an instance I of Π , it is NP-hard to distinguish between the following two cases:*

- (YES-INSTANCE:) $\text{OPT}_{\Pi}(I) \geq |I|^{1/d-\epsilon}$
- (NO-INSTANCE:) $\text{OPT}_{\Pi}(I) \leq |I|^{\epsilon}$

Theorem 4.2 (Meta-Theorem for Minimization Problems). *Let Π be a minimization problem for which there is a reduction R for Π that satisfies low χ -projection property with $|R[G]| = O(|V(G)|^d)$, for some constant $d \geq 0$. Then for any $\epsilon > 0$, given an instance I of Π , it is NP-hard to distinguish between the following two cases:*

- (YES-INSTANCE:) $\text{OPT}_{\Pi}(I) \leq |I|^{\epsilon}$
- (NO-INSTANCE:) $\text{OPT}_{\Pi}(I) \geq |I|^{1/d-\epsilon}$

4.1 Proof of Theorem 4.1 (Meta Theorem for Maximization Problems)

Consider a reduction R that transforms a graph G into an instance of Π that satisfies the low α -projection property. We analyze how the optimal value changes over ℓ -fold lexicographic products.

Lemma 4.3. *For any positive integer ℓ , $\text{OPT}_{\Pi}(R[G^{\ell}]) \leq \ell c_0 |V(G)|^{c_1+d+1} \alpha(G)^{2c_2\ell}$*

Proof. This is proved by induction on a positive integer ℓ . The base case $\ell = 1$ holds because $\text{OPT}_{\Pi}(R[G]) \leq c_0 |R[G]| \leq c_0 |V(G)|^d$. Assume that the induction hypothesis holds for any $\ell > 1$, and consider $\text{OPT}_{\Pi}(R[G^{\ell+1}])$. By writing $G^{\ell+1} = G \cdot G^{\ell}$ and applying the low α -projection property, we have

$$\text{OPT}_{\Pi}(R[G^{\ell+1}]) \leq |V(G)|^{c_1} + \alpha(G)^{c_2} \text{OPT}_{\Pi}(R[G^{\ell}])$$

Then, by applying induction hypothesis, we have

$$\begin{aligned} \text{OPT}_{\Pi}(R[G^{\ell+1}]) &\leq |V(G)|^{c_1} + \alpha(G)^{c_2} \left(\ell c_0 |V(G)|^{c_1+d+1} \alpha(G)^{2c_2\ell} \right) \\ &\leq |V(G)|^{c_1} + \alpha(G)^{c_2+2c_2\ell} \ell c_0 |V(G)|^{c_1+d+1} \\ &\leq (\ell + 1) c_0 |V(G)|^{c_1+d+1} \alpha(G)^{2c_2(\ell+1)} \end{aligned}$$

□

We note that the exponent of the term $\alpha(G)$ depends on ℓ (the number of times the product is applied), while that of $|V(G)|$ does not. Intuitively speaking, this is why the contribution of the term $|V(G)|^{c_1}$ vanishes after taking graph products.

Hardness of Approximation. Now we prove the hardness of approximation result claimed in Theorem 4.1. Start from graph G as given by Theorem 3.2. Then construct an instance $R[G^{\ell}]$ with $\ell = \lceil 1/\epsilon \rceil$. This results in the instance $R[G^{\ell}]$ of the problem Π of size $N = |R[G^{\ell}]| = O(|V(G)|^{\ell d})$.

In the YES-INSTANCE, we have

$$\text{OPT}_{\Pi}(R[G^{\ell}]) \geq \alpha(G^{\ell}) = \alpha(G)^{\ell} \geq |V(G)|^{(1-\epsilon)\ell} = N^{1/d-O(\epsilon)}.$$

In the NO-INSTANCE, we have

$$\text{OPT}_{\Pi}(R[G^{\ell}]) \leq O(|V(G)|^{d+c_1+1} \alpha(G)^{2c_2\ell}).$$

Since $\alpha(G) \leq |V(G)|^{\epsilon}$ in this case, we have

$$\alpha(G)^{2c_2\ell} \leq |V(G)|^{2c_2\epsilon} = |V(G)|^{O(1)} = N^{O(\epsilon)}.$$

This implies that $\text{OPT}_{\Pi}(R[G^{\ell}]) \leq |V(G)|^{O(1)} N^{O(\epsilon)} = N^{O(\epsilon)}$, and the gap between YES-INSTANCE and NO-INSTANCE is $N^{1/d-O(\epsilon)}$. This completes the proof.

4.2 Proof of Theorem 4.2 (Meta Theorem for Minimization Problems)

Similarly to the case of maximization problems, we can prove the following lemma by induction on integers ℓ . We shall skip the proof as it is the same as that of Lemma 4.3 except that $\alpha(G)$ is replaced by $\chi(G)$.

Lemma 4.4. *For any positive integer ℓ , $\text{OPT}_{\Pi}(R[G^{\ell}]) \leq \ell c_0 |V(G)|^{c_1+d+1} \chi(G)^{2c_2\ell}$.*

Hardness of Approximation Take the instance $R[G^\ell]$ with $\ell = \lceil 1/\epsilon \rceil$.

In the YES-INSTANCE, we have the following bound, which is slightly different from the case of the maximization problem.

$$\text{OPT}_{\Pi}(R[G^\ell]) \geq \chi(G^\ell) \geq (\chi_f(G))^\ell \geq |V(G)|^{\ell(1-2\epsilon)} = N^{1/d-O(\epsilon)}.$$

In the NO-INSTANCE, Lemma 4.4 gives $\text{OPT}_{\Pi}(R[G^\ell]) \leq N^{O(\epsilon)}$. Thus, we have the desired gap, completing the proof.

4.3 Overview of Applications

Most of the reductions in this paper are direct applications of the above two meta theorems. That is, we design the following reductions.

- A reduction R_{EDP} for EDP such that $|R_{\text{edp}}[G]| = O(|V(G)|^2)$ and satisfies α -projection property. This implies a tight $n^{1/2-\epsilon}$ hardness of approximating EDP on DAGs.
- A reduction R_{fa} for MinCon such that $|R_{fa}[G]| = O(|V(G)|^2)$ and satisfies χ -projection property. This gives $n^{1/2-\epsilon}$ hardness of approximating MinCon(NFA,ADFA).

Notice that the reduction R_{fa} above is not tight. To obtain a tight result, we need $|R_{fa}[G]| = O(|V(G)|)$, and it seems difficult to obtain such a reduction. We instead exploit the further structure of graph products and prove bounds of the form

$$\chi(G^k) \leq \text{OPT}(R'_{fa}[G^k]) \leq \chi(G)^{O(k)} |V(G)|^{O(1)}$$

Now our reduction size is smaller, i.e., $|R'_{fa}[G^k]| = |V(G)|^{(1+o(1))k}$ as opposed to $|R_{fa}[G^k]| = |V(G)|^{2k}$. Moreover, the reduction R'_{fa} exploits the fact that the input graph is written as a k -fold product of graphs. This more restricted form of graph products allows us to prove tight hardness (and PAC impossibility result) of DFA and DNF/CNF Minimization.

5 Hardnesses of Finite Automata Problems: Minimum Consistency and Proper PAC Learning

We show in this section the hardness of the consistency problems for finite automata, as well as the implications on impossibility results for PAC learning. We start our discussion by proving the hardness for MINCON(ADFA,NFA), which includes the minimum consistent NFA problem (MINCON(NFA,NFA)) as a special case. Then we proceed to prove the tight hardness of approximating MINCON(ADFA,DFA), which implies the tight hardness of approximating the minimum consistent DFA problem and also implies the impossibility result for proper PAC-learning DFA.

5.1 Hardness of MINCON(ADFA, NFA) via Graph Products

In this section, we show an $N^{1/2-\epsilon}$ hardness for MINCON(ADFA,NFA). Formally, we prove the following theorem.

Theorem 5.1. *Let $\epsilon > 0$ be any positive constant. Given two sets of positive and negative sample strings \mathcal{P}, \mathcal{N} over alphabet $\Sigma = \{0, 1\}$ with a total length of N bits, it is NP-hard to distinguish the following two cases:*

- *There is an acyclic deterministic finite automata of size N^ϵ that is consistent with all strings in $\mathcal{P} \cup \mathcal{N}$.*
- *Any non-deterministic finite automata consistent with $\mathcal{P} \cup \mathcal{N}$ must have at least $N^{1/2-\epsilon}$ states.*

This is done by designing a reduction $R[G]$ with χ -projection property and $|R[G]| = O(|V(G)|^2)$. Our proof in fact shows that the projection properties hold for both optimal DFA and NFA functions.

5.1.1 The Reduction R

We will be working with binary strings, i.e., the alphabet set $\Sigma = \{0, 1\}$. Given a graph $G = (V, E)$, we construct two sets \mathcal{P}, \mathcal{N} of positive and negative samples, which encode vertices and edges of the graph. We assume w.l.o.g. that $|V(G)| = 2^k$ for some integer k . Therefore, each vertex $u \in V(G)$ can be associated with a k -bit string $\langle u \rangle \in \{0, 1\}^k$.

Now our reduction $R[G]$ is defined as follows. The positive samples are given by

$$\mathcal{P} = \{\langle u \rangle 1 \langle u \rangle^R : u \in V(G)\}$$

and the negative samples are

$$\mathcal{N} = \{\langle u \rangle 1 \langle v \rangle^R : uv \in E(G)\}$$

We denote this instance of the consistency problem by an ordered pair $(\mathcal{P}, \mathcal{N})$. Now we proceed to prove property (I), that any NFA consistent with $(\mathcal{P}, \mathcal{N})$ must have at least $\chi(G)$ states.

Lemma 5.2. *Let $M = (Q, \Sigma, \delta, q_0, F)$ be an NFA that is consistent with $(\mathcal{P}, \mathcal{N})$. Then for any vertex $u \in V(G)$,*

$$\delta^*(q_0, \langle u \rangle) \not\subseteq \bigcup_{v: uv \in E(G)} \delta^*(q_0, \langle v \rangle).$$

Proof. Assume for contradiction that $\delta^*(q_0, \langle u \rangle) \subseteq \bigcup_{v: uv \in E(G)} \delta^*(q_0, \langle v \rangle)$. Since $\langle u \rangle 1 \langle u \rangle^R$ is a positive sample, there is a state $q \in \delta^*(q_0, \langle u \rangle)$ that leads to an accepting state (i.e., $\delta^*(q, 1 \langle u \rangle^R) \cap F \neq \emptyset$). By the assumption, the state q also belongs to another set $\delta^*(q_0, \langle v \rangle)$ for some $v : uv \in E(G)$.

Now consider the string $\langle v \rangle 1 \langle u \rangle^R$, which is a negative sample because $vu \in E(G)$. Since $q \in \delta^*(q_0, \langle v \rangle)$ and $\delta^*(q, 1 \langle u \rangle^R) \cap F \neq \emptyset$, the string $\langle v \rangle 1 \langle u \rangle^R$ must be accepted by M , a contradiction. \square

Lemma 5.2 implies in particular that, for each vertex $u \in V(G)$, the set $\delta^*(q_0, \langle u \rangle) \setminus \left(\bigcup_{v: uv \in E(G)} \delta^*(q_0, \langle v \rangle) \right)$ is not empty. Now denote by $\text{OPT}_{DFA}(R[G])$ and $\text{OPT}_{NFA}(R[G])$ the number of states in the minimum DFA and NFA that are consistent with the samples $R[G] = (\mathcal{P}, \mathcal{N})_G$, respectively.

Corollary 5.3. *Any NFA M that is consistent with $(\mathcal{P}, \mathcal{N})_G$ must have at least $\chi(G)$ states. Therefore, $\text{OPT}_{DFA}(R[G]) \geq \text{OPT}_{NFA}(R[G]) \geq \chi(G)$ for all G .*

Proof. For each state $q \in Q$, define a set $C_q = \left\{ u \in V(G) : q \in \delta^*(q_0, \langle u \rangle) \setminus \left(\bigcup_{v: uv \in E(G)} \delta^*(q_0, \langle v \rangle) \right) \right\}$. It is easy to see that C_q is an independent set and thus form a proper color class of G . Lemma 5.2 implies that each vertex $u \in V(G)$ belongs to at least one class. So, $\{C_q\}_{q \in Q}$ gives a proper $|Q|$ -coloring of G , implying that $|Q| \geq \chi(G)$. \square

5.1.2 χ -Projection Property

We will consider a specific class of DFA $M = (Q, \Sigma, \delta, q_0, F)$, which we call *canonical DFA*. Specifically, we say that a DFA is *canonical* if it has the following properties.

- The state diagram has exactly ℓ layers for some ℓ , and each path from q_0 to any sink has length exactly ℓ .
- All accepting states are in the last layer.

Denote shortly by $\text{OPT}(R[G])$ the number of states in the minimum canonical DFA consistent with $R[G]$. So we have that $\text{OPT}(R[G]) \geq \text{OPT}_{DFA}(R[G]) \geq \text{OPT}_{NFA}(R[G])$. The following lemma gives the χ -projection property for $\text{OPT}(\cdot)$

Lemma 5.4. $\text{OPT}(R[G \cdot H]) \leq \chi(G)(\text{OPT}(R[H]) + O(|V(G)|))$

To prove this lemma, we show how to construct, given a canonical DFA for $R[H]$, a “compact” canonical DFA for $R[G \cdot H]$. We note that one key idea here is to avoid exploiting the DFA for $R[G]$ but instead tries to use the color classes of G in its optimal coloring to “compress” the DFA for $R[G \cdot H]$.

Proof. Let $M_H = (Q_H, \{0, 1\}, \delta_H, q_H, F_H)$ be the minimum DFA for the instance $R[H]$ whose number of states is $s = \text{OPT}(R[H])$ and has $\ell_H = 2h + 1$ layers for $h = \lceil \log |V(H)| \rceil$. Let C_1, \dots, C_B be the color classes of G defined by the optimal coloring, so $B = \chi(G)$. Let $f : V(G) \rightarrow [B]$ be the corresponding coloring function. We will also be using several copies of a directed complete binary tree with 2^k leaves, where each leaf corresponds to a string in $\{0, 1\}^k$ and is associated with a vertex in $V(G)$. Call this directed binary tree T_k .

We will use M_H and T_k to construct a new acyclic DFA M that have at most $B(s + O(|V(G)|))$ states and exactly $\ell = 2(k + \ell_H) + 1$ layers. Now we proceed with the description of machine $M = (Q, \{0, 1\}, \delta, q, F)$. We start by taking a copy of directed tree T_k , and call this copy $T_k^{(0)}$. The starting state q is defined to be the root of $T_k^{(0)}$. This is the *first phase* of the construction. Notice that there are k layers in the first phase, so exactly k positions of any input string will be read after this phase. Each state in the last layer is indexed by $state(\langle v \rangle)$ for each $v \in V(G)$.

In the *second phase*, we take B copies of the machines M_H where the j^{th} copy, denoted by $M_H^{(j)} = (Q_H^{(j)}, \{0, 1\}, \delta_H^{(j)}, q_H^{(j)}, F_H^{(j)})$, is associated with color class C_j defined earlier. For each vertex $v \in V(G)$, we connect the corresponding state $state(\langle v \rangle)$ in the last layer of Phase 1 to the starting state $q_H^{(f(v))}$. This transition can be thought of as a “null” transition which can be removed afterward, but keeping it this way would make the analysis simpler. Since each copy of M_H has $2\ell_H + 1$ layers, now our construction has exactly $2\ell_H + k + 1$ layers.

In the final phase, we first extend all rejecting states in $M_H^{(j)}$ by a unified path until it reaches layer $2(\ell_H + k) + 1$. This is a rejecting state rej_0 . Now, for each $j = 1, \dots, B$, we connect each accepting state in the last layer of $M_H^{(j)}$ to the root in the copy $T_k^{(j)}$ again by a “null” transition, so we reach the desired number of layers now (notice that each root-to-leaf path has $2(k + \ell_H) + 1$ states.) The states in the last layer of $T_k^{(j)}$ are indexed by $state(j, \langle v \rangle)$. The accepting states of M are defined as $F = \bigcup_{j=1}^B \{state(j, \langle u \rangle^R) : u \in C_j\}$, and the rest of the states are defined as rejecting. This completes our construction. See Figure 2 for illustration.

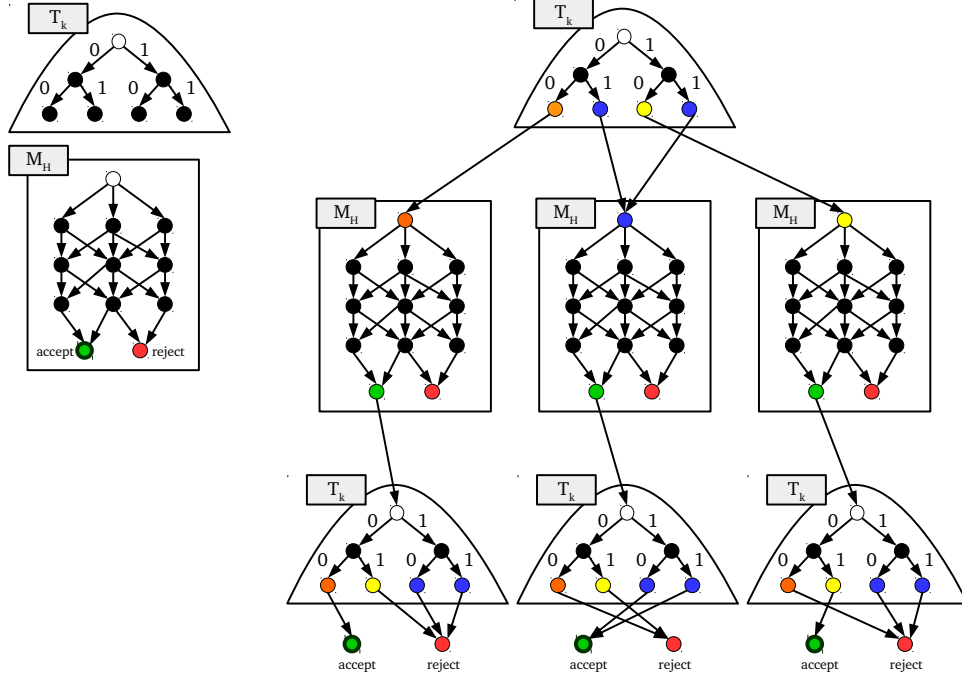


Figure 2: The illustration of the construction in the proof of Lemma 5.4.

The size of the construction is $|V(G)| + Bs + O(|V(G)|B) = B(s + O(|V(G)|))$. The next claim shows that the machine M is consistent with samples obtained from the product of G and H , which thus finish the proof.

Claim 5.5. *Given a machine M_H that is consistent with samples $R[H]$, the machine M constructed as above is consistent with samples $R[G \cdot H]$.*

Proof. First we check the positive sample. For each vertex $(u, a) \in V(G \cdot H)$, the corresponding string $\langle(u, a)\rangle 1 \langle(u, a)\rangle^R$ can be thought of as $x = \langle u \rangle \langle a \rangle 1 \langle a \rangle^R \langle u \rangle^R$. After the first k transitions, the machine M will stop at the state $q_H^{(f(u))}$. Then the substring $\langle a \rangle 1 \langle a \rangle^R$ will lead to an accepting state in $F_H^{(f(u))}$ (since M_H is consistent with samples in $R[H]$). Now, at the current state, we are at the root of the tree $T_k^{(f(u))}$, and we are left with the substring $\langle u \rangle^R$. Since $u \in C_{f(u)}$, the substring $\langle u \rangle^R$ leads to an accepting state. This proves that the machine M always accepts positive samples.

Next, consider a negative sample $\langle(u, a)\rangle 1 \langle(v, b)\rangle^R$ generated by the edge $(u, a)(v, b) \in E(G)$. Again, this can be thought of as $\langle u \rangle \langle a \rangle 1 \langle b \rangle^R \langle v \rangle^R$. There are two possible cases:

- If $uv \in E(G)$, then the machine will enter $M_H^{(f(u))}$ after reading the substring $\langle u \rangle$. Next, the machine reads the substring $\langle a \rangle 1 \langle b \rangle^R$. If it manages to reach the third phase without rejection (i.e., M_H accepts $\langle a \rangle 1 \langle b \rangle^R$), then it will enter the tree $T_k^{(f(u))}$. Note that there is no edge joining two vertices in $C_{f(u)}$ because it is a color class. Thus, the substring $\langle v \rangle$ leads to a rejection because $uv \in E(G)$ implies that $v \notin C_{f(u)}$.

(Notice that the rejection does not depend on what happens inside M_H .)

- If $u = v$ and $ab \in E(G)$, then after the first k transitions, the machine enters $M_H^{(f(u))}$ with the input string $\langle a \rangle 1 \langle b \rangle^R$ for $ab \in E(H)$. Since M_H is consistent with samples in $R[H]$, this would lead to a rejection in $M_H^{(f(u))}$ and therefore in M .

□

□

We will also need the base case condition as required by the low α -projection property.

Lemma 5.6. *For any graph G , $\text{OPT}(R[G]) \leq O(|V(G)|^2 \log |V(G)|)$*

Proof. We simply use the tree T_{2k+1} with the initial state q at the root of T_{2k+1} , where each vertex at the leaf can be associated with a string in $\{0, 1\}^{2k+1}$. We simply define the accepting states to be those that correspond to the strings of the form $\langle u \rangle 1 \langle u \rangle^R$. The size of the construction is $2^{2k+1} = O(|V(G)|^2 \log |V(G)|)$. □

5.2 Tight Hardness for MINCON(ADFA, DFA)

Notice that the construction in the previous section is not tight because the size of the negative samples in $R[G]$ is large compared to the number of vertices in graph G , i.e., $|\mathcal{N}| = \Theta(|V(G)|^2)$. To handle this problem, we take into account the structure of the lexicographic product and “encode” negative samples in a more compact form, i.e. we ideally want the construction size to be nearly linear on $|V(G)|$, i.e., $|\mathcal{N}| = O(|V(G)|^{1+o(1)})$, instead of quadratic.

To this end, we construct a reduction $R_k[G^k]$. We remark that, while the construction in this section gives tighter results for DFA, ADFA, and OBDD, it does not apply to NFA.

5.2.1 The Reduction $R_k[G^k]$

We show a reduction $R_k[G^k]$ of size $|R_k[G^k]| = |V(G)|^{k(1+o(1))}$. Consider a graph $H = G^k$ (the k -fold lexicographic product of G). We will encode the edge structures of H into the positive and negative samples as follows.

Positive Samples: For each $\vec{u} = (u_1, \dots, u_k) \in V(H)$, define a positive sample

$$\text{pos}(\vec{u}, i) = \langle u_1 \rangle \dots \langle u_k \rangle 1 \langle u_1 \rangle \dots \langle u_i \rangle.$$

The set of all positive samples is denoted by

$$\mathcal{P} = \{\text{pos}(\vec{u}, i) : u \in V(H), i = 1 \dots k\}$$

Negative Samples: For each a pair of vertices $\vec{u} \in V(H)$ and $v \in V(G)$ such that $u_i v \in E(G)$, define a negative sample

$$\text{neg}(\vec{u}, v, i) = \langle u_1 \rangle \dots \langle u_k \rangle 1 \langle u_1 \rangle \dots \langle u_{i-1} \rangle \langle v_i \rangle$$

The set of all negative samples is denoted by

$$\mathcal{N} = \{\text{neg}(\vec{u}, v, i) : \vec{u} \in V(H), v \in V(G), u_i v \in E(G), i = 1, \dots, k\}$$

Intuitively, an edge in the the input graph represents a *conflict* between two vertices. Negative samples are thus defined to capture a conflict (an edge) in the product of graphs between vertices \vec{u} and \vec{v} at coordinate i . Notice that the size of positive and negative samples are $|\mathcal{P}| = kn^k$ and $|\mathcal{N}| = kn^k|E(G)| \leq kn^{k+2}$.

Let $\text{OPT}_{ADFA}(\cdot)$ denote the number of states in the optimal acyclic DFA that is consistent with the samples. We will prove the following lemma.

Lemma 5.7. $\chi(H) \leq \text{OPT}_{DFA}(\mathcal{P}, \mathcal{N}) \leq \text{OPT}_{ADFA}(\mathcal{P}, \mathcal{N}) \leq \chi(G)^{2k}|V(G)|^4$.

The bound $\text{OPT}_{DFA}(\mathcal{P}, \mathcal{N}) \leq \text{OPT}_{ADFA}(\mathcal{P}, \mathcal{N})$ is trivial. For the other bounds, we will prove the left and right-hand side inequalities of Lemma 5.7 in Section 5.2.2 and Section 5.2.3, respectively. The hardness result then follows trivially from Theorem 3.3 and Theorem 3.1. In particular, taking the hard instance of the graph coloring problem as in Theorem 3.3, we have that

$$\text{YES-INSTANCE: } \text{OPT}_{DFA}(\mathcal{P}, \mathcal{N}) \leq \chi(G)^{2k}n^4 \leq n^{O(1)}.$$

$$\text{NO-INSTANCE: } \text{OPT}_{DFA}(\mathcal{P}, \mathcal{N}) \geq \chi(H) \geq \chi(G)^{k-o(1)} \geq n^{(k-O(1))}.$$

Since $|\mathcal{P}| + |\mathcal{N}| = O(kn^{k+2})$, this implies the hardness gap of $(|\mathcal{P}| + |\mathcal{N}|)^{1-\varepsilon}$, for any $\varepsilon > 0$.

5.2.2 The Lower Bound of OPT_{DFA}

First, we show the lower bound for $\text{OPT}_{DFA}(\mathcal{P}, \mathcal{N})$. Let $M = (Q, \Sigma, \delta, q_0, F)$ be a DFA consistent with $(\mathcal{P}, \mathcal{N})$. We construct from M a $|Q|$ -coloring of H : For each state $q \in Q$, we define a color class $C_q = \{\vec{u} : \delta^*(q_0, \langle \vec{u} \rangle) = q\}$. Since M is deterministic, each vertex must get at least one color.

Lemma 5.8. *For any vertices $\vec{u}, \vec{v} \in C_q$, $\vec{u}\vec{v} \notin E(H)$. That is, C_q is a proper color class of H .*

Proof. Suppose to a contrary that there is a pair of vertices $\vec{u}, \vec{v} \in C_q$ such that $\vec{u}\vec{v} \in E(H)$. Since H is obtained by the lexicographic product, there exists a coordinate i in which \vec{u} and \vec{v} conflict, i.e., $u_j = v_j$ for all $j < i$ and $u_i v_i \in E(G)$. We know that $\delta^*(q, 1\langle u_1 \rangle \dots \langle u_i \rangle) \in F$ because $\text{pos}(\vec{u}, i) = \langle \vec{u} \rangle 1\langle u_1 \rangle \dots \langle u_i \rangle$ is a positive sample. Since $\delta^*(q_0, \langle \vec{u} \rangle) = \delta^*(q_0, \langle \vec{v} \rangle) = q$, we must also have $\delta^*(q_0, \langle \vec{v} \rangle 1\langle u_1 \rangle \dots \langle u_i \rangle) = \delta^*(q, 1\langle u_1 \rangle \dots \langle u_i \rangle) \in F$. But, this contradicts the fact that $\text{neg}(\vec{v}, \vec{u}, i) = \langle \vec{v} \rangle 1\langle v_1 \rangle \dots \langle v_{i-1} \rangle \langle u_i \rangle = \langle \vec{v} \rangle 1\langle u_1 \rangle \dots \langle u_i \rangle$ is a negative sample. \square

5.2.3 The Upper bound of OPT_{ADFA}

Now we need to argue that there is an acyclic DFA M of size $\chi(G)^{2k}n^4$. Suppose $V(G) = \{0, 1\}^\ell$. Let $c = \chi(G)$, and $\sigma : V(G) \rightarrow [c]$ be an optimal coloring of G . Our construction has two steps. First, we construct a complete rooted c -ary tree with $2k$ level, namely S . Note that S is a directed tree whose edges are oriented toward leaves. Each vertex in S except the root is associated with one color class from σ . In particular, for each internal vertex a of S , each child x of a is associated with a distinct color from $[c]$. We define the coloring of S by $\rho : V(S) \rightarrow [c]$. Second, we replace each vertex a of S by a complete binary tree T with n leaves; we denote this copy of T by T_a . Each leaf q of T_a is associated with a vertex u of G and thus has a color $\sigma(u)$ assigned. (We abuse $\sigma(q) = \sigma(u)$ to mean a color of q .) For any vertex x in S that is a child of a , we join every leaf q of T_a with color $\sigma(q) = \rho(a)$ to the root r of T_x . The transition edge qr is a null transition unless a is a vertex at level k in S ; for the case that a is at level k , the transition edge qr is labeled “1”. (Note that a null transition edge qr means that we will merge q and r in the final construction. It is easy to see that this results in a DFA (not NFA) because S is a tree.) It can be seen that the

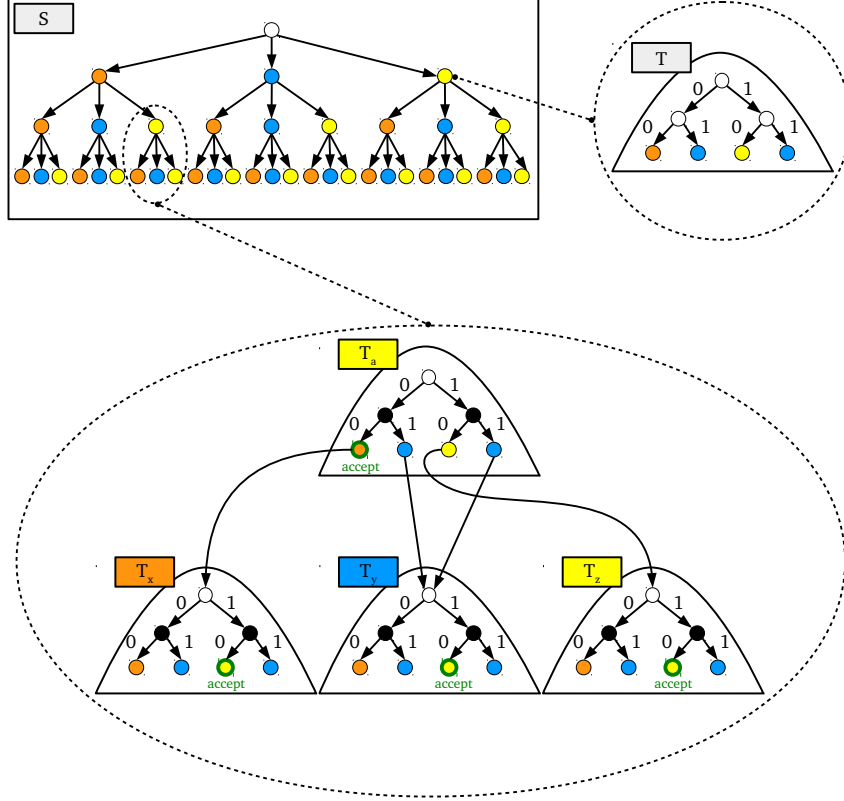


Figure 3: The illustration of the construction in the proof of Lemma 5.7.

constructed directed graph has a single source vertex (i.e., a vertex with no incoming edges), which we define as a starting state q_0 .

To finish the construction, we define accepting states. Let a_0 be the root of S . Consider a vertex a_i at level $i > k$ in S and its corresponding tree T_{a_i} . Since S is a tree, there is a unique path from a_1 to a_i , namely, $P = a_0, \dots, a_k, a_{k+1}, \dots, a_i$. For each leaf q of T_{a_i} , we define q as an accepting state if and only if $\sigma(q) = \rho(a_{i-k+1})$, i.e., q and a_{i-k+1} receive the same color. See Figure 3 for illustration.

Each copy of T has at most $2n$ vertices, and S has at most c^{2k} vertices. Thus, the size of the DFA M is at most $2nc^{2k}$. Also, observe that M is acyclic.

Lemma 5.9. *The DFA M is consistent with $(\mathcal{P}, \mathcal{N})$.*

Proof. Consider any sample $\vec{u} \in \mathcal{P} \cup \mathcal{N}$, which must be of the form:

$$\vec{u} = \langle u_1 \rangle \dots \langle u_k \rangle 1 \langle u_{k+1} \rangle \dots \langle u_{k+i} \rangle \quad \text{for } i : 1 \leq i \leq k \text{ and } u_j \in V(S) \text{ for all } j = 1, \dots, k+i.$$

Note that $u_j = u_{k+j}$ for all $j = 1, 2, \dots, i-1$. The transition $\delta^*(q_0, \vec{u})$ forms a path P in M , which traverses from the starting state q_0 to some state q_j . (That is, $q_j = \delta^*(q_0, \vec{u})$.) By construction, P corresponds to the path a_1, \dots, a_i in S and thus must visit a leaf q_j of tree T_{a_j} , for $j = 1, \dots, i$. Moreover, each q_j is associated with vertex $u_j \in V(G)$. Notice that $\rho(a_j) = \sigma(u_{j-1})$ because we have an edge from q_{j-1} to T_{a_j} if and only if q_{j-1} and a_j receive the same color.

If \vec{u} is a positive sample in \mathcal{P} , then we have $u_{i-k} = u_i$. (Note that q_j and u_j receive the same color for all $j = 1, 2, \dots, k$.) Since a_{i-k+1} has the same color as q_{i-k} (and so does $u_{i-k} = u_i$), we have $\rho(a_{i-k+1}) = \sigma(q_i)$. Thus, q_i is an accepting state.

If \vec{u} is a negative sample in \mathcal{N} , then we must have an edge $u_{i-k}u_i \in E(G)$. So, u_{i-k} and u_i receive different colors. Since $\rho(a_{i-k+1}) = \sigma(u_{i-k})$, it follows that $\rho(a_{i-k+1}) \neq \sigma(q_i)$. Thus, q_i is not an accepting state. This proves that M is consistent with both positive and negative samples. \square

5.3 Hardness of Proper PAC-Learning

Here we show that DFAs are not PAC-learnable. That is, we prove Theorem 1.2. We will use the connection between PAC learning and the existence of an *Occam algorithm*, defined as follows.

Definition 5.10. *An Occam algorithm for a hypothesis class \mathcal{H} in terms of function classes \mathcal{F} is an algorithm \mathcal{A} that for some constant $k \geq 0$ and $\alpha < 1$, the following guarantee holds. Let $h \in \mathcal{H}$ has size n and represents some language $L(h)$. Then on any input of s samples of $L(r)$, each of length at most m , the algorithm \mathcal{A} outputs an element $h \in \mathcal{H}$ of size at most $n^k m^k s^\alpha$ that is consistent with each of the s samples.*

Therefore, an Occam algorithm for DFA is the case when $\mathcal{H} = \text{DFA} = \mathcal{F}$, and the measure of the size of each hypothesis $h \in \text{DFA}$ is the number of states. It is known that PAC learnability of DFA implies the existence of an Occam algorithm for the same hypothesis class as stated formally in the following theorem.

Theorem 5.11 ([BP92], statement from [Pit89]). *If DFAs are properly PAC-learnable, then there exists a randomized Occam algorithm for DFA that runs in polynomial time.*

Theorem 5.11 implies that, to prove Theorem 1.2, it suffices to rule out the existence of a randomized Occam algorithm for DFA, which is shown in the next Theorem.

Theorem 5.12. *Unless $\text{NP} = \text{RP}$, there is no polynomial time randomized Occam algorithm for DFA.*

Proof. We prove by contrapositive. Assume that there is a randomized Occam algorithm \mathcal{A} for DFA with parameters (k, α) for some constants $k \geq 0$ and $0 \leq \alpha < 1$. Then we argue that there would exist an algorithm that distinguishes between the YES-INSTANCE and NO-INSTANCE given in Theorem 1.1. To see this, take an instance of $\text{MinCon}(\text{ADFA}, \text{DFA})$ as in Theorem 1.1. So, we have a pair of sets (P, N) of N samples, each of length $O(\log N)$. The parameters of the Occam algorithm \mathcal{A} are thus $s = N$ and $m = O(\log N)$.

We choose the parameter ϵ in Theorem 1.1 to be $\epsilon = (1 - \alpha)/(2k + 1)$.

In the YES-INSTANCE, there is a DFA of size N^ϵ consistent with the samples. Thus, our hypothesis class has size $n = N^\epsilon$. By definition, the Occam algorithm \mathcal{A} gives us a DFA M of size

$$|M| \leq N^{\epsilon k} \cdot (\log N)^k \cdot N^\alpha \leq N^{\alpha + (2k)\epsilon} \leq N^{\alpha + (2k)\frac{1-\alpha}{2k+1}} = N^{1 - \frac{1-\alpha}{2k+1}} = N^{1-\epsilon}$$

In the NO-INSTANCE, any DFA M consistent with (P, N) has size $|M| > N^{1-\epsilon}$.

Therefore, the randomized Occam algorithm \mathcal{A} can distinguish between the YES-INSTANCE and NO-INSTANCE in Theorem 1.1, implying that $\text{NP} = \text{RP}$. This completes the proof. \square

A similar but weaker theorem can be proven for the case of NFAs. Indeed, we rule out the existence Occam algorithm for NFA with parameter $0 \leq \alpha \leq 1/2$, assuming that $\text{NP} \neq \text{RP}$.

Theorem 5.13. *Unless $NP = RP$, there is no polynomial time randomized Occam algorithm for NFA with parameter $0 \leq \alpha \leq 1/2$.*

Proof. The proof is essentially the same as that of Theorem 5.12 with slightly different parameters.

We prove by contrapositive. Assume that there is a randomized Occam algorithm \mathcal{A} for NFA with parameters (k, α) for some constants $k \geq 0$ and $0 \leq \alpha \leq 1/2$. We will show that the algorithm \mathcal{A} can be used to distinguish between the YES-INSTANCE and NO-INSTANCE given in Theorem 1.4 and thus implying that $NP = RP$.

Take an instance of MinCon(ADFA,NFA) as in Theorem 1.1. So, we have a pair of sets (P, N) of N samples, each of length $O(\log N)$. The parameters of the Occam algorithm \mathcal{A} are thus $s = N$ and $m = O(\log N)$.

We choose the parameter ϵ in Theorem 1.1 to be $\epsilon = (1/2 - \alpha)/(2k + 1)$.

In the YES-INSTANCE, there is an NFA of size N^ϵ consistent with the samples. Thus, our hypothesis class has size $n = N^\epsilon$. By definition, the Occam algorithm \mathcal{A} gives us a NFA M with size

$$|M| \leq N^{\epsilon k} \cdot (\log N)^k \cdot N^\alpha \leq N^{\alpha + (2k)\epsilon} = N^{\alpha + (2k)\frac{1/2 - \alpha}{2k + 1}} = N^{1/2 - \frac{1/2 - \alpha}{2k + 1}} = N^{1/2 - \epsilon}$$

In the NO-INSTANCE, any NFA M consistent with (P, N) has size $|M| > N^{1/2 - \epsilon}$.

Therefore, the randomized Occam algorithm \mathcal{A} can distinguish between the YES-INSTANCE and NO-INSTANCE in Theorem 1.4, implying that $NP = RP$. This completes the proof. \square

Corollary 5.14. *Unless $NP = RP$, there are no Occam algorithms for the following hypothesis classes:*

- *Deterministic Finite Automata (DFA)*
- *Acyclic Deterministic Finite Automata (ADFA)*
- *Ordered Branching Decision Diagram (OBDD)*

In particular, for any $\epsilon \in (0, 1), k > 0$, the minimum consistent hypothesis problems for these classes are $N^{1 - \epsilon} \text{OPT}^k$ -hard to approximate unless $NP = RP$.

6 Hardness of EDP on DAGs

In this section, we prove the $|V(G)|^{1/2 - \epsilon}$ hardness of approximating EDP on DAGs and packing vertex-disjoint bounded length cycles. We will first show the construction for EDP, and later we argue that a slight modification of the construction yields the hardness of packing vertex-disjoint bounded length cycles.

6.1 Reduction R

We first define the canonical reduction $\vec{R}[G]$ formally. Given a graph $G = (V, E)$ on n vertices, the switching graph of G , denoted by $\vec{R}[G]$, is a graph defined on a plane and constructed in two steps as follows. The coordinates of graph $\vec{R}[G]$ lie in the box formed by the corners $(0, 0)$ and (n, n) .

FIRST STEP: For each vertex $i \in V(G)$, we draw a line segment ℓ_i on the plane connecting vertices s_i and t_i as shown in Figure 4. To be precise, the line ℓ_i goes from the coordinate $(n + 1 - i, 0)$

to the coordinate $(n + 1 - i, i)$ of the grid and then goes to the coordinate $(0, i)$. For each pair of vertices $i, j \in V(G)$, we have an *intersection point* $y_{i,j}$ at the crossing point of lines ℓ_i and ℓ_j . Some of these intersection points will be later defined as vertices in the switching graphs whereas others are just a crossing points in the plane embedding. We call this graph $\widehat{R}[G]$ which will also be crucial in our analysis. Edges in $\widehat{R}[G]$ are directed from left to right and top to bottom.

SECOND STEP: For each edge $ij \in E(G)$, we split $y_{i,j}$ into two vertices $x_{i,j}^{\text{in}}$ and $x_{i,j}^{\text{out}}$ and have a directed edge $e_{i,j} = x_{i,j}^{\text{in}}x_{i,j}^{\text{out}}$ in the graph $\vec{R}[G]$. Otherwise, if $ij \notin E(G)$, the intersection point $y_{i,j}$ is replaced by an uncrossing as in Figure 4.

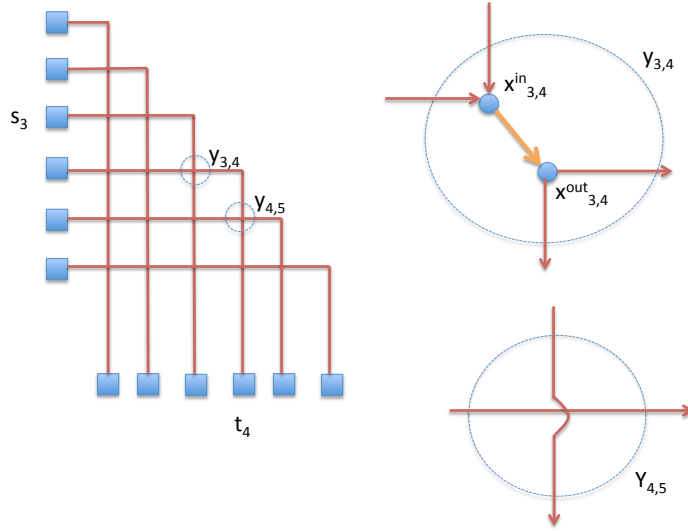


Figure 4: The graph $\vec{R}[G]$ where $(3, 4) \in E(G)$ but $(4, 5) \notin E(G)$

First, the following lemma establishes a (simple) connection between EDP and the maximum independent set problem.

Lemma 6.1. *For any graph H , $\text{edp}(\vec{R}[H]) \geq \alpha(H)$.*

Proof. Let $S \subseteq V(H)$ be any independent set in H . We define the collection of paths $\mathcal{P}_S = \{P_i\}_{i \in S}$ in graph $\vec{R}[H]$. Since S is an independent set, any pair of paths P_i and P_j for $i, j \in S$ are disjoint by construction. \square

Unfortunately, the converse of this inequality does not hold within any reasonably small factor. In fact, there is a graph H for which $\alpha(H) = 2$ but $\text{edp}(\vec{R}[H]) = \Omega(n)$; see Appendix A. Therefore, we focus on proving the low α -projection property.

6.2 α -Projection Property

For technical reasons, we will need to analyze a slightly different measure from the optimal value $\text{edp}(\vec{R}[G])$. This notion will be a weaker notion of feasible solutions for EDP. We say that a

collection of disjoint paths $\mathcal{P} = \{P_1, \dots, P_\ell\}$ is *orderly feasible* if for any pair $P = (s_i, \dots, t_j)$ and $P' = (s_{i'}, \dots, t_{j'})$ such that $i < i'$, then it must be the case that $j < j'$; for instance, in an orderly feasible set, if we connect s_1 to t_3 , it must be the case that s_2 is connected to t_j for $j > 3$. Intuitively, in an orderly feasible set \mathcal{P} , a path is allowed to start from s_i and ends at some sink t_j for $j \neq i$, but every pair of paths in \mathcal{P} is forced to “cross” at some point. Observe that any collection of feasible edge disjoint paths must also be orderly feasible. As a consequence, if we define $\widetilde{\text{edp}}(\vec{R}[G])$ as the maximum cardinality of all orderly feasible collections of paths, then we have that $\widetilde{\text{edp}}(\vec{R}[G]) \geq \text{edp}(\vec{R}[G])$.

The following observation is more or less obvious.

Observation 6.2. *For any graph G , $\widetilde{\text{edp}}(\vec{R}[G]) \leq |\vec{R}[G]|$*

Next, the following lemma will finish the proof of the low α -projection property.

Lemma 6.3. *For any two graphs G and H ,*

$$\widetilde{\text{edp}}(\vec{R}[G \cdot H]) \leq 3|V(G)|^2 + \alpha(G)\widetilde{\text{edp}}(\vec{R}[H])$$

We will spend the rest of this section to prove the lemma.

6.3 Geometry of Paths: Regions, switching boxes, and configurations

This section discusses the structure of the graph $\vec{R}[G \cdot H]$ and a feasible solution for EDP in $\vec{R}[G \cdot H]$. We define some terminologies that will be needed in the analysis.

Ordering of Paths We need a notion of “ordering” of edge-disjoint paths with respect to certain curve. We think of graph $\widehat{R}[G]$ as being drawn on the plane with standard x and y coordinates. All sources and sinks are on y and x axes respectively.

For any collection of edge-disjoint paths \mathcal{P} in $\vec{R}[G]$, one can naturally map these paths on the graph $\widehat{R}[G]$ and think of them as curves on the plane. A continuous curve $C : [0, 1] \rightarrow \mathbb{R}^2$ is said to be *good* if for all $t < t'$, point $C(t)$ is dominated by point $C(t')$ in the plane and the curve C does not go through any intersection point $y_{i,j}$ (informally, the curve is directed to the top and right). Let C be any good curve. The ordering \preceq_C is defined on the set of paths \mathcal{P}' intersecting C as follows: Paths $P \prec_C P'$ if and only if C intersects P before it intersects P' . Since C does not intersect point $y_{i,j}$, either $P \prec_C P'$ or $P' \prec_C P$.

Regions and Switching Boxes. In $\vec{R}[G \cdot H]$, we have canonical paths P_{ia} for $i \in V(G)$ and $a \in V(H)$. For each $i \in V(G)$, we define a *region* R_i on the plane that contains all paths $P_{(i,a)}$ for $a \in [r]$. For $i, j \in V(G)$, the intersection between regions R_i and R_j is called a *bounding box* $B(i, j)$ which contains $|V(H)|^2$ virtual vertices of the form $Y(i, a), (j, b)$ for $a, b \in V(H)$. Notice that a canonical path $P_{(i,a)}$ is completely contained inside region R_i , and as we walk on the path from $s(i, a)$ to $t(i, a)$, we will visit the bounding boxes $B(i, 1), \dots, B(i, n)$ in this order. For convenience, the region in R_i between $B(i, i - 1)$ and $B(i, i + 1)$ is called $B(i, i)$. See Figure 6 for illustration.

Proposition 6.4. *Consider any box $B(i, j)$ for $i \neq j$. One of the following two cases holds:*

- *For all $a, b \in V(H)$, the virtual vertex $y_{(i,a),(j,b)}$ is a directed edge $e_{(i,a),(j,b)}$. This happens when $ij \in E(G)$, and we say that the box $B(i, j)$ is a non-switching box.*

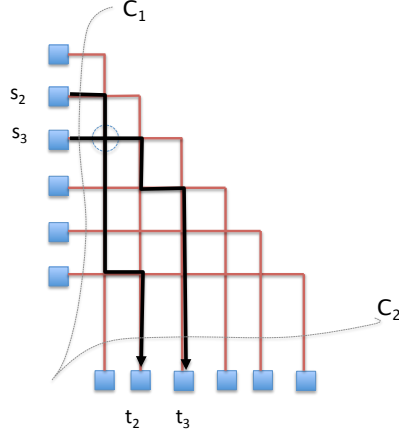


Figure 5: Both C_1 and C_2 are good curves that originated from $(0,0)$ (illustrated by dotted lines). We have $Q' \prec_{C_1} Q$ while $Q \prec_{C_2} Q'$.

- For all $a, b \in V(H)$, the virtual vertex is an uncrossing, in which case, we say that the box $B(i, j)$ is a switching box.

The term switching box is coined from an intuitive reason: Consider a switching box $B(i, j)$ and a collection of edge-disjoint paths that are routed in a solution. Let \mathcal{P}_{top} and \mathcal{P}_{left} be the paths in the solution that enter this box from the top and left respectively, so paths in \mathcal{P}_{top} (resp. \mathcal{P}_{left}) must leave the box from the bottom (resp. right). Define the curves C_{in} (and C_{out}) as the union of left and top boundaries of $B(i, j)$ (resp. the union of right and bottom boundaries). With respect to the curve C_{in} all paths in \mathcal{P}_{top} are ordered after paths in \mathcal{P}_{left} , while this becomes the opposite for C_{out} . In other words, the box $B(i, j)$ “switches” the order of these paths.

6.4 Proof

Now we prove Lemma 6.3. Let I be the set of indices of edge-disjoint paths in $\vec{R}[G \cdot H]$ where, for each $(i, a) \in I$, there is a path $Q_{(i,a)}$ connecting $s(i, a)$ to $t(\psi(i, a))$ and the paths $Q_{(i,a)}$ are edge-disjoint and orderly feasible (recall that orderly feasible solutions may connect $s(i, a)$ to some other sink $t(i', a')$). We say that a path $Q_{(i,a)}$ is *semi-canonical* if it is completely contained in region R_i . Let $I' \subseteq I$ be the set of semi-canonical paths.

Lemma 6.5. $|I'| \leq \alpha(G) \widetilde{\text{edp}}(\vec{R}[H])$

Proof. We first define the partition of I' by the first coordinates of paths. Define $I'_i = \{(i, a) : (i, a) \in I'\}$, so we will have $I' = \bigcup_{i \in V(G)} I'_i$. We count the number of indices $i \in V(G)$ such that $I'_i \neq \emptyset$. Define $\Lambda = \{i \in V(G) : I'_i \neq \emptyset\}$. We claim that Λ is an independent set and therefore $|\Lambda| \leq \alpha(G)$: Suppose $i, j \in V(G)$ such that $I'_i, I'_j \neq \emptyset$. Let $(i, a) \in I'_i$ and $(j, b) \in I'_j$ be any two paths. Due to the fact that this is an orderly feasible solution, these two paths must cross at some virtual vertex $y_{(i,a'),(j,b')}$ inside box $B(i, j)$, and they must share an edge $e_{(i,a'),(j,b')}$, a contradiction. This implies that $|\Lambda| \leq \alpha(G)$.

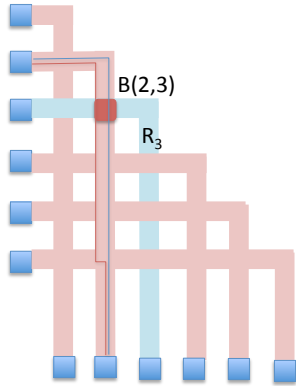


Figure 6: Regions and switching boxes in $\vec{R}[G]$. There are two paths routed inside region R_2 .

Next, we argue that $|I'_i| \leq \widetilde{\text{edp}}(\vec{R}[H])$ for all $i \in \Lambda$, which will complete the proof of the lemma. If we consider the box $B(i, i)$, we see an isomorphic copy H' of graph $\vec{R}[H]$, in which each path $Q_{(i,a)}$ corresponds to another path $Q'_{\varphi(a)}$ that connects some “source” $s'(\varphi(a))$ to “sink” $t'(\xi(a))$. We claim that the collection of paths $Q'_{\varphi(a)}$ is orderly feasible in the instance $(H', \{(s'(a), t'(a))\}_{a \in V(H)})$: Assume otherwise that some paths $Q'_{\varphi(a)}$ and $Q'_{\varphi(b)}$ do not cross inside H' , so the paths $Q_{(i,a)}$ and $Q_{(i,b)}$ must cross at some other box $B(i, j)$ for $i \neq j$. If such box is a switching box, it is impossible for these two paths to cross because they must enter and leave the box from the same direction; otherwise, if box $B(i, j)$ is not a switching box, it is also impossible for them to cross. \square

Let $I'' = I \setminus I'$. For convenience, let us renumber I'' such that $I'' = \{1, \dots, t\}$ such that the source of path 1 is above that of path 2 and so on. Now we will show that $|I''| \leq 3|E(G)| + 1$. Our proof will rely on the notion of configurations. We first define the order of boxes $B(i, j)$ for $i > j$ such that $B(2, 1)$ is the first box, which precedes $B(3, 1)$ (the second box), and so on. More formally, the box $B(i, j)$ precedes $B(i', j')$ if and only if $i < i'$ or $i = i'$ and $j < j'$; in short, this is simply a lexicographic order of boxes. This defines a total order over boxes.

We define a number of good curves C_1, \dots, C_z for $z = \binom{|V(G)|}{2}$, where the curve C_h is any good curve such that (i) $C_h(0) = (0, 0)$, (ii) $C_h(1) = (x_{\max}, y_{\max})$ and (iii) the first $h - 1$ boxes are above C_h , while $z - h + 1$ curves are below it (notice that C_h partitions the region $[0, n + 1] \times [0, n + 1]$ into two parts, i.e. one above the curve and the other below it).

Observation 6.6. *For each $h = 1, \dots, z$ and path $i \in I''$, the curve C_h intersects path $i \in I''$.*

Proof. This is just because any path in the orderly feasible solution starts from the region above the curve C_h , while it ends in the region below the curve. \square

For each $h = 1, \dots, z$, a curve C_h can be used to define a configuration $\sigma_h = (x_1, \dots, x_t)$ of paths in I'' where $x_j \in I''$ is the index of the j th path that intersects with the curve C_h (this order is well-defined because the curve has directions). Notice that $\sigma_1 = (t, \dots, 1)$, and $\sigma_z = (1, \dots, t)$.

The number of *reversals* of a configuration σ_h is the number of locations j such that $\sigma_j > \sigma_{j+1}$. Denote this number by $rev(\sigma_h)$, so we have that $rev(\sigma_1) = t - 1$, and $rev(\sigma_z) = 0$. Our proof proceeds by analyzing how the number of reversals changes over configurations $\sigma_1, \dots, \sigma_z$. We will show that, for any h , we have $rev(\sigma_h) - rev(\sigma_{h+1}) \leq 3$, which implies that $t - 1 = rev(\sigma_1) - rev(\sigma_z) = \sum_{h=1}^{z-1} (rev(\sigma_h) - rev(\sigma_{h+1})) \leq 3(z - 1)$; in other words, $|I''| = t \leq 3z \leq 3|V(G)|^2$. So the last thing we need to prove is the following lemma:

Lemma 6.7. *For any $h = 1, \dots, z - 1$, we have $rev(\sigma_h) - rev(\sigma_{h+1}) \leq 3$.*

Proof. Let $B(i, j)$ be the h th box and $J \subseteq I''$ be the indices of paths entering this box. If the box $B(i, j)$ is a non-switching box, then it must be the case that $\sigma_{h+1} = \sigma_h$ due to the fact that paths cannot cross inside region $B(i, j)$. This implies that $rev(\sigma_h) = rev(\sigma_{h+1})$ in this case.

Now we consider the other case when $B(i, j)$ is a switching box. We write $J = J_{top} \cup J_{left}$ where J_{top} (and J_{left}) is the set of indices of paths entering box $B(i, j)$ from the top (and left respective). It is clear that paths coming out of the bottom and right of the box are exactly J_{top} and J_{left} respectively. Notice that, while the curve C_h crosses J_{top} after J_{left} , the curve C_{h+1} would cross paths in J_{left} before those in J_{top} . The configurations σ_h and σ_{h+1} can be written as $\sigma_h = \sigma' \circ \sigma^{left} \circ \sigma^{top} \circ \sigma''$ and $\sigma_{h+1} = \sigma' \circ \sigma^{top} \circ \sigma^{left} \circ \sigma''$ respectively. See Figure 7 for illustration. \square

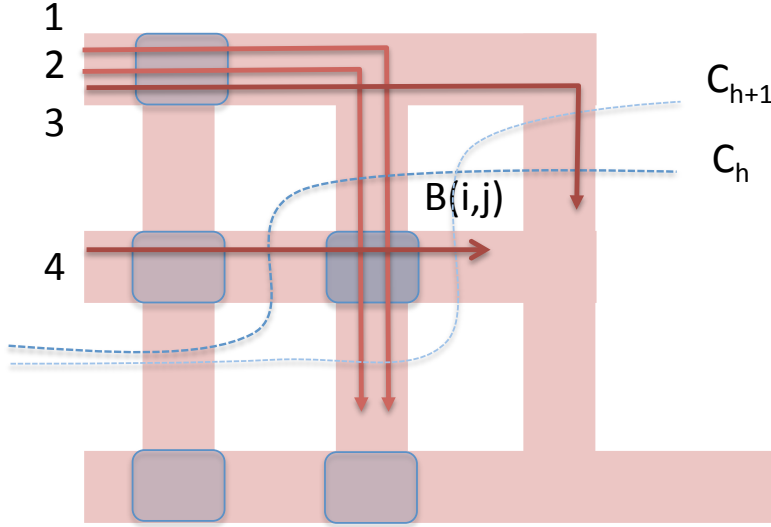


Figure 7: Configurations of curve C_h and C_{h+1} are $\sigma_h = (4, 2, 1, 3)$ and $\sigma_{h+1} = (2, 1, 4, 3)$ respectively. In this case, $\sigma^{top} = (2, 1)$ and $\sigma^{left} = (4)$. Box $B(i, j)$ is a switching box.

7 Other Problems

In this section, we prove the hardness of k -Cycle Packing and DNF/CNF Minimization. As noted previously, our proof for CNF minimization is an alternative proof of Aleknovich et al. [ABF⁺08].

7.1 Hardness of k -Cycle Packing for Large k

We consider the problem of packing edge-disjoint k -cycles in which our goal is to pack as many cycles of length at most k as possible. We only need to slightly change the reduction $\vec{R}[G]$ as used in Section 6 in the following way: In the second step, for each pair $i, j \in V(G)$, if $ij \in E(G)$, we do the same, but for $ij \notin E(G)$ (including the case when $i = j$), we make two new vertices on each line before and after the jump (see Figure 8). Also, we have a *back edge* from t_i to s_i for each $i \in V(G)$.

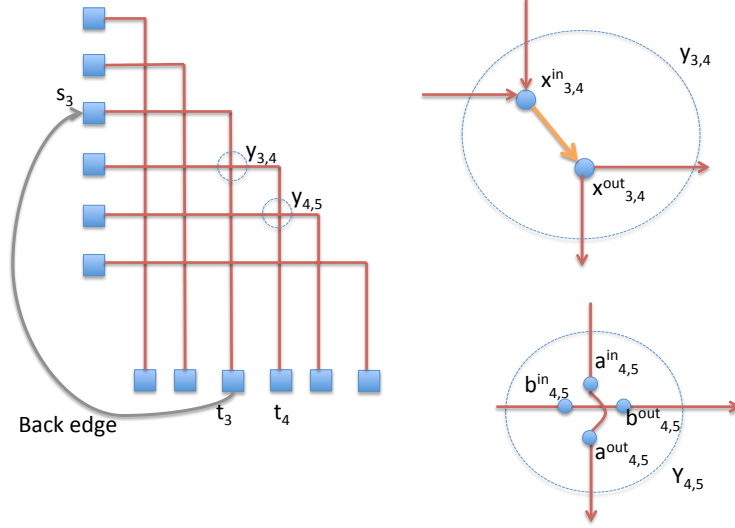


Figure 8: The graph $\vec{R}[G]$ where $(3,4) \in E(G)$ but $(4,5) \notin E(G)$. The differences between this gadget and EDP gadget only lies in the new vertices $a_{i,j}^{in}, a_{i,j}^{out}, b_{i,j}^{in}, b_{i,j}^{out}$

With this reduction, any “canonical” cycle between source s_i to sink t_i (and taking back edge to s_i) must have length exactly $2n + 2$, so we choose the value $k = 2n + 2$. Let $edc(\vec{R}'[G])$ denote the optimal value of k -cycle packing. We now establish the connection between the optimal value of EDP solution in $\vec{R}[G]$ and the k -EDC solution in $\vec{R}'[G]$.

Notice that for any cycle that uses only one back edge, there is a corresponding path from some s_i to t_i . The number of these cycles corresponds exactly to $\text{EDP}(\vec{R}[G])$, so we can write

$$\text{EDP}(\vec{R}[G]) \leq \text{edc}(\vec{R}'[G]) \leq \text{EDP}(\vec{R}[G]) + \widetilde{\text{edc}}(\vec{R}'[G])$$

where $\widetilde{\text{edc}}(\vec{R}'[G])$ is the number of cycles that use more than one back edge. The following lemma says that these cycles must be longer than k , i.e. $\widetilde{\text{edc}}(\vec{R}'[G]) = 0$. In other words, this implies that $\text{edc}(\vec{R}'[G]) = \text{EDP}(\vec{R}[G])$.

Lemma 7.1. *Let C be a cycle in $\vec{R}'[G]$ that uses more than one back edge. Then $|C| > k$.*

Proof. Any cycle C must start at some s_i and ends at s_i . Let $i = i_0, i_1, \dots, i_\ell = i$ be the indices of the source-sink pairs visited in the cycle C , i.e. the cycle goes through $s_{i_0} \rightarrow t_{i_1} s_{i_1} \rightarrow s_{i_2} \rightarrow \dots \rightarrow$

$t_{i_\ell} s_{i_\ell}$. Observe that a path that goes from s_j to $t_{j'}$ visit exactly $(n - j + j')$ vertices of the form $y_{i,j}$, because such path must go right j' times and go down $n - j$ times (in arbitrary order). Combining this with s_j and $t_{j'}$, such path would visit $2(n - j + j' + 1)$ vertices. Therefore, the total length of the cycle C is

$$\sum_{j=0}^{\ell-1} 2(n - i_j + i_{j+1} + 1) = 2\ell(n + 1) + 2(i_\ell - i_0) = 2\ell(n + 1)$$

So this cycle would have been longer than the threshold k if $\ell > 1$. \square

7.2 Learning CNF Formula

We present an alternative proof for the hardness of properly learning CNF using our framework. Our reduction is quite similar to Alekhovich et al.'s (see [ABF⁺08]), but our proof highlights the role of graph products in the proof (while their construction cannot be seen as a standard graph product in any way).

Let G be any graph. We think of a vertex $u \in V(G)$ as an integer in $\{1, \dots, n\}$. For each vertex $u \in V(G)$, we define an encoding $\langle u \rangle = 0^{u-1}10^{n-u}$. For each edge $uv \in E(G)$, the encoding of an edge uv has two 1s at the positions corresponding to u and v . Our reduction encodes the k -fold graph product $H = G^k$ into samples as follows. For each $\vec{u} = (u_1, \dots, u_k) \in V(H)$, we define a negative sample $neg(\vec{u}) = \langle u_1 \rangle \dots \langle u_k \rangle$. For each $\vec{u} \in E(H)$, $i \in [k]$ and $u_i v \in E(G)$, we define a positive sample $pos(\vec{u}, v, i) = \langle u_1 \rangle \dots \langle u_{i-1} \rangle \langle u_i v \rangle \langle u_{i+1} \rangle \dots \langle u_k \rangle$.

Notice that the total number of variables is nk , where we think of them as k blocks; in each of which, there are n variables. Denote by $z(i, u)$ the variable in block $i \in [k]$ that corresponds to a vertex $u \in V(G)$.

Lemma 7.2. $\text{OPT}(R[H]) \geq \chi(H)$

Proof. Suppose $\bigwedge_{j=1}^M C_j$ be a CNF formula that is consistent with all samples. We claim that the number of clauses M is at least $\chi(H)$. For each $\vec{u} \in V(H)$, let $\sigma(\vec{u})$ be the index such that $C_{\sigma(\vec{u})}$ evaluates to false on sample $neg(\vec{u})$; this clause must exist since this is a negative sample (if there are many such indices for vertex \vec{u} , we choose any arbitrary one). Now for each $s \in [M]$, we define the set of vertices $Q_s \subseteq V(H)$ as $Q_s = \{\vec{u} : \sigma(\vec{u}) = s\}$.

Claim 7.3. Q_s is an independent set for all $s \in [M]$.

Proof. Assume otherwise that some $\vec{u}\vec{v} \in E(H)$ such that $\vec{u}, \vec{v} \in Q_s$. Let i be the index such that $u_j = v_j$ for all $j < i$ and $u_i v_i \in E(G)$. Let $X, Y \subseteq [k] \times V(G)$ be the subset of variable indices that appear positively and negatively in clause C_s , so we can rewrite $C_s = \left(\bigvee_{(j,u) \in X} z(j, u) \right) \vee \left(\bigvee_{(j,u) \in Y} \overline{z(j, u)} \right)$. Since C_s evaluates to false on both $neg(\vec{u})$ and $neg(\vec{v})$, we can neither have variable $z(i, u_i)$ nor $z(i, v_i)$ in the clause C_s : Suppose otherwise that $(i, u_i) \in X$ or $(i, u_i) \in Y$, then either $neg(\vec{u})$ or $neg(\vec{v})$ would have evaluated to true on clause C_s (contradicting $\vec{u} \in Q_s$). Similarly, if $(i, v_i) \in X$ or $(i, v_i) \in Y$, then either $neg(\vec{v})$ or $neg(\vec{u})$ would have been true in clause C_s .

In other words, $(i, u), (i, v) \notin X \cup Y$. But notice that $pos(\vec{u}, v_i, i)_{(i', u')} = neg(\vec{u})_{(i', u')}$ for all $(i', u') \notin \{(i, u_i), (i, v_i)\}$, so C_s must evaluate to false on input $pos(\vec{u}, v_i, i)$, a contradiction. \square

We have just shown that $\{Q_s\}_{s \in [M]}$ is a valid M -coloring for graph H , so we must have $M \geq \chi(H)$, as desired. \square

Next, we prove the upper bound.

Lemma 7.4. $\text{OPT}(R[H]) \leq \chi(G)^k n^{O(1)}$

Proof. We construct the same formula as in Aleknovich et al. That is, let I_1, \dots, I_M be color classes of G and $\sigma : V(G) \rightarrow [M]$ be the corresponding coloring function. Define the formula $f_i(z) = \bigwedge_{c=1}^M \bigvee_{u \notin I_c} z(i, u)$, for $i = 1, 2, \dots, k$, and define $f(z) = \bigvee_{i=1}^k f_i(z)$. This formula can be turned into a CNF of size at most $\chi(G)^k |V(G)|^{O(1)}$.

Claim 7.5. *The formula f is consistent with all the samples.*

Proof. Consider each negative sample $neg(\vec{u})$ for $\vec{u} = (u_1, \dots, u_k) \in V(H)$. For each $i \in [k]$, notice that $f_i(\langle u_i \rangle)$ evaluates to false because $\bigvee_{u \notin I_{\sigma(u_i)}} \langle u_i \rangle_u$ is false (since $u_i \in I_{\sigma(u_i)}$ is the only bit of $\langle u_i \rangle$ that is “1”). This implies that $\bigvee_{i=1}^k f_i(neg(\vec{u}))$ is false.

Now consider, for each $\vec{u} \in V(H)$, $v : u_i v \in E(G)$ and $i \in [k]$, a positive sample $pos(\vec{u}, v, i)$. We claim that $f_i(pos(\vec{u}, v, i))$ is true, which causes $f(pos(\vec{u}, v, i))$ to be true: Assume to the contrary that f_i is false, so some term $\bigvee_{w \notin I_c} \langle u_i v \rangle_w$ is false for some c ; notice that it can be false only if $\langle u_i v \rangle_w = 0$ for all $w \notin I_c$; since we have $\langle u_i v \rangle_{u_i} = \langle u_i v \rangle_v = 1$, it must be the case that both u_i and v belong to I_c , contradicting the fact that I_c is a color class. \square

\square

8 Conclusion and Open Problems

We have shown applications of pre-reduction graph product techniques in proving hardness of approximation. For some applications, such as EDP, proving α -projection property implies tight hardness, but for some others, we need a more careful reduction of the form $R[G^\ell]$ (taking into account the fact that the input is an ℓ -fold product of graphs).

There are many open problems on edge-disjoint paths. Most notably can one narrow down the gap of undirected EDP between $O(\sqrt{n})$ upper bound and $\log^{1/2-\epsilon} n$ lower bound? For directed EDP, there is still a (relatively large) gap in the case of low congestion routing, between the upper bound of $n^{1/c}$ [KS04] and the lower bound of $n^{1/(3c+23)}$ [CGKT07] if we allow routing with congestion c . We believe that our techniques are likely to work there (in a much more sophisticated manner), and it would potentially close this gap. This would resolve an open question in Chuzhoy et al. [CGKT07].

Another interesting problem is the cycle packing problem. For this problem, the approximability is pretty much settled on undirected graphs with an upper bound of $O(\log^{1/2} n)$ and a lower bound of $\log^{1/2-\epsilon} n$ [FS11, KNS⁺07]. On directed graphs, there is still a large gap between $n^{1/2}$ and $\Omega(\frac{\log n}{\log \log n})$. For k -cycle packing problem, it is interesting to see whether our technique gives $k^{1-\epsilon}$ hardness for small k .

Acknowledgement: We thank Julia Chuzhoy for suggesting the EDP reduction and for related discussions when the first author was still at the University of Chicago.

References

- [Aar08] Scott Aaronson. 6.080/6.089 Great Ideas in Theoretical Computer Science, Spring 2008, Lecture 21. MIT OpenCourseWare, 2008. Available at <http://stellar.mit.edu/S/course/6/sp08/6.080/courseMaterial/topics/topic1/lectureNotes/lec21/lec21.pdf>. 3
- [ABF⁺08] Michael Alekhnovich, Mark Braverman, Vitaly Feldman, Adam R. Klivans, and Toniann Pitassi. The complexity of properly learning simple concept classes. *J. Comput. Syst. Sci.*, 74(1):16–34, 2008. Announced at FOCS 2004. 5, 6, 29, 30
- [ABX08] Benny Applebaum, Boaz Barak, and David Xiao. On basing lower-bounds for learning on worst-case assumptions. In *FOCS*, pages 211–220, 2008. 3
- [ACG⁺10] Matthew Andrews, Julia Chuzhoy, Venkatesan Guruswami, Sanjeev Khanna, Kunal Talwar, and Lisa Zhang. Inapproximability of edge-disjoint paths and low congestion routing on undirected graphs. *Combinatorica*, 30(5):485–520, 2010. 4
- [Ang78] Dana Angluin. On the complexity of minimum inference of regular sets. *Information and Control*, 39(3):337–350, 1978. 2
- [AZ06] Matthew Andrews and Lisa Zhang. Logarithmic hardness of the undirected edge-disjoint paths problem. *J. ACM*, 53(5):745–761, 2006. Also, in STOC’05. 4
- [BP92] Raymond Board and Leonard Pitt. On the necessity of occam algorithms. *Theoretical Computer Science*, 100(1):157 – 184, 1992. 3, 22
- [BS92] Piotr Berman and Georg Schnitger. On the complexity of approximating the independent set problem. *Inf. Comput.*, 96(1):77–94, 1992. Also, in STACS’89. 1
- [CGKT07] Julia Chuzhoy, Venkatesan Guruswami, Sanjeev Khanna, and Kunal Talwar. Hardness of routing with congestion in directed graphs. In *STOC*, pages 165–178, 2007. 2, 4, 31
- [Chu12] Julia Chuzhoy. Routing in undirected graphs with constant congestion. In *STOC*, pages 855–874, 2012. 3
- [CK07] Chandra Chekuri and Sanjeev Khanna. Edge-disjoint paths revisited. *ACM Transactions on Algorithms*, 3(4), 2007. 4
- [CKS05] Chandra Chekuri, Sanjeev Khanna, and F. Bruce Shepherd. Multicommodity flow, well-linked terminals, and routing problems. In *STOC*, pages 183–192, 2005. 3
- [CKS06] Chandra Chekuri, Sanjeev Khanna, and F. Bruce Shepherd. An $O(\sqrt{n})$ approximation and integrality gap for disjoint paths and unsplittable flow. *Theory of Computing*, 2(1):137–146, 2006. 2, 4
- [CKS09] Chandra Chekuri, Sanjeev Khanna, and F. Bruce Shepherd. Edge-disjoint paths in planar graphs with constant congestion. *SIAM J. Comput.*, 39(1):281–301, 2009. 3
- [CL12] Julia Chuzhoy and Shi Li. A polylogarithmic approximation algorithm for edge-disjoint paths with congestion 2. In *FOCS*, pages 233–242, 2012. 3

- [CLN13a] Parinya Chalermsook, Bundit Laekhanukit, and Danupon Nanongkai. Graph products revisited: Tight approximation hardness of induced matching, poset dimension and more. In *SODA*, pages 1557–1576, 2013. [9](#), [35](#)
- [CLN13b] Parinya Chalermsook, Bundit Laekhanukit, and Danupon Nanongkai. Independent set, induced matching, and pricing: Connections and tight (subexponential time) approximation hardnesses. In *FOCS*, pages 370–379, 2013. [9](#)
- [CLN14] Parinya Chalermsook, Bundit Laekhanukit, and Danupon Nanongkai. Coloring graph powers: Graph product bounds and hardness of approximation. In *LATIN*, pages 409–420, 2014. [9](#)
- [DIH10] Colin De la Higuera. *Grammatical inference: learning automata and grammars*. Cambridge University Press, 2010. [2](#), [3](#)
- [DLSS14] Amit Daniely, Nati Linial, and Shai Shalev-Shwartz. From average case complexity to improper learning complexity. In *STOC*, pages 441–448, 2014. [3](#), [5](#)
- [Fei02] Uriel Feige. Relations between average case complexity and approximation complexity. In *STOC*, pages 534–543, 2002. [3](#)
- [Fel08] Vitaly Feldman. Hardness of proper learning. In *Encyclopedia of Algorithms*. Springer, 2008. [3](#)
- [FK98] Uriel Feige and Joe Kilian. Zero knowledge and the chromatic number. *J. Comput. Syst. Sci.*, 57(2):187–199, 1998. Also, in CCC 1996. [11](#), [12](#)
- [FS11] Zachary Friggstad and Mohammad R. Salavatipour. Approximability of packing disjoint cycles. *Algorithmica*, 60(2):395–400, 2011. [31](#)
- [GKR⁺03] Venkatesan Guruswami, Sanjeev Khanna, Rajmohan Rajaraman, F. Bruce Shepherd, and Mihalis Yannakakis. Near-optimal hardness results and approximation algorithms for edge-disjoint paths and related problems. *J. Comput. Syst. Sci.*, 67(3):473–496, 2003. Also, in STOC 1999. [4](#)
- [GL14] Venkatesan Guruswami and Euiwoong Lee. Inapproximability of feedback vertex set for bounded length cycles. *Electronic Colloquium on Computational Complexity (ECCC)*, 21:6, 2014. [2](#), [5](#)
- [Gol78] E. Mark Gold. Complexity of automaton identification from given data. *Information and Control*, 37(3):302–320, 1978. [2](#)
- [Hås96] Johan Håstad. Clique is hard to approximate within $n^{1-\epsilon}$. In *FOCS*, pages 627–636, 1996. [11](#)
- [KK10] Kenichi Kawarabayashi and Yusuke Kobayashi. The edge disjoint paths problem in eulerian graphs and 4-edge-connected graphs. In *SODA*, pages 345–353, 2010. [3](#)
- [Kle96] Jon Michael Kleinberg. *Approximation algorithms for disjoint paths problems*. PhD thesis, Citeseer, 1996. [4](#)

- [Kle05] Jon M. Kleinberg. An approximation algorithm for the disjoint paths problem in even-degree planar graphs. In *FOCS*, pages 627–636, 2005. [3](#)
- [KMR97] David R. Karger, Rajeev Motwani, and G. D. S. Ramkumar. On approximating the longest path in a graph. *Algorithmica*, 18(1):82–98, 1997. [1](#)
- [KNS⁺07] Michael Krivelevich, Zeev Nutov, Mohammad R. Salavatipour, Jacques Yuster, and Raphael Yuster. Approximation algorithms and hardness results for cycle packing problems. *ACM Transactions on Algorithms*, 3(4), 2007. [4](#), [5](#), [31](#)
- [KS04] Stavros G. Kolliopoulos and Clifford Stein. Approximating disjoint-path problems using packing integer programs. *Math. Program.*, 99(1):63–87, 2004. [31](#)
- [KT98] Jon M. Kleinberg and Éva Tardos. Approximations for the disjoint paths problem in high-diameter planar networks. *J. Comput. Syst. Sci.*, 57(1):61–73, 1998. [3](#)
- [KV94] Michael J. Kearns and Leslie G. Valiant. Cryptographic limitations on learning boolean formulae and finite automata. *J. ACM*, 41(1):67–95, 1994. Announced at STOC 1989. [3](#), [5](#)
- [LV88] Ming Li and Umesh V. Vazirani. On the learnability of finite automata. In *COLT*, pages 359–370, 1988. [2](#)
- [Pit89] Leonard Pitt. *Inductive inference, DFAs, and computational complexity*. Springer, 1989. [2](#), [3](#), [22](#)
- [PV88] Leonard Pitt and Leslie G. Valiant. Computational limitations on learning from examples. *J. ACM*, 35(4):965–984, 1988. [3](#)
- [PW93] Leonard Pitt and Manfred K. Warmuth. The minimum consistent DFA problem cannot be approximated within any polynomial. *J. ACM*, 40(1):95–142, 1993. Announced at STOC 1989. [2](#), [3](#), [5](#)
- [RS95] Neil Robertson and Paul D. Seymour. Graph minors .xiii. the disjoint paths problem. *J. Comb. Theory, Ser. B*, 63(1):65–110, 1995. [3](#)
- [VV04] Kasturi R. Varadarajan and Ganesh Venkataraman. Graph decomposition and a greedy algorithm for edge-disjoint paths. In *SODA*, pages 379–380, 2004. [4](#)
- [Zuc07] David Zuckerman. Linear degree extractors and the inapproximability of max clique and chromatic number. *Theory of Computing*, 3(1):103–128, 2007. Also, in STOC 2006. [11](#), [12](#)

Appendix

A List of Bad Examples

In this section, we provide the evidences that all the reductions considered in this paper are neither trivially “working” nor subadditive in the sense of our previous SODA paper [CLN13a]. Therefore, we will need the new conceptual ideas introduced in this paper.

A.1 Edge Disjoint Paths

We show a graph G in which $\alpha(G) = 2$ but $\text{EDP}(\vec{R}[G]) = n/3$. To ensure that G does not have an independent set of size 3, we define graph G by defining a triangle-free graph H and let $G = K_n \setminus H$. We consider a set of vertices $V(G) = A \cup B \cup C$, where $|A| = \{1, \dots, n/3\}$, $B = \{n/3 + 1, \dots, 2n/3\}$ and $C = \{2n/3 + 1, \dots, n\}$. Graph H only have edges between A and B in such a way that, for any $i \in A$ and $j \in B$, we have $ij \in E(H)$ if and only if $i - j > n/3$. It is obvious that H is bipartite, so if we define $G = K_n \setminus H$, we have $\alpha(G) = 2$.

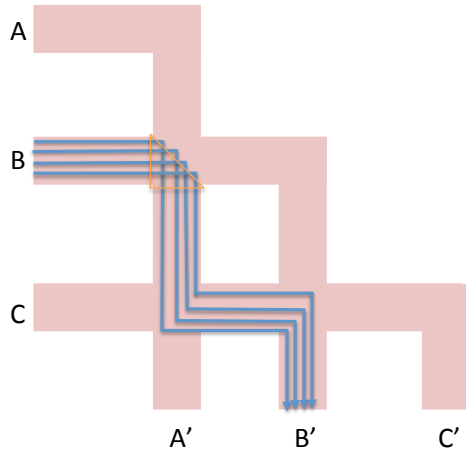


Figure 9: Bad Example for EDP reduction

Now we check that $\text{EDP}(\vec{R}[G]) = n/3$. For each $i \in [n]$, let E_i denote the set of edges on the canonical path from s_i to t_i in $\vec{R}[G]$. For each $i \in B$, we define a path P_i that:

- Start at s_i and go straight until it meets with an edge in $E_{i-n/3}$, at which the path turns downward (the turning is possible because we have an edge $i(i - n/3) \in E(G)$).
- The path goes downward until it meets with an edge in $E_{n-i+n/3}$, at which the path turns again towards the right.
- Once path P_i meets with an edge in E_i again, it takes a turn downward and remains so until it reaches t_i .

Observation A.1. For any $i, j \in B$, $P_i \cap P_j = \emptyset$.