Answering Related Questions

Édouard Bonnet
CNRS, ENS de Lyon, Université Claude Bernard Lyon 1, LIP UMR 5668, 69342 Lyon, France

Abstract

We introduce the meta-problem $\textsc{Sidestep}(\Pi, \mathsf{dist}, d)$ for a problem Π, a metric dist over its inputs, and a map $d : \mathbb{N} \to \mathbb{R}_+ \cup \{\infty\}$. A solution to $\textsc{Sidestep}(\Pi, \mathsf{dist}, d)$ on an input I of Π is a pair $(J, \Pi(J))$ such that $\mathsf{dist}(I, J) \leqslant d(|I|)$ and $\Pi(J)$ is a correct answer to Π on input J. This formalizes the notion of answering a related question (or sidestepping the question), for which we give some practical and theoretical motivations, and compare it to the neighboring concepts of smoothed analysis, planted problems, and edition problems. Informally, we call hardness radius the "largest" d such that $\textsc{Sidestep}(\Pi, \mathsf{dist}, d)$ is NP-hard. This framework calls for establishing the hardness radius of problems Π of interest for the relevant distances dist.

We exemplify it with graph problems and two distances $\mathsf{dist}_\triangle$ and dist_e (the edge edit distance) such that $\mathsf{dist}_\triangle(G, H)$ (resp. $\mathsf{dist}_e(G, H)$) is the maximum degree (resp. number of edges) of the symmetric difference of G and H if these graphs are on the same vertex set, and $+\infty$ otherwise. Thus when solving $\textsc{Sidestep}(\Pi, \mathsf{dist}_\triangle, d)$ (resp. $\textsc{Sidestep}(\Pi, \mathsf{dist}_e, d)$) on an n-vertex input G, acceptably-close graphs H are obtained by XORing G with a graph of maximum degree at most $d(n)$ (resp. having at most $d(n)$ edges). We show that the decision problems Independent Set, Clique, Vertex Cover, Coloring, Clique Cover have hardness radius $n^{\frac{1}{2}-o(1)}$ for $\mathsf{dist}_\triangle$, and $n^{\frac{4}{3}-o(1)}$ for dist_e, that Hamiltonian Cycle (or Hamiltonian Path) has hardness radius 0 for $\mathsf{dist}_\triangle$, and somewhere between $n^{\frac{1}{2}-o(1)}$ and $n/3$ for dist_e, and that Dominating Set has hardness radius $n^{1-o(1)}$ for dist_e. We leave several open questions.

1 Introduction

At the end of a talk or in everyday life, we can be asked three kinds of well-formed questions: *easy* ones, to which we can quickly provide an answer, *hard* ones, which we cannot directly address but are still able to mention something related, and *very hard* ones, for which even the latter is beyond us. In this paper, we formalize the distinction between *hard* and *very hard* questions for computational problems. To agree on the meaning of "*related*" we need to specify a metric on the input space of the problem at hand, as well as a radius—possibly depending on the input size—below which we are acceptably close to the original question. If the radius is the constant zero function, we are in fact solving the original problem. At the other extreme, if the radius is infinite, then we shall simply find a trivially solvable instance; an easy task for most problems. It is then natural to determine how small this radius can be while keeping the meta-problem of "sidestepping the question" tractable. To a first approximation, we call *hardness radius* the supremum of the radius below tractability.

This paper serves as a theoretical basis for *answering related questions* with the following motivations in mind.

First and main motivation. On the one hand, efficiently solving the meta-problem for some task Π (in the regime where the radius allows it) permits to build a customized dataset of graphs each labeled by the adequate answer to Π. Indeed, if one wants the dataset on which Π is solved to be some precise set $\mathcal{D}$ or to be sampled according to a chosen distribution, one at least gets a dataset $\mathcal{D}'$ that is pointwise close to the desired target. This can be useful to create benchmarks, curate instances for a programming competition, or enable supervised learning for Π. Depending on the exact use case, we may ask for more properties from the instances of $\mathcal{D}'$. Ideally, solving Π on the inputs of $\mathcal{D}'$ should remain as challenging as on general inputs. Or, at the very least, $\mathcal{D}'$ should not contain too many *trivial* instances. This can in principle be favored by tweaking the distance function between inputs.

Second motivation. On the other hand, the magnitude of the hardness radius somehow informs, at least for relevant input metrics, on the degree of difficulty of a problem. For instance, for a chosen distance between inputs of NP-hard graph problems, this draws a hierarchy of *increasingly hard* problems. We wish to explore how this hierarchy[1] compares with those obtained through the prisms of approximability, parameterized complexity, and smoothed analysis.

We will later elaborate on our motivations, and how our new meta-problem and the hardness radius relate to well-established concepts: smoothed analysis, planted problems, and edge modification problems. We now get more formal. Given a problem Π, a distance dist over its inputs, and a non-decreasing function $d : \mathbb{N} \to \mathbb{R}_+ \cup \{\infty\}$, we introduce the following problem.

> SIDESTEP(Π, dist, d)
> **Input:** An input I of Π.
> **Output:** A pair $(J, \Pi(J))$ with $\mathsf{dist}(I, J) \leqslant d(|I|)$.

In the above, $|I|$ denotes the *size* of input I, defined as a natural number. When the main part of I is a graph, $|I|$ denotes its number of vertices. Note that Π can be a *decision* or a *function* problem. And in both cases SIDESTEP(Π, dist, d) is itself a function problem. Thus $\Pi(J)$ lies in $\{\mathsf{true}, \mathsf{false}\}$ if Π is a decision problem, and it is the correct solution to Π on input J, or a special symbol nil indicating the absence of solution, when Π is a function problem. Often, there may be several correct solutions. We denote the set of all correct solutions to Π on J by $\mathrm{Sol}_\Pi(J)$, and $\Pi(J)$ should be thought as any fixed element of $\mathrm{Sol}_\Pi(J)$. We make the slight abuse of notation of writing the expression $d(n)$ instead of the function d in the third field of SIDESTEP.

Considered metrics. We will exclusively deal with problems Π whose input is a weightless undirected graph, and optionally an integer threshold, and consider two distances between graphs. Both require that two graphs have the same vertex set for the distance between them to be finite. For every two graphs G, H, $\mathsf{dist}_e(G, H) := \infty$ if $V(G) \neq V(H)$ and

$$\mathsf{dist}_e(G, H) := |E(G) \triangle E(H)|,$$

otherwise, where $\triangle$ denotes the symmetric difference. For two graphs G, H on the same vertex set, $V(G)$, we may denote by $G \triangle H$ the graph with vertex set $V(G)$ and edge set $E(G) \triangle E(H)$. We set $\mathsf{dist}_\Delta(G, H) := \infty$ if $V(G) \neq V(H)$ and

$$\mathsf{dist}_\Delta(G, H) := \Delta(G \triangle H),$$

otherwise, where $\Delta(\cdot)$ denotes the maximum degree of its graph argument. If the whole input consists of a graph together with some additional integral, rational, or real parameters (typically serving as thresholds for a graph property), a finite distance between $(G, \overline{k})$ and $(H, \overline{\ell})$ implies that $\overline{k} = \overline{\ell}$. Specifically, for every considered distance dist, $\mathsf{dist}((G, \overline{k}), (H, \overline{\ell})) =$

[1] or *these hierarchies* since more than one input distance can be considered

$\mathsf{dist}(G, H)$ if $\overline{k} = \overline{\ell}$, and $\mathsf{dist}((G, \overline{k}), (H, \overline{\ell})) = \infty$ otherwise. In particular, we will often exclude the parameters (when they exist at all) from the expressed distances.

Hardness and tractability radii. A function problem is said NP-*hard* if every problem in NP reduces to it via a polynomial-time (not necessarily many-one) reduction. The *hardness radius* of a problem Π for (or *under*) distance dist is at least some non-negative, non-decreasing function d if Sidestep(Π, dist, d) is NP-hard. Similarly, the *tractability radius* of Π for distance dist is at most some non-negative, non-decreasing function d if Sidestep(Π, dist, d) is in FP, the polynomial-time class for function problems. As the hardness and tractability radii are *functions*, formalizing what they are *equal to* through infimum and supremum would require a total order over maps. However, two non-decreasing functions $f, g : \mathbb{N} \to \mathbb{R}_+$ can be asymptotically incomparable, in the sense that for every $n \in \mathbb{N}$ there are $n_f, n_g \geqslant n$ such that $f(n_f) > g(n_f)$ and $f(n_g) < g(n_g)$. Rather than defining the family of allowed maps, we will leave the definition as is: mere lower and upper bounds.

In some situations, though, we will be able to "pinpoint" more or less precisely the hardness or tractability radius. We may say that Π has for distance dist *hardness radius*

- γ if Sidestep(Π, dist, γ) is NP-hard, and for all $\varepsilon > 0$ such that $\gamma + \varepsilon$ is in the image of dist, Sidestep(Π, dist, $\gamma + \varepsilon$) is in FP,
- $n^{\gamma - o(1)}$ if Sidestep(Π, dist, n^{γ}) is in FP, and for all $\varepsilon \in (0, \gamma]$, Sidestep($\Pi$, dist, $n^{\gamma-\varepsilon}$) is NP-hard.

Note that the hardness and tractability radii can in principle be separated by a "zone" of NP-intermediate problems. While our main distances, dist_{Δ} and dist_e, only take integral values, and we could have given simplified definitions for them, the above supports the general case when the metric is valued in $\mathbb{R}_+ \cup \{\infty\}$.

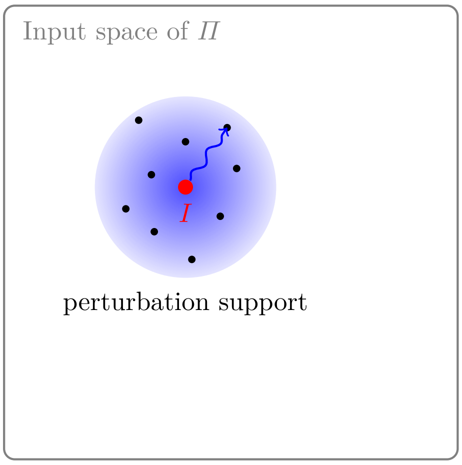
1.1 Related concepts

Here we detail how our meta-problem relates to and differs from well-established concepts.

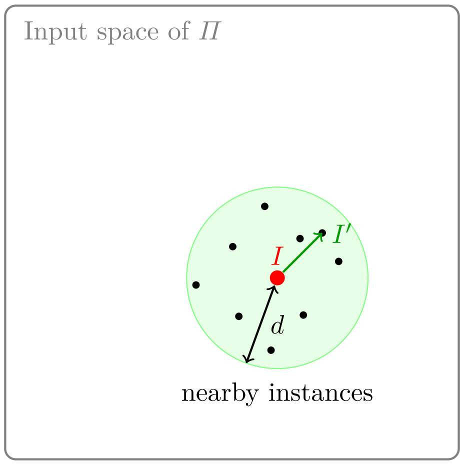
Smoothed analysis. Introduced by Spielman and Teng to explain the practical effectiveness of the simplex algorithm despite its worst-case exponential runtime [14], smoothed analysis has been successful in serving a similar purpose for several algorithms operating over a continuous input space,[2] such as the k-means method [2] or the 2-opt heuristic for the Euclidean TSP [8]; also see [15]. The (worst-case) complexity of Sidestep(Π, dist, d) can be seen as an adversary choosing the worst input I, and our replying the most convenient input I' at distance at most $d(|I|)$ from I, and then solving Π on input I'. Smoothed analysis can be thought as following the above process with the twist that I' is instead chosen by Nature, according to some probability distribution. We refer the interested reader to [14, 15] for formal definitions.

While the aim of smoothed analysis is to model real data, and thus to better describe the practical performance of an algorithm, the main motivation of solving Sidestep(Π, dist, d) lies elsewhere. However, both endeavors are comparable in that they hierarchize NP-hard problems by a finer-grained complexity notion departing from worst-case analysis. In fact, smoothed analysis also has its own "hardness radius" in the supremum of the variance for

[2] hence, in practice, taking rationals or floats as input, which approximate the reals that would ideally constitute the instance



(a) Smoothed analysis.



(b) $\textsc{Sidestep}(\Pi, \mathsf{dist}, d)$.

Figure 1 Comparison of the smoothed analysis of Π and $\textsc{Sidestep}(\Pi, \mathsf{dist}, d)$. The *moves* of the adversary, Nature, and ours are depicted in red, blue, and green, respectively.

which the average-complexity around the worst instance I is at least superpolynomial.[3]

We believe that, at least for problems on weightless graphs, our framework has its merits compared to smoothed analysis. Even if theory has been developed to extend smoothed analysis to algorithms on exclusively[4] discrete inputs [5], the resulting smoothed complexity is less appealing than that of continuous problems or combinatorial problems with a continuously perturbed objective function. Bläser and Manthey [5] consider the random perturbation consisting of XORing the edge set of the input graph with a graph drawn from the Erdős-Rényi model $\mathcal{G}(n, p = \varepsilon)$ with $\varepsilon \in (0, 1/2]$. Applied to k-Coloring, this leads to the perhaps undesired conclusion that this problem is "easy"—as far as smoothed complexity is concerned—for the somewhat unsatisfactory reason that if ε is above a constant ε_0 bounded away from 0, then the perturbed n-vertex graph has cliques of size $\Omega(\log n)$ (like $\mathcal{G}(n, \varepsilon_0)$ itself) hence is a negative instance, whereas if $\varepsilon \leqslant \varepsilon_0$, the allowed time is actually $2^{\Theta(n^2)}$, way above the time the exact exponential algorithms take.

Admittedly, also leveraging that the clique number lower bounds the chromatic number, $\textsc{Sidestep}(k\text{-}\textsc{Coloring}, \mathsf{dist}_\Delta, k-1)$ and $\textsc{Sidestep}(k\text{-}\textsc{Coloring}, \mathsf{dist}_e, \binom{k+1}{2} - k)$ are tractable. One can indeed check that these radius bounds allow to plant a clique of size $k+1$ in a previously connected vertex subset (of size $k+1$). This makes for a rather low hardness radius for a NP-hard problem that is not particularly easy to solve in practice. More unsettlingly, the fact that for any fixed k, most graphs have clique number larger than k should not presume anything on the actual complexity of k-Coloring. This is handier to remediate in our setting: simply tweak the metric by setting an infinite distance between two graphs of distinct clique numbers. (In the next couple of paragraphs, we further incentivize to tweak the distance function for such a problem.) Lastly, the hardness radius is easier to extract in our framework—where it appears naturally—than in that of smooth analysis.

Planted problems. Planted problems, and their main representative, the planted clique

[3] And *smoothed polynomial time* is reached when below the "hardness radius" the variance is low enough that the *allowed time* exceeds that of the brute-force approach.

[4] By that we exclude combinatorial problems with real (or rational) weights that can naturally be continuously perturbed, as in [3, 4].

problem [12], ask to distinguish between a (plain) random graph (usually $\mathcal{G}(n, 1/2)$) and the "same" random graph augmented by some edges forming a solution to an NP-hard problem; be it by adding all the edges on a selected vertex subset (Clique) or the edges of a spanning cycle (Hamiltonian Cycle). The search variant consists of actually reporting this planted solution. In some way, Sidestep(Π, dist, d) is a *planting* problem instead. It is about editing the inputs with features that make them—at least to the *editor*—easily identifiable as positive or as negative instances.

For the sake of our first motivation, it is desirable that the edited instances remain as typical as possible. In particular, it is highly preferable that they are challenging to solve for other parties than the editor. In that respect, the abovementioned suggested solution to Sidestep(k-Coloring, dist_Δ, $k-1$) and Sidestep(k-Coloring, dist_e, $\binom{k+1}{2} - k$) is dissatisfying. As previously observed, a way to not pollute the k-Coloring dataset with trivial instances all containing a clique of size $k+1$ is to tune the metric. A more relevant problem is then Sidestep(k-Coloring, dist, d) where dist sets a large distance (possibly infinite) between two graphs with distinct clique numbers.

Another similarly displeasing solution is that of Theorem 7: when handling Sidestep(Hamiltonian Cycle, dist_e, d) with $d = n/3$, we sometimes create trivially negative instances by disconnecting the graph, and more specifically by isolating a vertex. Again, this can be prevented by modifying the distance function. For example, one could impose that two graphs at a finite distance from each other have the same vertex subsets of connected and 2-connected components.

These more problem-specific metrics are suggestions for potential future work. In this paper, we chose to limit the distances between graphs to dist_Δ and dist_e. Besides, we find interesting the question of the minimum d for which Sidestep(Hamiltonian Cycle, dist_e, d) is polynomial-time solvable, that is, in plain English: How many edge editions are necessary and sufficient to be efficiently conclusive on whether a graph is Hamiltonian? We like that our framework encompasses many fairly natural questions like this one, which, to our knowledge, have not been explored in this exact form.

Edge modification problems. Edge modification problems ask, given a graph G and a non-negative integer k, whether at most k edges of G can be edited (i.e., added or removed) for the resulting graph to land on a particular graph class $\mathcal{C}$. Most often, the parameterized complexity of these problems with respect to k is investigated. There is a recent survey on the topic [6].

Conceptually, this is close to Sidestep($\Pi_\mathcal{C}$, dist_e, k), where $\Pi_\mathcal{C}$ is the recognition problem of class $\mathcal{C}$. Let us highlight the key differences. The most obvious one is that, unlike in the edge modification problem to $\mathcal{C}$, instances of Sidestep($\Pi_\mathcal{C}$, dist_e, k) can be addressed by creating a negative instance. Another important difference is that edge modification problems almost exclusively focus on classes $\mathcal{C}$ that are easy to recognize, such as forests, interval graphs, or cographs. Indeed, if the problem is already NP-hard when $k = 0$, there is not much to be done. On the contrary, in Sidestep($\Pi_\mathcal{C}$, dist_e, k), k is part of the problem *not* of the input, and we try and establish the smallest k such that a polynomial-time algorithm exists. Therefore in our case, $\Pi_\mathcal{C}$ *shall be* hard.

1.2 Our results

In Section 3 we illustrate the new framework with the hamiltonicity problem. We first observe that Sidestep(Hamiltonian Cycle, dist_Δ, 1) is tractable, while Sidestep(Hamiltonian Cycle, dist_Δ, 0) is not, since it coincides with Hamiltonian Cycle.

▶ **Theorem 1.** *Hamiltonian Cycle has hardness radius 0 under* dist_Δ.

For establishing the hardness radius of Hamiltonian Cycle, dist_e proves to be a more challenging distance. We improve the tractability radius from $n/2$ (due to a simple perfect matching argument used to obtain Theorem 1) to $n/3$.

▶ **Theorem 2.** *Sidestep(Hamiltonian Cycle,* $\mathsf{dist}_e, n/3$*) is in* FP.

The algorithm of Theorem 2 uses a path-growing argument reminiscent of some proofs of Dirac's theorem [7] or Ore's theorem [18]. We do not manage to match Theorem 2 with a tight lower bound. However we show that the hardness radius of Hamiltonian Cycle under dist_e is at least $n^{\frac{1}{2}-o(1)}$.

▶ **Theorem 3.** *For any* $\beta > 0$*, Sidestep(Hamiltonian Cycle,* $\mathsf{dist}_e, n^{\frac{1}{2}-\beta}$*) is* NP*-hard.*

Theorem 3 offers a first example of what we call a *robust reduction*, i.e., in the particular case of the metric dist_e, a reduction whose positive (resp. negative) instances can withstand the edition of some number of edges (in that case $n^{\frac{1}{2}-\beta}$). In the proof of Theorem 3, we indeed design a reduction where the produced Hamiltonian graphs cannot be made non-Hamiltonian by removing at most $n^{\frac{1}{2}-\beta}$ edges, and the produced non-Hamiltonian graphs cannot be made Hamiltonian by adding at most $n^{\frac{1}{2}-\beta}$ edges. The former is ensured by applying Ore's theorem on some specific induced subgraphs, while the latter is arranged by piecing together copies of the same 2-connected component.

We further give some evidence (see Theorem 12) that lowering the tractability radius at $n^{1-\beta}$ for some fixed $\beta > 0$ may be challenging since that would imply beating the current polynomial-time best approximation algorithm for Longest Path in Hamiltonian graphs.

Turning our attention to the Dominating Set problem, we prove the following.

▶ **Theorem 4.** *Dominating Set has hardness radius* $n^{1-o(1)}$ *for* dist_e.

The upper bound of $n - 1$ for the tractability radius of Dominating Set under dist_e is trivial, we give a simple argument to bring this upper bound down to n/e, where e is Euler's number.

Section 4 is devoted to establishing the hardness (or tractability) radii for dist_Δ and dist_e of five central graph problems.

▶ **Theorem 5.** *Independent Set, Clique, Vertex Cover, Coloring, and Clique Cover have hardness radius* $n^{\frac{1}{2}-o(1)}$ *under* dist_Δ *and hardness radius* $n^{\frac{4}{3}-o(1)}$ *under* dist_e.

Theorem 5 comprises twenty substatements: two algorithms and two hardness results for five problems. Let us first note that all the problems of Theorem 5 should be thought as decision problems. They take a graph and an integer threshold, and ask for a feasible solution at that threshold; see Section 4 for the definitions. However, since Theorem 5 even works when a feasible solution is required in output (not just a yes/no answer), we deal with this *function variant* of the decision problem.

Solving Sidestep(Max Independent Set, dist, d), where Max Independent Set is the optimization variant, is in principle much harder (see Section 1.3) than solving Sidestep(Independent Set, dist, d). In the latter, one can edit the input graph to be clearly below or clearly above the threshold. In the former, whichever nearby graph ends up being selected for the output, its independence number should be precisely known.

We start by establishing Theorem 5 for Independent Set. The algorithms are simple win-wins planting witnesses of low or of high independence number. The matching lower bounds are based on the famously high inapproximability of Max Independent Set [11, 17]. We then show that a restricted form of polynomial-time reductions, namely polynomial-time isomorphisms both preserving the input size and the distance between inputs, transfers the hardness radius; see Theorem 21 for a precise statement. As a consequence, we obtain Theorem 5 for Clique and Vertex Cover.

The algorithms for Independent Set almost readily work for Clique Cover, and hence for Coloring via Theorem 21. We finally adapt the hardness results for Coloring (and hence for Clique Cover via Theorem 21). This requires a bit more work, especially for the distance dist_e where we leverage a classical upper bound of the chromatic number of m-edge graphs (Observation 25), and the Cauchy–Schwarz inequality.

As far as approximation algorithms and parameterized complexity are concerned, Vertex Cover, which admits fixed-parameter tractable algorithms and polynomial-time 2-approximation algorithms, is "much more tractable" than the other problems of Theorem 5. On the contrary, the complexity measure of ascending hardness radius (for, say, dist_e) groups Vertex Cover together with Independent Set and Coloring, higher up than Dominating Set and Hamiltonian Cycle.

1.3 Open questions and potential future work

Section 3 suggests the following question.

▶ **Question 1.** *What is the hardness radius of Hamiltonian Cycle under* dist_e*?*

Toward answering Question 1, is it true that for any positive k, Sidestep(Hamiltonian Cycle, $\mathsf{dist}_e, n/k$) is in FP? We ask the same question for Dominating Set.

▶ **Question 2.** *For any positive k, is Sidestep(Dominating Set, $\mathsf{dist}_e, n/k$) in* FP*?*

In Theorem 13, we show Question 2 for $k = e$.

▶ **Question 3.** *What is the hardness radius of Dominating Set under* dist_Δ*?*

It is easy to see that for every integer $s > 0$, Sidestep(Dominating Set, $\mathsf{dist}_\Delta, n/s$) is in FP. On inputs (G, k) with $k \leqslant s$, use the brute-force approach to decide if G has a dominating set of size k. When instead $k > s$, pick any k vertices and make each adjacent to $n/k < n/s$ distinct vertices; the original k vertices now form a dominating set.

▶ **Question 4.** *What is the hardness radius of Max Clique under* dist_Δ *and under* dist_e*?*

More generally, we ask Question 4 for the optimization versions of the problems of Theorem 5.

The NP-hard Triangle Partition problem inputs an n-vertex graph G with n divisible by 3, and asks for a partition of $V(G)$ into $n/3$ triangles. It is straightforward that Sidestep(Triangle Partition, $\mathsf{dist}_\Delta, 2$) is in FP.

▶ **Question 5.** *Is Sidestep(Triangle Partition, $\mathsf{dist}_\Delta, 1$) in* FP*?*

It is easy to see that Sidestep(Triangle Partition, $\mathsf{dist}_e, 2n/3$) is in FP. How much can the tractability radius be pushed?

▶ **Question 6.** *What is the hardness radius of Triangle Partition under* dist_e*?*

For some applications when Π is a decision problem, one may prefer the one-sided variants of Sidestep(Π, dist, d), where one forces nearby instances to always be positive or always be negative.

SIDESTEP-POS($\mathit{\Pi}, \mathsf{dist}, d$)
Input: An input I of $\mathit{\Pi}$.
Output: A positive instance J for $\mathit{\Pi}$ with $\mathsf{dist}(I, J) \leqslant d(|I|)$.

SIDESTEP-NEG($\mathit{\Pi}, \mathsf{dist}, d$)
Input: An input I of $\mathit{\Pi}$.
Output: A negative instance J for $\mathit{\Pi}$ with $\mathsf{dist}(I, J) \leqslant d(|I|)$.

These problems are one step closer to edition problems, especially SIDESTEP-POS($\mathit{\Pi}, \mathsf{dist}, d$), but still the settings are somewhat distinct. One can ask for hardness/tractability radii for these one-sided variants.

Finally, even if we only consider graph problems in the current paper, one can tackle the hardness/tractability radii of SIDESTEP($\mathit{\Pi}, \mathsf{dist}, d$) for formula, hypergraph, string, or geometric problems (to name a few) with the relevant distances dist between inputs. Besides, the chosen dividing line to define these radii need not be FP vs NP-hard.

2 Graph and set notation and definitions

If i and j two integers, we denote by $[i, j]$ the set of integers that are at least i and at most j. For every integer i, $[i]$ is a shorthand for $[1, i]$.

We denote by $V(G)$ and $E(G)$ the vertex and the edge set, respectively, of a graph G. If G is a graph and $S \subseteq V(G)$, we denote by $G[S]$ the subgraph of G induced by S, and use $G - S$ as a short-hand for $G[V(G) \setminus S]$. We denote the open and closed neighborhoods of a vertex v in G by $N_G(v)$ and $N_G[v]$, respectively. For $S \subseteq V(G)$, we set $N_G(S) := \bigcup_{v \in S} N_G(v) \setminus S$ and $N_G[S] := N_G(S) \cup S$. A *dominating set* of G is a subset $S \subseteq V(G)$ such that $N_G[S] = V(G)$. We say that $v \in V(G)$ (resp. $S \subseteq V(G)$) dominates $X \subseteq V(G)$ if $X \subseteq N_G[v]$ (resp. $X \subseteq N_G[S]$). In every notation with a graph subscript, we may omit it if the graph is clear from the context.

If $e \in E(G)$, we denote by $G - e$ the graph G deprived of edge e, but the endpoints of e remain. More generally, if $F \subseteq E(G)$, $G - F$ is the graph obtained from G by removing all the edges of F (but not their endpoints). We denote by $G + e$ (resp. $G + F$) the graph with vertex set $V(G)$ and edge set $E(G) \cup \{e\}$ (resp. $E(G) \cup F$).

The *clique number* of a graph G, denoted by $\omega(G)$, is the maximum size of a clique of G, i.e., a set of pairwise adjacent vertices. The *independence number* of G, denoted by $\alpha(G)$, can be defined as the clique number of $\overline{G}$, *the complement of* G, which flips edges and non-edges. The *chromatic number* of G, denoted by $\chi(G)$, is the least number of colors required to properly color G, i.e., give a color to each vertex such that no edge has its two endpoints of the same color, or equivalently the least number of parts in a partition of $V(G)$ into independent sets of G. We will often use the simple fact that any graph G of maximum degree $\mathit{\Delta}$ has chromatic number at most $\mathit{\Delta} + 1$, and has an independent set of size at least $|V(G)|/(\mathit{\Delta} + 1)$.

A *vertex cover* of G is a subset $S \subseteq V(G)$ such that every edge of G has at least one endpoint in S. A *Hamiltonian cycle* (resp. *Hamiltonian path*) is a cycle (resp. path) passing through each vertex exactly once. A graph is said *Hamiltonian* if it admits a Hamiltonian cycle. A *perfect matching* of G is a matching of G with $\lfloor |V(G)|/2 \rfloor$ edges.

We denote by $A \triangle B$ the *symmetric difference* of A and B, i.e., $(A \setminus B) \cup (B \setminus A)$. If two graphs G, H have the same vertex set, then $G \triangle H$ denotes the graph with vertex set

$V(G) = V(H)$ and edge set $E(G) \triangle E(H)$. A useful simple fact about the symmetric difference is that $A = B \triangle C$ implies $A \triangle B = C$. This holds when A, B, C are sets or graphs.

3 Hamiltonicity and Domination

We treat Hamiltonian Cycle (and similarly Hamiltonian Path) as a function problem. In particular, an algorithm for Sidestep(Hamiltonian Cycle, dist, d) returns, on input G, a pair (G', C) where C is a Hamiltonian cycle of G', or (G', nil) if G' is not Hamiltonian.

We first observe that the hardness radius of Hamiltonian Cycle under $\mathsf{dist}_\triangle$ is 0. (The same holds with some small adjustments for Hamiltonian Path.) As the distance $\mathsf{dist}_\triangle$ only takes integral values, we shall simply show that Sidestep(Hamiltonian Cycle, $\mathsf{dist}_\triangle, 1$) is tractable. Indeed, Sidestep(Π, dist, 0) is equivalent to Π for every problem Π and distance dist; hence Sidestep(Hamiltonian Cycle, $\mathsf{dist}_\triangle, 0$) is NP-hard.

▶ **Theorem 6.** *Sidestep(Hamiltonian Cycle, $\mathsf{dist}_\triangle, 1$) is in* FP.

Proof. If G is disconnected, simply output (G, nil) as a disconnected graph cannot be Hamiltonian. We now assume that G is connected. Compute a perfect matching of G in polynomial time. If this step reports that G has no perfect matching, again output (G, nil). Indeed a Hamiltonian cycle yields a perfect matching by taking every other edge on the cycle.

We now assume that G is connected, and that we have found a perfect matching $M := \{u_1v_1, u_2v_2, \ldots, u_pv_p\}$ of G. If $V(G)$ is odd, we further assume without loss of generality (as G is connected) that the unique vertex u of G outside M is adjacent to u_1. We build the following graph J: Start with the edgeless graph on vertex set $V(G)$, and for every $i \in [p-1]$ if $v_iu_{i+1} \notin E(G)$ then add the edge v_iu_{i+1} to $E(J)$. Finally if $V(G)$ is odd and $v_pu \notin E(G)$, add v_pu to $E(J)$. If $V(G)$ is even and $v_pu_1 \notin E(G)$, add v_pu_1 to $E(J)$.

We set $H := G \triangle J$, thus $G \triangle H = J$. It can be observed that J has maximum degree at most 1. Therefore $\mathsf{dist}_\triangle(G, H) \leqslant 1$. And we can output $(H, (u)u_1v_1u_2v_2 \ldots u_pv_p)$. ◀

Note that the proof of Theorem 6 shows that Sidestep(Hamiltonian Cycle, $\mathsf{dist}_e, n/2$) is in FP. We now see that we can improve on the $n/2$ bound.

▶ **Theorem 7.** *Sidestep(Hamiltonian Cycle, $\mathsf{dist}_e, n/3$) is in* FP.

Proof. Greedily grow a path P in the n-vertex input G. That is, while an endpoint of P has a neighbor outside $V(P)$, add this neighbor to the path. When this process ends, we get a path $P = v_1v_2 \ldots v_h$. If $h = n$, we output $(G + v_1v_h, v_1v_2 \ldots v_h)$. Similarly, if $|V(G) \setminus V(P)| = n - h$ is smaller than $n/3$, it is easy to add at most $n/3$ edges to G to make it Hamiltonian. We therefore assume that $h \leqslant 2n/3$, and further assume, as before, that G is connected.

We will now show that either at least one endpoint v_1, v_h has degree at most $h/2 \leqslant n/3$, or we can find a strictly longer path P' and restart the argument. First, we observe that v_1 and v_h have, by construction, no neighbors in $V(G) \setminus V(P)$ (otherwise P could be extended).

▷ Claim 8. If both v_1v_{i+1}, v_iv_h are edges of G, then one can effectively find in G a path on $h+1$ vertices.

Proof: As G is connected (and P is not a Hamiltonian path), there is a vertex $v_j \in V(P)$ with a neighbor $v \in V(G) \setminus V(G)$. By symmetry, we can assume that $1 \leqslant j \leqslant i$. Then, $vv_jv_{j+1} \ldots v_{i-1}v_iv_hv_{h-1} \ldots v_{i+2}v_{i+1}v_1v_2 \ldots v_{j-2}v_{j-1}$ is the desired P' on $h+1$ vertices. ◊

While Claim 8 applies, we proceed with the longer path P' (this will happen fewer than n times). If it does not apply, then the sum of the degrees of v_1 and v_h is upper bounded by h. Hence one of v_1, v_h has degree at most $h/2 \leqslant n/3$. In this case, we isolate v_1 or v_h by at most $n/3$ edge deletions, thereby producing a clearly negative instance. ◀

We next see how to make robust reductions for HAMILTONIAN CYCLE, which can essentially sustain a square root number of edge editions (in the number of vertices). There is still some room between this lower bound and the upper bound of Theorem 7. We will later explain why a truly sublinear upper bound might be difficult to obtain.

▶ **Theorem 9.** *For any $\beta > 0$, SIDESTEP(HAMILTONIAN CYCLE, $\mathsf{dist}_e, n^{\frac{1}{2}-\beta}$) is NP-hard.*

Proof. We reduce from HAMILTONIAN PATH on subcubic graphs with two vertices of degree 1. In particular, any Hamiltonian path has to have these vertices as endpoints. This problem is NP-hard; see for instance [10, 1], which in fact proves that HAMILTONIAN *CYCLE* is NP-complete in subcubic graphs. In the latter paper—which shows a stronger result—the constructed graphs have two adjacent vertices of degree 2. Thus the edge between them can be removed to show the NP-hardness of our Hamiltonian path problem.

Let H be an ν-vertex subcubic graph, and $s, t \in V(H)$ be its two vertices of degree 1. We build an n-vertex instance G of SIDESTEP(HAMILTONIAN CYCLE, $\mathsf{dist}_e, n^{\frac{1}{2}-\beta}$).

Construction of G. Let $q := \lfloor n^{\frac{1}{2}-\beta} \rfloor + 1$, and make $2q$ copies $H_1, \ldots, H_{2q}$ of H. For every $i \in [2q]$, denote by s_i, t_i the copies of s, t, respectively, in H_i. Identify t_i and s_{i+1} for every $i \in [2q-1]$, and s_1 and t_{2q}. Let us call G' the graph obtained at this stage.

Now turn every vertex u in $V(G') \setminus \bigcup_{i \in [2q]} \{s_i, t_i\}$ (actually $\bigcup_{i \in [2q]} \{s_i, t_i\} = \bigcup_{i \in [2q]} \{s_i\}$) into a clique C_u of size $q+2$. The vertices s_i and t_i remain single vertices. For every edge $uv \in E(G')$ not incident to an s_i or t_i, add a new vertex x_{uv} ($= x_{vu}$). Make x_{uv} fully adjacent to C_u and to C_v (there is no edge between C_u and C_v). For every $i \in [2q]$, let s'_i (resp. t'_i) be the unique neighbor of s_i (resp. of t_i) in H_i. Make s_i fully adjacent to $C_{s'_i}$, and t_i fully adjacent to $C_{t'_i}$. For convenience, we may use $x_{s_i s'_i}, x_{t_i t'_i}$ as aliases for s_i, t_i, respectively. This finishes the construction of G.

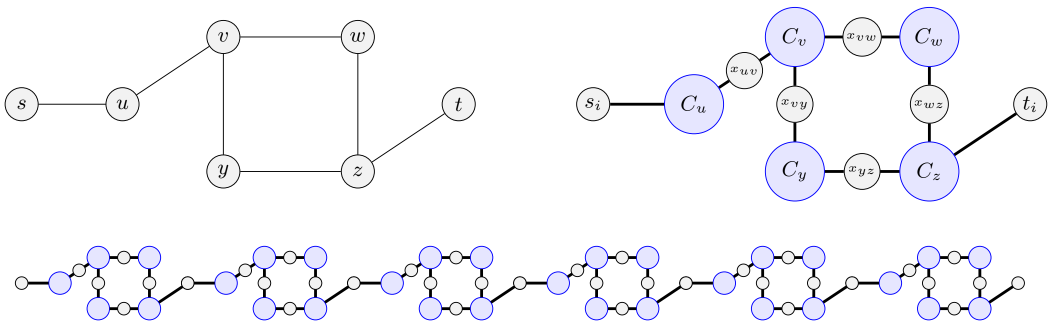


Figure 2 Top left: A subcubic graph H with two degree-1 vertices s and t. Top right: The graph isomorphic to each G_i. Larger blue vertices are cliques of size $q+2$, and edges incident to them go toward all their vertices. Bottom: The obtained graph G with $q = 3$, where the leftmost vertex s_1 and the rightmost vertex t_{2q} are in fact the same vertex.

We denote by G_i the induced subgraph of G derived from H_i. Observe that

$$|V(G)| = 2q((|V(H)| - 2)(q+2) + |E(H)| - 1) = \Theta(q^2 \nu),$$

thus $n = \Theta(n^{1-2\beta} \nu)$ and $n = \Theta(\nu^{\frac{1}{2\beta}})$. As $\beta > 0$ is a fixed constant, this defines a polynomial-time reduction. See Figure 2 for illustrations.

Correctness. The reduction relies on the following two claims.

▷ Claim 10. If H admits an s–t Hamiltonian path, then deleting (equivalently, editing) fewer than q edges in G cannot make it non-Hamiltonian.

PROOF: Let G^- be any graph obtained from G by deleting at most $q-1$ edges. We show the following two properties that for every $x_{uv} \neq x_{vw} \in V(G)$

- there is a path in G^-, starting at x_{uv}, ending at x_{vw}, and whose vertex set is $C_v \cup \{x_{uv}, x_{vw}\}$, and
- if v has a third neighbor y (in its copy H_i), then there is a path in G^-, starting at x_{uv}, ending at x_{vw}, and whose vertex set is $C_v \cup \{x_{uv}, x_{vw}, x_{vy}\}$.

We can then conclude by mimicking an s–t Hamiltonian path P of H in every G_i. More precisely, we start at $s_i = x_{s_i s'_i}$ (for i going from 1 to $2q$), and obey the following rules, where P_i is the path P in copy H_i. When in x_{uv}, if the successor of v along P_i is w, we continue in G_i with the x_{uv}–x_{vw} path of the second item if v has a third neighbor y in H_i and x_{vy} was not traversed yet. We instead go with the x_{uv}–x_{vw} path of the first item, otherwise.

First item. Recall that Ore's theorem [18] is that every n-vertex simple graph such that the sum of the degrees of any two non-adjacent vertices is at least n is Hamiltonian. This implies that $G^-[C_v]$ is Hamiltonian. Indeed, a non-adjacent pair of vertices in $G^-[C_v]$ has combined degree at least $2q-(q-2)=q+2$. We fix a Hamiltonian cycle C of $G^-[C_v]$.

The bipartite graph $G^-[\{x_{uv}, x_{vw}\}, C_v]$ has at least $2(q+2)-(q-1)=q+5$ edges. This ensures that there are two consecutive vertices u', w' along C such that $x_{uv}u', x_{vw}w'$ are edges of $G^-[C_v \cup \{x_{uv}, x_{vw}\}]$. If not, $G^-[\{x_{uv}, x_{vw}\}, C_v]$ would have at most $q+2$ edges. The edge $x_{uv}u'$, the $(q+2)$-vertex path along C from u' to w', and the edge $x_{vw}w'$ define our desired spanning path.

Second item. Observe that $C_v \cup \{x_{vy}\}$ is a clique on $q+3$ vertices. Again by Ore's theorem there is a Hamiltonian cycle C in $G^-[C_v \cup \{x_{vy}\}]$. In G, vertices x_{uv}, x_{vw} both have $q+2$ neighbors in C. Let $s \geqslant 3$ be the number of neighbors of x_{uv} in C remaining in G^-; hence $q+2-s$ edges between x_{uv} and C_v are removed in G^-. These s neighbors have themselves a combined neighborhood N in $C - \{x_{vy}\}$ of size at least $s-1$. As at most $s-3$ edges incident to x_{vw} can be deleted, at least one vertex of N is still adjacent to x_{vw}. We make the x_{uv}–x_{vw} Hamiltonian path in $G^-[C_v \cup \{x_{uv}, x_{vw}, x_{vy}\}]$ as in the previous paragraph. ◊

▷ Claim 11. If H has no s–t Hamiltonian path, then adding (equivalently, editing) fewer than q edges in G cannot make it Hamiltonian.

PROOF: Let G^+ be any graph obtained from G by adding at most $q-1$ edges. By the pigeonhole principle, there is at least one G_i such that no vertex of $G_i - \{s_i, t_i\}$ is incident to an edge added in G^+. Thus the set $\{s_i, t_i\}$ separates $V(G_i) \setminus \{s_i, t_i\}$ from the rest of G^+. So, for G^+ to be Hamiltonian, it should be that G_i admits an s_i–t_i Hamiltonian path. We shall then just prove that if G_i admits an s_i–t_i Hamiltonian path, then so does H_i (hence H has an s–t Hamiltonian path).

Let $u_1, u_2, \ldots, u_{\nu-2}$ be the vertices of $H_i - \{s_i, t_i\}$ in the order in which the cliques C_u are visited for the first time, in some fixed s_i–t_i Hamiltonian path P of G_i. For each clique C_u, there is a vertex subset X_u of size at most 3 (the up to 3 vertices of the form $x_{u,\bullet}$) that disconnects C_u from the rest of the graph in G_i. This implies that $C_u \cup X_u$ cannot be exited and reentered by P. In particular, this means that $u_1 = s'_i$, $u_{\nu-2} = t'_i$, and for every $j \in [\nu-3]$, $u_j u_{j+1} \in E(H_i)$. Therefore $s_i, u_1, u_2, \ldots, u_{\nu-2}, t_i$ is an s_i–t_i Hamiltonian path in H_i. ◊

Thus, by Claims 10 and 11, if a positive (resp. negative) instance is reported by the algorithm solving SIDESTEP(HAMILTONIAN CYCLE, $\mathsf{dist}_e, n^{\frac{1}{2}-\beta}$), then H has an s–t Hamiltonian path (resp. has no s–t Hamiltonian path). ◀

We observe that pushing the tractability radius for dist_e (from $n/3$) down to $n^{.99}$ requires to improve the current polynomial-time approximation factor for LONGEST PATH in Hamiltonian graphs. The current best polytime algorithm returns paths of length $\Omega(\frac{\log^2 n}{\log^2 \log n})$ in n-vertex Hamiltonian graphs [16]. While better guarantees can be obtained in bounded-degree Hamiltonian graphs [9], for any $\beta > 0$, no polytime algorithm is known to output paths of length n^β even in subcubic Hamiltonian graphs.

▶ **Theorem 12.** *For any $\beta > 0$, a polynomial-time algorithm for SIDESTEP(HAMILTONIAN CYCLE, $\mathsf{dist}_e, n^{1-\beta}$) implies a polynomial-time detection of paths of length $\lfloor n^\beta \rfloor$ in Hamiltonian graphs.*

Proof. We make the same reduction as in Theorem 9 with only two copies. Let H, s, t be as in the previous proof, and ν be the number of vertices of H. Let $q := \lfloor n^{1-\beta} \rfloor + 3$. We build an n-vertex instance G of SIDESTEP(HAMILTONIAN CYCLE, $\mathsf{dist}_e, n^{1-\beta}$) as shown in Figure 2 with cliques C_u of size q and solely two copies H_1, H_2 (and G_1, G_2). So $s_1 = t_2$ and $s_2 = t_1$. We observe that G has $2((|V(H)| - 2)q + |E(H)| - 1) = O(q\nu)$ vertices. This implies that $n = O(\nu^{1/\beta})$, so the reduction is indeed polynomial for any fixed $\beta > 0$.

If the algorithm $\mathcal{A}$ for SIDESTEP(HAMILTONIAN CYCLE, $\mathsf{dist}_e, n^{1-\beta}$) outputs a negative instance (of the form (G', nil)), then we know from the proof of Theorem 9 that H has no s–t Hamiltonian path, thus G is not Hamiltonian; which does not happen within Hamiltonian graphs. The only remaining case is that $\mathcal{A}$ outputs a positive instance (G', C). Among the n edges of C, at most $\lfloor n^{1-\beta} \rfloor$ were added to G. This still leaves at least n^β consecutive edges of G in C. We thus obtain the sought path of length $\lfloor n^\beta \rfloor$. ◀

We now move to the DOMINATING SET problem, where given a graph G and an integer $k \in [0, |V(G)|]$, one is asked for a dominating set of size k, or to output nil if none exists.

▶ **Theorem 13.** *SIDESTEP(DOMINATING SET, $\mathsf{dist}_e, n/e$) is in* FP *(where e is Euler's number).*

Proof. Let (G, k) be any input of SIDESTEP(DOMINATING SET, $\mathsf{dist}_e, n/e$) with $n = |V(G)|$. We first run the known polynomial-time $(1 - \frac{1}{e})$-approximation algorithm (actually the greedy algorithm) for MAX k-COVERAGE [13], the problem of picking k sets from a given list so as to maximize the cardinality of their union, on the closed-neighborhood set system of G, i.e., $\{N_G[v] \ : v \in V(G)\}$.

If this outputs k sets whose union has size less than $(1 - \frac{1}{e})n$, we can safely output $((G, k), \mathsf{nil})$, as this gives a guarantee that no dominating set of size k exists in G. Otherwise, we obtain a set S of k vertices such that $X := V(G) \setminus N_G[S]$ is of size at most $\frac{n}{e}$. Let us fix some $v \in S$. Build the graph H by adding $|X|$ edges to G: the edge vx for every $x \in X$. We can output $((H, k), S)$ as $\mathsf{dist}_e(G, H) = |X| \leqslant n/e$ and S is a dominating set of H of size k. ◀

The *blow-up* operation (i.e, replacing every vertex by a clique) provides a simple robust reduction for DOMINATING SET under dist_e.

▶ **Theorem 14.** *For any $\beta > 0$, SIDESTEP(DOMINATING SET, $\mathsf{dist}_e, n^{1-\beta}$) is* NP*-hard.*

Proof. Let H be an ν-vertex instance of DOMINATING SET. We build the n-vertex graph G by replacing each vertex $v \in V(H)$ by a clique C_v of size $q := 2\lfloor n^{1-\beta} \rfloor + 1$. Thus, two distinct vertices $x \in C_u, y \in C_v$ are adjacent in G whenever $u = v$ or $uv \in E(H)$. We have $n = q\nu$, thus $n = O(\nu^{1/\beta})$. This reduction is polynomial for any fixed $\beta > 0$.

We give to G the same threshold as H, say k. We claim that a polynomial-time algorithm solving SIDESTEP(DOMINATING SET, $\mathsf{dist}_e, n^{1-\beta}$) on input (G, k) would also solve DOMINATING SET on input (H, k). The correctness is given by the following two claims.

▷ Claim 15. If H admits a dominating set of size k, then every graph G' obtained from G by removing at most $\lfloor n^{1-\beta} \rfloor$ edges has a dominating set of size k.

PROOF: Note that for every $v \in V(H)$, there is at least one vertex of C_v, say x_v, not incident to any edge of $G \triangle G'$ (removed edge). This is because $|C_v| = 2\lfloor n^{1-\beta} \rfloor + 1$. Let $S \subseteq V(H)$ be a dominating set of H, and $S' := \{x_v \ : \ v \in S\}$. By construction, $|S'| = |S|$ and x_v dominates in G' the set $\bigcup_{w \in N_H[v]} C_v$. As S is a dominating set of H, S' is a dominating set of G'. ◊

▷ Claim 16. If H has no dominating set of size k, then every graph G' obtained from G by adding at most $\lfloor n^{1-\beta} \rfloor$ edges has no dominating set of size k.

PROOF: Let $S \subseteq V(G')$ be any set of size k. Consider the set $S' := \{v \in V(H) \ : \ S \cap C_v \neq \emptyset\}$ of size at most k. By assumption, S' is not a dominating set of H. So there is a $v \in V(H)$ such that $N_H[v] \cap S' = \emptyset$. By the observation in Claim 15, there is a vertex $x_v \in C_v$ such that x_v is not incident to any edge of $G \triangle G'$ (added edge). Vertex x_v is thus not dominated by S'. ◊

We can conclude as a positive (resp. negative) answer to SIDESTEP(DOMINATING SET, dist_e, $n^{1-\beta}$) on input (G, k) implies a positive (resp. negative) answer to DOMINATING SET on input (H, k). ◀

4 Independent Set, Clique, Vertex Cover, and Coloring

We first show that INDEPENDENT SET has hardness radius $n^{\frac{1}{2}-o(1)}$ under $\mathsf{dist}_\triangle$, and $n^{\frac{4}{3}-o(1)}$ under dist_e. We recall that INDEPENDENT SET inputs a pair (G, k) where G is a graph and k is a non-negative integer, and asks for an independent set of size k in G. We could treat INDEPENDENT SET purely as a decision problem (with output in $\{\mathsf{true}, \mathsf{false}\}$). However all the results presented in this section extend to the setting where the desired output is an actual independent set of size k in G, and nil if none exists.

We start by observing that $\sqrt{n}$ upper bounds the tractability radius for $\mathsf{dist}_\triangle$.

▶ **Theorem 17.** *SIDESTEP(INDEPENDENT SET, $\mathsf{dist}_\triangle, n^{\frac{1}{2}}$) is in* FP.

Proof. Let (G, k) be the input, with $n = |V(G)|$. If $k \leqslant \sqrt{n} + 1$, fix any subset $S \subseteq V(G)$ of size $\lfloor \sqrt{n} \rfloor + 1$. Build the graph H with vertex set $V(G)$ such that $H[S]$ is an independent set, and $H - S = G - S$. By construction, $\mathsf{dist}_\triangle(G, H) \leqslant \sqrt{n}$. We can thus output $((H, k), S)$.

If instead $k > \sqrt{n} + 1$, arbitrarily partition $V(G)$ into $k - 1$ parts $V_1, \ldots, V_{k-1}$ of size $\lfloor \frac{n}{k-1} \rfloor$ or $\lceil \frac{n}{k-1} \rceil$. Build the graph H, obtained from G by turning each V_i into a clique. As $|V_i| - 1 \leqslant \sqrt{n}$, it holds that $\mathsf{dist}_\triangle(G, H) \leqslant \sqrt{n}$. The partition into $k - 1$ cliques $V_1, \ldots, V_{k-1}$ in H implies that $\alpha(H) \leqslant k - 1$. We can thus output $((H, k), \mathsf{nil})$. ◀

Perhaps surprisingly, the easy scheme of Theorem 17 is essentially best possible. Using the known strong inapproximability of MAX INDEPENDENT SET, we show that the sidestepping problem becomes NP-hard as soon as the radius gets only slightly smaller.

▶ **Theorem 18.** *For any $\beta > 0$, $\textsc{Sidestep}(\textsc{Independent Set}, \mathsf{dist}_\Delta, n^{\frac{1}{2}-\beta})$ is* NP*-hard.*

Proof. For any $\varepsilon \in (0, 1/2]$, given n-vertex input graphs G satisfying either $\alpha(G) \leqslant n^\varepsilon$ or $\alpha(G) \geqslant n^{1-\varepsilon}$ it is NP-hard to tell which of the two outcomes holds [11, 17].

Consider inputs $(G, \lfloor\sqrt{n}\rfloor)$ of $\textsc{Sidestep}(\textsc{Independent Set}, \mathsf{dist}_\Delta, n^{\frac{1}{2}-\beta})$ with $n = |V(G)|$ and G satisfies the above promise. We show that if these instances could be solved in polynomial time, this would contradict, under $\mathsf{P} \neq \mathsf{NP}$, the latter hardness-of-approximation result with $\varepsilon := \beta/2$. Assume that the supposed algorithm returns $((H, \lfloor\sqrt{n}\rfloor), S)$ such that $\mathsf{dist}_\Delta(G, H) \leqslant n^{\frac{1}{2}-\beta}$ and S is an independent set of H of size $\lfloor\sqrt{n}\rfloor$. An important observation is that $G[S]$ has maximum degree at most $n^{\frac{1}{2}-\beta}$. Hence, in polynomial time, one can compute a subset $S' \subseteq S$ such that S' is an independent set of G of size at least

$$\frac{\lfloor\sqrt{n}\rfloor}{n^{\frac{1}{2}-\beta}+1} \geqslant \frac{n^\beta}{2} > n^\varepsilon.$$

For the last inequality to hold, we assume that $n > 2^{1/\varepsilon}$, which we can safely do, as otherwise the instance can be solved by the brute-force approach. Therefore S' witnesses that $\alpha(G) \leqslant n^\varepsilon$ does not hold. We can thus report that $\alpha(G) \geqslant n^{1-\varepsilon}$.

Now we assume instead that the algorithm returns $((H, \lfloor\sqrt{n}\rfloor), \mathsf{nil})$ such that $\mathsf{dist}_\Delta(G, H) \leqslant n^{\frac{1}{2}-\beta}$ and $\alpha(H) < \sqrt{n}$. We claim that $\alpha(G) < n^{1-\varepsilon}$. Indeed, symmetrically to what has been previously observed, if G had an independent set S of size at least $n^{1-\frac{\beta}{2}}$, then a $(n^{\frac{1}{2}-\beta}+1)^{-1}$ fraction of S would form an independent set in H. Again, assuming that $n > 2^{1/\varepsilon}$, this would contradict $\alpha(H) < \sqrt{n}$. Hence, we can report that the outcome $\alpha(G) \leqslant n^\varepsilon$ holds. ◀

We can adapt the proof of Theorem 17 to yield an essentially optimal upper bound for distance dist_e. For that, we simply set the demarcation on the independent-set size at $n^{\frac{2}{3}}$ rather than $\sqrt{n}$.

▶ **Theorem 19.** *$\textsc{Sidestep}(\textsc{Independent Set}, \mathsf{dist}_e, n^{\frac{4}{3}})$ is in* FP*.*

Proof. Let (G, k) be the input, with $n = |V(G)|$. If $k \leqslant n^{\frac{2}{3}}$, fix any subset $S \subseteq V(G)$ of size $n^{\frac{2}{3}}$. Remove the at most $\binom{n^{2/3}}{2} \leqslant n^{\frac{4}{3}}$ edges of $G[S]$ from G, and let H be the obtained graph. We can output $((H, k), S)$ since $\mathsf{dist}_e(G, H) \leqslant n^{\frac{4}{3}}$ and S is an independent set of size k in H.

If instead $k > n^{\frac{2}{3}}$, arbitrarily partition $V(G)$ into $\lfloor n^{\frac{2}{3}}\rfloor$ parts $V_1, \ldots, V_{\lceil n^{2/3}\rceil}$ each of size at most $\lceil n^{\frac{1}{3}}\rceil + 1$. Turn each V_i into a clique, and call the obtained graph H. This adds to G at most $\lfloor n^{2/3}\rfloor \cdot \binom{\lceil n^{1/3}\rceil+1}{2} \leqslant n^{\frac{4}{3}}$ edges, for large enough n. Thus we can output $((H, k), \mathsf{nil})$. ◀

Finally we show the counterpart of Theorem 18 for dist_e.

▶ **Theorem 20.** *For any $\beta > 0$, $\textsc{Sidestep}(\textsc{Independent Set}, \mathsf{dist}_e, n^{\frac{4}{3}-\beta})$ is* NP*-hard.*

Proof. As in Theorem 18, we use the hardness of distinguishing, for any $\varepsilon \in (0, 1/2]$, n-vertex graphs G such that $\alpha(G) \leqslant n^\varepsilon$ from those such that $\alpha(G) \geqslant n^{1-\varepsilon}$.

Consider inputs $(G, \lfloor n^{2/3}\rfloor)$ of $\textsc{Sidestep}(\textsc{Independent Set}, \mathsf{dist}_e, n^{\frac{4}{3}-\beta})$ with $n = |V(G)|$ and G satisfies one of the previous outcomes. We show that solving these instances in polynomial time would contradict, unless $\mathsf{P} = \mathsf{NP}$, the above hardness-of-approximation result with $\varepsilon := \beta/4$. Assume that the supposed algorithm returns $((H, \lfloor n^{2/3}\rfloor), S)$ such that $\mathsf{dist}_e(G, H) \leqslant n^{\frac{4}{3}-\beta}$ and S is an independent set of H of size $\lfloor n^{2/3}\rfloor$.

We observe that $G[S]$ has at most $n^{\frac{4}{3}-\beta}$ edges. Hence every (induced) subgraph of $G[S]$ has a vertex of degree at most $2(n^{\frac{4}{3}-\beta})^{1/2} = 2n^{\frac{2}{3}-2\varepsilon}$. Thus $G[S]$ has an independent set of size at least

$$\frac{\lfloor n^{2/3} \rfloor}{2n^{\frac{2}{3}-2\varepsilon}+1} > n^{\varepsilon},$$

for n at least a fixed function of ε. Therefore $\alpha(G) \leqslant n^{\varepsilon}$ cannot hold, so we can conclude that $\alpha(G) \geqslant n^{1-\varepsilon}$.

We now assume that the algorithm returns instead $((H, \lfloor n^{2/3} \rfloor), \mathsf{nil})$ such that $\mathsf{dist}_e(G, H) \leqslant n^{\frac{4}{3}-\beta}$ and $\alpha(H) < \lfloor n^{2/3} \rfloor$. We claim that $\alpha(G) < n^{1-\varepsilon}$. Suppose for the sake of contradiction that G has an independent set S of size $n^{1-\frac{\beta}{4}}$. Note that $H[S]$ has at most $n^{\frac{4}{3}-\beta}$ edges. In particular, $H[S]$ has at most $n^{1-2\varepsilon}$ vertices of degree at least $2n^{\frac{1}{3}-2\varepsilon}$, thus more than $n^{1-\varepsilon} - n^{1-2\varepsilon}$ vertices of degree less than $2n^{\frac{1}{3}-2\varepsilon}$. Therefore $H[S]$ has an independent set of size at least

$$\frac{n^{1-\varepsilon} - n^{1-2\varepsilon}}{2n^{\frac{1}{3}-2\varepsilon}+1} > \frac{1}{10} n^{\frac{2}{3}+\varepsilon} > n^{2/3},$$

for some large enough n. This would contradict that $\alpha(H) < \lfloor n^{2/3} \rfloor$. Hence, we can conclude that $\alpha(G) \leqslant n^{\varepsilon}$ holds. ◀

We now introduce the necessary formalism to show that all the previous results of this section carry over to Clique and Vertex Cover.

A *polynomial-time isomorphism* φ from Π to Π' (where Π and Π' are not necessarily decision problems) is a bijective map from the inputs of Π to the inputs of Π' that is a polynomial-time reduction, with inverse φ^{-1} also computable in polynomial time.[5] Specifically when Π and Π' are function problems, the bijection φ is accompanied by a polynomial-time function ψ, called *output pullback* (of φ), such that for any $x \in \mathrm{Sol}_{\Pi'}(\varphi(I))$, it holds that $\psi(x) \in \mathrm{Sol}_{\Pi}(I)$. Note that if instead Π and Π' are decision problems the function ψ would simply map true to true and false to false, and can remain implicit. We say that φ is *length-preserving* if for every input I of Π, $|\varphi(I)| = |I|$.

▶ **Theorem 21.** *Let Π and Π' be two problems, φ, a length-preserving polynomial-time isomorphism from Π to Π', ψ, its output pullback, dist, a metric on the inputs of Π, and dist', a metric on the inputs of Π' such that for every pair of inputs I, J of Π it holds that $\mathsf{dist}(I, J) = \mathsf{dist}'(\varphi(I), \varphi(J))$.*

For any map d, if Sidestep$(\Pi', \mathsf{dist}', d)$ is in FP, then Sidestep(Π, dist, d) is in FP.

Proof. The polynomial-time algorithm solving Sidestep(Π, dist, d) works as follows. On input I of Π, compute the instance $I' := \varphi(I)$ of Π' in polynomial time. Run the polynomial-time algorithm for Sidestep$(\Pi', \mathsf{dist}', d)$ on input I'. It outputs $(J', \Pi'(J'))$ with $\mathsf{dist}'(I', J') \leqslant d(|I'|)$. Now compute $J := \varphi^{-1}(J')$, and $x := \psi(\Pi'(J'))$. Finally output (J, x). By definition of the output pullback, x is indeed a correct solution to Π on J. Furthermore, $\mathsf{dist}(I, J) = \mathsf{dist}'(\varphi(I), \varphi(J)) = \mathsf{dist}'(I', J') \leqslant d(|I'|) = d(|\varphi(I)|) = d(|I|)$, since φ is length-preserving. ◀

We now use Theorem 21 for Clique and Vertex Cover. Similarly to Independent Set, Clique (resp. Vertex Cover) inputs a pair (G, k) where G is a graph and $k \in [0, |V(G)|]$,

[5] For decision problems, this notion is at the core of the Berman–Hartmanis conjecture stating that there is a polynomial-time isomorphism between any two NP-complete languages.

and a clique of G (resp. vertex cover of G) of size k is expected as output, and nil if none exists.

▶ **Theorem 22.** *Clique and Vertex Cover also have hardness radius $n^{\frac{1}{2}-o(1)}$ for dist_{Δ}, and $n^{\frac{4}{3}-o(1)}$ for dist_e.*

Proof. We start with the case of Clique. The map $\varphi : (G,k) \mapsto (\overline{G},k)$, where $\overline{G}$ is the complement of graph G and k is a non-negative integer, is a length-preserving polynomial-time isomorphism from Clique to Independent Set and from Independent Set to Clique (as φ is an involution), and output pullback ψ defined as the identity map. Indeed $|V(G)| = |V(\overline{G})|$, $\overline{\overline{G}} = G$, and an independent set in $\overline{G}$ is a clique in G. Moreover $G \triangle H = \overline{G} \triangle \overline{H}$, so $\mathsf{dist}((G,k),(H,k')) = \mathsf{dist}(\varphi(G,k),\varphi(H,k'))$ for $\mathsf{dist} \in \{\mathsf{dist}_e, \mathsf{dist}_{\Delta}\}$.

Thus the assumptions of Theorem 21 are met by φ and ψ for

$$(\Pi, \Pi') \in \{(\text{Clique}, \text{Independent Set}), (\text{Independent Set}, \text{Clique})\}$$

and $\mathsf{dist} = \mathsf{dist}' \in \{\mathsf{dist}_e, \mathsf{dist}_{\Delta}\}$. Hence Theorems 17, 19, and 21 imply that Sidestep(Clique, $\mathsf{dist}_{\Delta}, n^{\frac{1}{2}}$) and Sidestep(Clique, $\mathsf{dist}_e, n^{\frac{4}{3}}$) are in FP. Now considering the reduction from Independent Set to Clique, Theorems 18, 20, and 21 imply that, for any $\beta > 0$, Sidestep(Clique, $\mathsf{dist}_{\Delta}, n^{\frac{1}{2}-\beta}$) and Sidestep(Clique, $\mathsf{dist}_e, n^{\frac{4}{3}-\beta}$) are NP-hard.

We now deal with Vertex Cover. We define the involutive map $\varphi : (G,k) \mapsto (G, |V(G)| - k)$. Map φ is a length-preserving polynomial-time isomorphism from Vertex Cover to Independent Set and from Independent Set to Vertex Cover, with output pullback ψ defined as the involution $X \mapsto V(G) \setminus X$. Indeed X is a vertex cover of G if and only if $V(G) \setminus X$ is an independent set of G. We have that

$$\mathsf{dist}((G,k),(H,k')) = \mathsf{dist}((G,|V(G)|-k),(H,|V(H)|-k')) = \mathsf{dist}(\varphi(G,k),\varphi(H,k'))$$

for $\mathsf{dist} \in \{\mathsf{dist}_e, \mathsf{dist}_{\Delta}\}$. This holds since these distances are infinite if $V(G) \neq V(H)$ or $k \neq k'$, and are equal to $\mathsf{dist}(G,H)$ otherwise.

Thus Theorems 17, 19, and 21 imply that Sidestep(Vertex Cover, $\mathsf{dist}_{\Delta}, n^{\frac{1}{2}}$) and Sidestep(Vertex Cover, $\mathsf{dist}_e, n^{\frac{4}{3}}$) are in FP, whereas Theorems 18, 20, and 21 imply that, for any $\beta > 0$, Sidestep(Vertex Cover, $\mathsf{dist}_{\Delta}, n^{\frac{1}{2}-\beta}$) and Sidestep(Vertex Cover, $\mathsf{dist}_e, n^{\frac{4}{3}-\beta}$) are NP-hard. ◀

We finish this section with the case of Coloring and Clique Cover.

We first observe that Theorems 17 and 19 also hold for Coloring, which takes as input a graph G and an integer $k \in [0, |V(G)|]$, and asks for a partition of $V(G)$ into k independent sets if one exists, and to output nil otherwise. To do so, one can adapt these theorems for the Clique Cover problem, which boils down to Coloring in the complement graph, and invoke Theorem 21. Indeed, in the algorithms of Theorems 17 and 19, the used witnesses of small independence number (negative instances for Independent Set) are vertex-partitions into cliques (positive instances for Clique Cover), whereas a large independent set (positive instances for Independent Set) is always a witness that the graph needs a large number of cliques to be vertex-partitioned (negative instances for Clique Cover).

▶ **Theorem 23.** *For any $\Pi \in \{$Coloring, Clique Cover$\}$, Sidestep($\Pi, \mathsf{dist}_{\Delta}, n^{\frac{1}{2}}$) and Sidestep($\Pi, \mathsf{dist}_e, n^{\frac{4}{3}}$) are in* FP.

Theorems 18 and 20 also hold for Coloring (and Clique Cover), but their proofs require a bit more adjustments.

▶ **Theorem 24.** *For any* $\beta > 0$, *Sidestep*$(\textsc{Coloring}, \mathsf{dist}_\Delta, n^{\frac{1}{2}-\beta})$ *is* NP*-hard.*

Proof. For any $\varepsilon \in (0, 1/2]$, given n-vertex input graphs G satisfying either $\chi(G) \leqslant n^\varepsilon$ or $\chi(G) \geqslant n^{1-\varepsilon}$, it is NP-hard to tell which of the two outcomes holds [11, 17]. Consider inputs $(G, \lfloor\sqrt{n}\rfloor)$ of $\textsc{Sidestep}(\textsc{Coloring}, \mathsf{dist}_\Delta, n^{\frac{1}{2}-\beta})$ with $n = |V(G)|$ and G satisfies that $\chi(G) \leqslant n^\varepsilon$ or $\chi(G) \geqslant n^{1-\varepsilon}$. We show that a polynomial-time algorithm for these instances would contradict, unless P = NP, the above hardness-of-approximation result with $\varepsilon := \beta/2$.

Assume that the supposed algorithm returns $((H, \lfloor\sqrt{n}\rfloor), \mathcal{P})$ such that $\mathsf{dist}_\Delta(G, H) \leqslant n^{\frac{1}{2}-\beta}$ and $\mathcal{P}$ is a partition of $V(H)$ into $\lfloor\sqrt{n}\rfloor$ independent sets of H. As $G' := G - E(H)$ (the graph G deprived of the edges of H) has maximum degree at most $n^{\frac{1}{2}-\beta}$, there is a partition $\mathcal{Q}$ of $V(G') = V(H)$ into $n^{\frac{1}{2}-\beta} + 1$ independent sets of G'. Partition $\mathcal{Q}$ refines $\mathcal{P}$ (split each part of $\mathcal{P}$ into its intersections with parts of $\mathcal{Q}$) into a partition $\mathcal{P}'$ of $V(G) = V(H)$ with at most $\lfloor\sqrt{n}\rfloor \cdot (n^{\frac{1}{2}-\beta} + 1)$ parts. By construction, each part of $\mathcal{P}'$ is an independent set of G. Moreover $\lfloor\sqrt{n}\rfloor \cdot (n^{\frac{1}{2}-\beta} + 1) \leqslant n^{1-\beta} + \sqrt{n} < n^{1-\varepsilon}$ for large enough n. Thus we can report that $\chi(G) \leqslant n^\varepsilon$.

Now we assume instead that the algorithm returns $((H, \lfloor\sqrt{n}\rfloor), \mathsf{nil})$ such that $\mathsf{dist}_\Delta(G, H) \leqslant n^{\frac{1}{2}-\beta}$ and $\chi(H) > \sqrt{n}$. We claim that $\chi(G) > n^\varepsilon$. Assume, for the sake of contradiction, that there is a partition $\mathcal{P}$ of $V(G) = V(H)$ into $\lfloor n^\varepsilon \rfloor$ independent sets of G. As $H' := H - E(G)$ has maximum degree at most $n^{\frac{1}{2}-\beta}$, there is a partition $\mathcal{Q}$ of $V(H') = V(H)$ into $n^{\frac{1}{2}-\beta} + 1$ independent sets of H'. Then the refinement $\mathcal{P}'$ of $\mathcal{P}$ by $\mathcal{Q}$ is a partition of $V(H)$ into at most $\lfloor n^\varepsilon \rfloor \cdot (n^{\frac{1}{2}-\beta} + 1) \leqslant n^{\frac{1}{2}-\beta+\varepsilon} + n^\varepsilon < \sqrt{n}$ independent sets of H; a contradiction to $\chi(H) > \sqrt{n}$. Hence, we can report that $\chi(G) \geqslant n^{1-\varepsilon}$ holds. ◀

We finally adapt Theorem 20 for Coloring. We need the following folklore observation.

▶ **Observation 25.** *Every* m*-edge graph has chromatic number at most* $\lceil\sqrt{2m}\rceil$.

Proof. It is in fact enough to show that any m-edge graph G has a vertex of degree at most $\lceil\sqrt{2m}\rceil - 1$. If all the vertices of G have degree at least $\lceil\sqrt{2m}\rceil$, then $m \geqslant \frac{1}{2}n\lceil\sqrt{2m}\rceil$. But also $n \geqslant \lceil\sqrt{2m}\rceil + 1$ for a vertex to possibly have $\lceil\sqrt{2m}\rceil$ neighbors. So $m \geqslant \frac{1}{2}\sqrt{2m}(\sqrt{2m}+1) > m$; a contradiction. ◀

The main additional difference compared to the adaptation of Theorem 24 is the use of the Cauchy–Schwarz inequality to get better upper bounds on some chromatic numbers.

▶ **Theorem 26.** *For any* $\beta > 0$, *Sidestep*$(\textsc{Coloring}, \mathsf{dist}_e, n^{\frac{4}{3}-\beta})$ *is* NP*-hard.*

Proof. As in Theorem 24, we rely on the hardness of distinguishing, for any $\varepsilon \in (0, 1/2]$, n-vertex graphs G such that $\chi(G) \leqslant n^\varepsilon$ from those such that $\chi(G) \geqslant n^{1-\varepsilon}$. Consider inputs $(G, \lfloor n^{2/3} \rfloor)$ of $\textsc{Sidestep}(\textsc{Coloring}, \mathsf{dist}_e, n^{\frac{4}{3}-\beta})$ with $n = |V(G)|$ and G satisfies $\chi(G) \leqslant n^\varepsilon$ or $\chi(G) \geqslant n^{1-\varepsilon}$. We show that solving these instances in polynomial time would contradict, unless P = NP, the above hardness-of-approximation result with $\varepsilon := \beta/4$.

Assume that the supposed algorithm returns $((H, \lfloor n^{2/3} \rfloor), \mathcal{P})$ such that $\mathsf{dist}_e(G, H) \leqslant n^{\frac{4}{3}-\beta}$ and $\mathcal{P} := \{P_1, \ldots, P_t\}$ is a partition of $V(H)$ into $t := \lfloor n^{2/3} \rfloor$ independent sets of H. Here we need to be more delicate than in the proof of Theorem 24 in upper bounding the chromatic number of G. We will rely on Observation 25 and the Cauchy–Schwarz inequality.

For every $i \in [t]$, let a_i be the number of edges in $G[P_i]$. As $\mathsf{dist}_e(G, H) \leqslant n^{\frac{4}{3}-\beta}$,

$$\sum_{i \in [t]} a_i \leqslant n^{\frac{4}{3}-\beta}.$$

By Observation 25, $G[P_i]$ can be properly colored with at most $\lceil \sqrt{2a_i} \rceil$ colors. Hence the chromatic number of G is at most

$$\sum_{i \in [t]} \lceil \sqrt{2a_i} \rceil \leqslant \sum_{i \in [t]} (\sqrt{2a_i} + 1) \leqslant t + \sqrt{2} \sum_{i \in [t]} \sqrt{a_i}.$$

In $\mathbb{R}^t$, the Cauchy–Schwarz inequality is that

$$\left(\sum_{i \in [t]} x_i y_i \right)^2 \leqslant \left(\sum_{i \in [t]} x_i^2 \right) \left(\sum_{i \in [t]} y_i^2 \right).$$

Setting $x_i := \sqrt{a_i}$ and $y_i := 1$, we get that

$$\chi(G) \leqslant t + \sqrt{2} \sqrt{\left(\sum_{i \in [t]} \sqrt{a_i}^2 \right) \left(\sum_{i \in [t]} 1^2 \right)} = t + \sqrt{2t \left(\sum_{i \in [t]} a_i \right)} \leqslant t + \sqrt{2t \cdot n^{\frac{4}{3} - \beta}}.$$

Thus

$$\chi(G) \leqslant \lfloor n^{2/3} \rfloor + \sqrt{2} \cdot n^{1/3} \cdot n^{\frac{2}{3} - 2\varepsilon} \leqslant \lfloor n^{2/3} \rfloor + \sqrt{2} n^{1 - 2\varepsilon} < n^{1-\varepsilon},$$

for large enough n. And we can conclude that $\chi(G) \leqslant n^{\varepsilon}$.

We now assume that the algorithm returns instead $((H, \lfloor n^{2/3} \rfloor), \mathsf{nil})$ such that $\mathsf{dist}_e(G, H) \leqslant n^{\frac{4}{3} - \beta}$ and $\chi(H) > n^{2/3}$. We claim that $\chi(G) > n^{\varepsilon}$. Suppose for the sake of contradiction that there is a partition $\mathcal{Q} := \{Q_1, \ldots, Q_s\}$ of $V(G)$ into $s := \lfloor n^{\varepsilon} \rfloor$ independent sets of G. For every $i \in [s]$, let b_i be the number of edges in $H[Q_i]$. As $\mathsf{dist}_e(G, H) \leqslant n^{\frac{4}{3} - \beta}$, $\sum_{i \in [s]} b_i \leqslant n^{\frac{4}{3} - \beta}$. By Observation 25, we have $\chi(H) \leqslant \sum_{i \in [s]} \lceil \sqrt{2b_i} \rceil$. A similar application of the Cauchy–Schwarz inequality with $x_i := \sqrt{b_i}$ and $y_i := 1$ yields

$$\chi(H) \leqslant s + \sqrt{2s \cdot n^{\frac{4}{3} - \beta}} \leqslant \lfloor n^{\varepsilon} \rfloor + \sqrt{2} \cdot n^{\varepsilon/2} \cdot n^{\frac{2}{3} - 2\varepsilon} = \lfloor n^{\varepsilon} \rfloor + \sqrt{2} \cdot n^{\frac{2}{3} - \frac{3\varepsilon}{2}} < n^{2/3},$$

for large enough n; a contradiction to $\chi(H) > n^{2/3}$. Hence, we can conclude that $\chi(G) \leqslant n^{1-\varepsilon}$ holds. ◀

One can extend Theorems 24 and 26 to Clique Cover by Theorem 21. This finishes the proof of Theorem 5.

References

1 Esther M. Arkin, Sándor P. Fekete, Kamrul Islam, Henk Meijer, Joseph S. B. Mitchell, Yurai Núñez Rodríguez, Valentin Polishchuk, David Rappaport, and Henry Xiao. Not being (super)thin or solid is hard: A study of grid hamiltonicity. *Comput. Geom.*, 42(6-7):582–605, 2009. URL: `https://doi.org/10.1016/j.comgeo.2008.11.004`, `doi:10.1016/J.COMGEO.2008.11.004`.

2 David Arthur, Bodo Manthey, and Heiko Röglin. Smoothed analysis of the k-means method. *J. ACM*, 58(5):19:1–19:31, 2011. `doi:10.1145/2027216.2027217`.

3 Cyril Banderier, René Beier, and Kurt Mehlhorn. Smoothed analysis of three combinatorial problems. In Branislav Rovan and Peter Vojtás, editors, *Mathematical Foundations of Computer Science 2003, 28th International Symposium, MFCS 2003, Bratislava, Slovakia, August 25-29, 2003, Proceedings*, volume 2747 of *Lecture Notes in Computer Science*, pages 198–207. Springer, 2003. `doi:10.1007/978-3-540-45138-9\_14`.

4 René Beier and Berthold Vöcking. Typical properties of winners and losers in discrete optimization. *SIAM J. Comput.*, 35(4):855–881, 2006. `doi:10.1137/S0097539705447268`.

5 Markus Bläser and Bodo Manthey. Smoothed complexity theory. *ACM Trans. Comput. Theory*, 7(2):6:1–6:21, 2015. doi:10.1145/2656210.

6 Christophe Crespelle, Pål Grønås Drange, Fedor V. Fomin, and Petr A. Golovach. A survey of parameterized algorithms and the complexity of edge modification. *Comput. Sci. Rev.*, 48:100556, 2023. URL: https://doi.org/10.1016/j.cosrev.2023.100556, doi:10.1016/J.COSREV.2023.100556.

7 Gabriel Andrew Dirac. Some theorems on abstract graphs. *Proceedings of the London Mathematical Society*, 3(1):69–81, 1952.

8 Matthias Englert, Heiko Röglin, and Berthold Vöcking. Smoothed analysis of the 2-opt algorithm for the general TSP. *ACM Trans. Algorithms*, 13(1):10:1–10:15, 2016. doi:10.1145/2972953.

9 Tomás Feder, Rajeev Motwani, and Carlos S. Subi. Finding long paths and cycles in sparse hamiltonian graphs. In F. Frances Yao and Eugene M. Luks, editors, *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing, May 21-23, 2000, Portland, OR, USA*, pages 524–529. ACM, 2000. doi:10.1145/335305.335368.

10 M. R. Garey, David S. Johnson, and Larry J. Stockmeyer. Some simplified NP-complete graph problems. *Theor. Comput. Sci.*, 1(3):237–267, 1976. doi:10.1016/0304-3975(76)90059-1.

11 Johan Håstad. Clique is hard to approximate within $n^{1-\varepsilon}$. In *37th Annual Symposium on Foundations of Computer Science, FOCS '96, Burlington, Vermont, USA, 14-16 October, 1996*, pages 627–636. IEEE Computer Society, 1996. doi:10.1109/SFCS.1996.548522.

12 Mark Jerrum. Large cliques elude the metropolis process. *Random Struct. Algorithms*, 3(4):347–360, 1992. URL: https://doi.org/10.1002/rsa.3240030402, doi:10.1002/RSA.3240030402.

13 George L. Nemhauser, Laurence A. Wolsey, and Marshall L. Fisher. An analysis of approximations for maximizing submodular set functions - I. *Math. Program.*, 14(1):265–294, 1978. doi:10.1007/BF01588971.

14 Daniel A. Spielman and Shang-Hua Teng. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. *J. ACM*, 51(3):385–463, 2004. doi:10.1145/990308.990310.

15 Daniel A. Spielman and Shang-Hua Teng. Smoothed analysis: an attempt to explain the behavior of algorithms in practice. *Commun. ACM*, 52(10):76–84, 2009. doi:10.1145/1562764.1562785.

16 Sundar Vishwanathan. An approximation algorithm for finding long paths in hamiltonian graphs. *J. Algorithms*, 50(2):246–256, 2004. doi:10.1016/S0196-6774(03)00093-2.

17 David Zuckerman. Linear degree extractors and the inapproximability of max clique and chromatic number. *Theory Comput.*, 3(1):103–128, 2007. URL: https://doi.org/10.4086/toc.2007.v003a006, doi:10.4086/TOC.2007.V003A006.

18 Øystein Ore. Note on hamilton circuits. *The American Mathematical Monthly*, 67(1):55–55, 1960. URL: http://www.jstor.org/stable/2308928.