H-Free Graphs, Independent Sets, and Subexponential-Time Algorithms*

Gábor Bacsó[1], Dániel Marx[2], and Zsolt Tuza[3]

1 Institute for Computer Science and Control, Hungarian Academy of Sciences, Budapest, Hungary
2 Institute for Computer Science and Control, Hungarian Academy of Sciences, Budapest, Hungary
3 Alfréd Rényi Institute of Mathematics, Budapest, Hungary; and Department of Computer Science and Systems Technology, University of Pannonia, Veszprém, Hungary

Abstract

It is an old open question in algorithmic graph theory to determine the complexity of the Maximum Independent Set problem on P_t-free graphs, that is, on graphs not containing any induced path on t vertices. So far, polynomial-time algorithms are known only for $t \le 5$ [Lokshtanov et al., SODA 2014, pp. 570–581, 2014]. Here we study the existence of subexponential-time algorithms for the problem: by generalizing an earlier result of Randerath and Schiermeyer for $t = 5$ [*Discrete Appl. Math.,* 158 (2010), pp. 1041–1044], we show that for any $t \ge 5$, there is an algorithm for Maximum Independent Set on P_t-free graphs whose running time is subexponential in the number of vertices.

Scattered Set is the generalization of Maximum Independent Set where the vertices of the solution are required to be at distance at least d from each other. We give a complete characterization of those graphs H for which d-Scattered Set on H-free graphs can be solved in time subexponential in the *size of the input* (that is, in the number of vertices plus number of edges):

- If every component of H is a path, then d-Scattered Set on H-free graphs with n vertices and m edges can be solved in time $2^{(n+m)^{1-O(1/|V(H)|)}}$, even if d is part of the input.
- Otherwise, assuming ETH, there is no $2^{o(n+m)}$-time algorithm for d-Scattered Set for any fixed $d \ge 3$ on H-free graphs with n-vertices and m-edges.

1998 ACM Subject Classification F.1.3 Complexity Measures and Classes, F.2.2 Nonnumerical Algorithms and Problems

Keywords and phrases independent set, scattered set, subexponential algorithms, H-free graphs

Digital Object Identifier 10.4230/LIPIcs.IPEC.2016.3

1 Introduction

The Maximum Independent Set problem (MIS, for short) is one of the fundamental problems in discrete optimization. It takes a graph G as input, and asks for the maximum number $\alpha(G)$ of mutually nonadjacent (i.e., independent) vertices in G. On unrestricted input, it is not only NP-hard (its decision version "Is $\alpha(G) \ge k$?" being NP-complete), but

* Research of Gábor Bacsó and Dániel Marx was supported by ERC Starting Grant PARAMTIGHT (No. 280152) and OTKA grant NK105645. Research of Zsolt Tuza was supported by the National Research, Development and Innovation Office – NKFIH under the grant SNN 116095.

APX-hard as well, and, in fact, even not approximable within $O(n^{1-\varepsilon})$ in polynomial time for any $\varepsilon > 0$ unless P=NP, as proved by Zuckerman [21]. For this reason, classes of graphs are of definite interest on which MIS becomes tractable. One direction of this area is to study the complexity of MIS on *H-free graphs*, that means graphs not containing any *induced* subgraph isomorphic to a given graph H.

What do we know about the complexity of MIS on H-free graphs? One the hardness side, it is easy to see that if G' is obtained from G by subdividing each edge with $2t$ new vertices, then $\alpha(G') = \alpha(G) + t|E(G)|$ holds. This can be used to show that MIS is NP-hard on H-free graphs whenever H is not a forest, and also if H contains a tree component with at least two vertices of degree larger than 2 (first observed in [2], see, e.g., [11]). As MIS is known to be NP-hard on graphs of maximum degree at most 3, the case when H contains a vertex of degree at least 4 is also NP-hard.

The only case not covered by the above observations is when every component of H is either a path, or a tree with exactly one degree-3 vertex c with three paths of arbitrary lengths starting from c. Even this collection means infinitely many cases. For decades, on these graphs H only partial results have been obtained, proving polynomial-time solvability in some cases. A classical algorithm of Minty [16] and its corrected form by Sbihi [19] solved the problem when H is a claw (3 paths of length 1 in the model above). This happened in 1980. Much later, in 2004, Alekseev [3] generalized this result by an algorithm for H isomorphic to a fork (2 paths of length 1 and one path of length 2).

Somewhat embarrassingly, even the seemingly easy case of P_t-free graphs is poorly understood (where P_t is the path on t vertices). MIS on P_t-free graphs is not known to be NP-hard for any t; for all we know, it could be polynomial-time solvable for every fixed $t \geq 1$. P_4-free graphs (also known as cographs) have very simple structure, which can be used to solve MIS in way that is very simple, but does not generalize to P_t-free graphs for larger t. In 2010, it was a breakthrough when Randerath and Schiermeyer [17] stated that MIS was solvable in subexponential time, more precisely within $O(C^{n^{1-\varepsilon}})$ for any constants $C > 1$ and $\varepsilon < 1/4$, on P_5-free graphs. Designing an algorithm based on deep results, Lokshtanov [11] finally proved that MIS is polynomial-time solvable on P_5-free graphs. More recently, a *quasipolynomial* ($n^{\log^{O(1)} n}$-time) algorithm was found for P_6-free graphs [13].

In this paper, we explore MIS and some variants on H-free graphs from the viewpoint of *subexponential-time algorithms*. That is, instead of aiming for algorithms with running time $n^{O(1)}$ on n-vertex graphs, we ask if $2^{o(n)}$ algorithms are possible. Our first result shows that there is indeed such an algorithm for P_t-free graphs.

▶ **Theorem 1.** *For every fixed $t \geq 5$,* MIS *on n-vertex P_t-free graphs is subexponential, namely, it can be solved by a $2^{O(n^{1-1/\lfloor t/2 \rfloor + o(1)})}$-time algorithm.*

In particular, for $t = 5$, this improves the result of Randerath and Schiermeyer [17]. The algorithm is based on the obsevation that a connected P_t-free graph always has a high-degree vertex, which can be used for efficient branching. However, the algorithm does not seem to be extendable to H-free graphs where H is the subdivision of a $K_{1,3}$, hence the existence of subexponential-time algorithms on such graphs remains an open question.

Scattered Set (also known under other names such as dispersion or distance-d independent set [14, 20, 1, 18, 6, 9]) is the natural generalization of MIS where the vertices of the solution are required to be at distance at least d from each other; the size of the largest such set will be denoted by $\alpha_d(G)$. We can consider with d being part of the input, or assume that $d \geq 2$ is a fixed constant, in which case we call it d-Scattered Set. Clearly, MIS is exactly the same as 2-Scattered Set. Despite its similarity to MIS, the branching algorithm of Theorem 1 cannot be generalized: we give evidence that there is no subexponential-time algorithm for 3-Scattered Set on P_5-free graphs. For the lower bound, we assume the Exponential-Time Hypothesis (ETH) of Impagliazzo, Paturi, and Zane, which can be informally stated as n-variable 3SAT cannot be solved in $2^{o(n)}$ time (see [7, 12, 10]).

▶ **Theorem 2.** *Assuming ETH, there is no $2^{o(n)}$-time algorithm for d-Scattered Set with $d = 3$ on P_5-free graphs with n vertices.*

In light of the negative result of Theorem 2, we slightly change our objective by aiming for an algorithm that is subexponential in the *size of the input,* that is, in the total number of vertices and edge of the graph G. As the number of edges of G can be up to quadratic in the number of vertices, this is a weaker goal: an algorithm that is subexponential in the number of edges is not necessarily subexponential in the number of vertices. We give a complete characterization when such algorithms are possible for Scattered Set.

▶ **Theorem 3.** *For every fixed graph H, the following holds.*

1. *If every component of H is a path, then d-Scattered Set on H-free graphs with n vertices and m edges can be solved in time $2^{(n+m)^{1-O(1/|V(H)|)}}$, even if d is part of the input.*
2. *Otherwise, assuming ETH, there is no $2^{o(n+m)}$-time algorithm for d-Scattered Set for any fixed $d \geq 3$ on H-free graphs with n-vertices and m-edges.*

The algorithmic side of Theorem 3 is based on the combinatorial observation that the treewidth of P_t-free graphs is sublinear in the number of edges, which means that standard algorithms on bounded-treewidth graphs can be invoked to solve the problem in time subexponential in the number of edges. It has not escaped our notice that this approach is completely generic and could be used for many other problems (e.g., Hamiltonian Cycle, 3-Coloring, ...) where $2^{O(t)} \cdot n^{O(1)}$ or even $2^{t \cdot \log^{O(1)} t} \cdot n^{O(1)}$-time algorithms are known on graphs of treewidth t. For the lower bound part of Theorem 3, we need to examine only two cases: claw-free graphs and C_t-free graphs (where C_t is the cycle on t vertices); the other cases then follow immediately.

The algorithm described in Section 3 implies Theorem 1, while Theorems 2 and 3 are implied by Sections 4 and 5.

2 Preliminaries

This work investigates simple undirected graphs throughout. The vertex set of graph G will be denoted by $V(G)$, the edge set by $E(G)$. When we deal with a fixed graph, we write simply V and E respectively.

A graph is H-free if it does not contain H as an induced subgraph.

A distance-d set (d-scattered) set) in a graph G is a vertex set $S \subseteq V(G)$ such that for every pair of vertices in S, the distance between them is at least d in the graph. For $d = 2$, we obtain the traditional notion of independent set (stable set). For $d > c$, a distance-d set is a distance-c set as well, for example, any distance-d set is independent for $d \geq 2$.

The algorithmic problem Weighted Independent Set is the problem of maximizing the sum of weights in a graph with nonnegative vertex weights w. The maximum is denoted by $\alpha_w(G)$. For a weight w everywhere 1, we obtain the usual problem Independent Set (MIS) with maximum $\alpha(G)$.

Several definitions are used in the literature under the name *subexponential function.* Each of them means some condition: this function (with variable $p > 1$, called the parameter) may not be larger than some bound, depending on p. Here we use two versions, where the bound is of type $exp(o(p))$ and $exp(p^{1-\epsilon})$ respectively, with some $\epsilon > 0$. (Clearly, the second one is the more strict.) Throughout the paper, we state our results emphasizing, which version we mean.

An algorithm A is *subexponential* in parameter $p > 1$ if the number of steps executed by A is a subexponential function of the parameter p. We will use here this notion for graphs, mostly in the following cases: p is the number n of vertices, the number m of edges, or $p = n + m$ (which is considered to be the size of the input generally).

A problem Π is *subexponential* if there exists some *subexponential* algorithm solving Π.

The notation $d_G(x, y)$ and $\mathrm{diam}(G)$ will have the usual meaning. For a vertex x of G, its *radius* $r_G(x)$ is $\max\{d_G(x, y) | y \in V(G)\}$ and for the *radius* of graph G, $r(G) := \min\{r_G(x) | x \in V(G))\}$. $\Delta(G)$ is the maximal degree in G.

P_t (C_t) is the chordless path (cycle) on t vertices.

3 Algorithm for MIS on P_t-free graphs

The method used here will be similar to that of [17]. There a special dominating set is found (applying [5]), here a vertex of small radius will help. More precisely, the algorithm is based on the observation that a connected P_t-free graph always has a high-degree vertex. The following definition formalizes this property.

▶ **Definition 4.** For a fixed real $\delta > 0$ and a natural number n_0, let $\mathcal{C} := \mathcal{C}(n_0, \delta)$ be the class of graphs G with the following property: For every connected induced subgraph G' of G with $k := |V(G')| \geq n_0$, $\Delta(G') \geq k^{\delta}$.

Clearly, each class $\mathcal{C} := \mathcal{C}(n_0, \delta)$ is contained in the class of P_t-free graphs for $t = n_0$. But if we extend $\mathcal{C}$, the result below will be stronger than a statement merely for graphs without some long induced path.

▶ **Definition 5.** For a fixed real $\delta > 0$ and a natural number n_0, let $\mathcal{G} := \mathcal{G}(n_0, \delta)$ be the class of graphs G with the following property: For every connected induced subgraph G' of G having maximum degree at least 3, with $k := |V(G')| \geq n_0$, $\Delta(G') \geq k^{\delta}$.

The following result presents the connection of $P_t$free graphs with the classes above.

▶ **Lemma 6.** *For every $t \geq 5$, every P_t-free graph is in $\mathcal{C}(N_0, \delta)$ (and thus in $\mathcal{G}(N_0, \delta)$ as well) with $\delta = \lfloor t/2 \rfloor^{-1}$ and an appropriate $N_0 = N_0(t)$.*

Proof. Every connected P_t-free graph has radius at most $\mathrm{diam}(G) \leq t - 2$. To obtain stronger constants, we use a result of Erdős, Saks, and Sós [8, Theorem 2.1], which states, in an alternative formulation, that every connected P_t-free graph has radius at most $\lfloor t/2 \rfloor$.[1]

Assuming that G is connected and has maximum degree Δ, the number of vertices at distance i from a vertex c with minimal radius is at most $\Delta \cdot (\Delta - 1)^{i-1}$. Thus, if G is connected, P_t-free, moreover it has n vertices and maximum degree $\Delta = \Delta(G)$, then for any $t \geq 6$, we have

$$n \leq 1 + \Delta \cdot \sum_{i=1}^{r} (\Delta - 1)^{i-1} < \Delta^{\lfloor t/2 \rfloor}, \tag{1}$$

[1] A subset of the present authors [4] established a stronger property which is equivalent to being P_t-free.

Algorithm 1 Algorithm DEGALPHA

Input: a graph G

1. If $|V(G)| = 1$ then $\alpha(G) = 1$.

2. If $|V(G)| > 1$ and G is disconnected:

  a. Determine a connected component G' of G, and set $G'' = G - G'$.

  b. Determine $\alpha(G')$ and $\alpha(G'')$, calling Algorithm DEGALPHA for G' and G'' separately, and write $\alpha(G) = \alpha(G') + \alpha(G'')$.

3. If $|V(G)| > 1$ and G is connected:

  a. Determine a vertex v of maximum degree, $d_G(v) = \Delta(G)$.

  b. $\Delta(G) \leq 2$ then $\alpha(G)$ is the maximal size of independent set in the corresponding path or cycle respectively.

  c. Determine $\alpha(G - v)$ and $\alpha(G - N[v])$ where $N[v]$ is the closed neighborhood of v, calling Algorithm DEGALPHA for $G - v$ and $G - N[v]$ separately, and write $\alpha(G) = \max(\alpha(G - v), \alpha(G - N[v]) + 1)$.

which corresponds to the standard Moore bound (see, e.g., inequality (1) on page 8 of [15]). As a consequence, for $t \geq 6$, we obtain

$$\Delta(G) \geq n^{\lfloor t/2 \rfloor^{-1}} \tag{2}$$

For $t = 5$, we get the slightly weaker bound $n \leq 1 + \Delta + \Delta(\Delta - 1) = \Delta^2 + 1$. However, with additional arguments, we can show that $n \leq \Delta^2$ holds if $\Delta > 2$, thus the statement is true if $n > 5$. (Sketch of the proof: the only way that $n = \Delta^2 + 1$ can hold is when c has exactly Δ neighbors, each of which has exactly $\Delta - 1$ neighbors at distance two from c, and they do not share any of these neighbors. Let u and v be two neighbors of c. If a neighbor $u' \neq c$ of u is nonadjacent to a neighbor $v' \neq c$ of v, then u', u, c, v, v' form an induced P_5. This shows that u' has degree at least $1 + (\Delta - 1)^2$, which is more than Δ if $\Delta > 2$.) ◀

Next we show that subexponential-time algorithms exists for the class $\mathcal{G}(n_0, \delta)$.

▶ Remark. The class $\mathcal{G}(n_0, \delta)$ with appropriate parameters contains non-P_t-free graphs for any t.

▶ **Lemma 7.** *For any fixed real $0 < \delta < 1$ and a natural number n_0, the independent set problem is subexponential (in the strong sense) for the class $\mathcal{G}(n_0, \delta)$, namely, it can be solved by an algorithm executing at most $O(\exp(c(\delta) \cdot n^{1-\delta} \cdot \ln n))$ steps, where $c(\delta)$ is any real constant greater than $\frac{\delta}{1-\delta}$.*

Proof. The conditions lead to a simple exact algorithm solving MIS (see Algorithm 1), which is also the basis for the analysis in [17] (except that here we need not deal with isolated vertices separately) and whose variants also appear in enumeration algorithms for independent sets.

It is a direct consequence of the definitions that Algorithm DEGALPHA properly determines the independence number of G.

Time analysis. We may and will assume that the number n of vertices is larger than a suitably fixed threshold value $n_0 = n_0(\delta)$. Connectivity test and separation of a connected component – as well as the determination of a maximum-degree vertex – can be performed in $O(n^2)$ steps. Therefore, a non-decreasing integer function $f(n)$ surely is a valid upper bound on the running time of Algorithm 1 on any input graph G on n vertices whenever, for any $n > n_0$ and all integers n' in the range $n/2 \leq n' < n$, we have

$$f(n) \geq kn^2 + f(n') + f(n - n') \tag{3}$$

$$f(n) \geq kn^2 + f(n-1) + f(n - \lceil n^\delta \rceil) \tag{4}$$

where k is a suitably chosen (not large) constant. Throughout this proof, square brackets [ ] will be used as parentheses, with the same meaning as ( ), for making some expressions more transparent.

Note that the time bound in Lemma 7 is superpolynomial, therefore writing f in the form

$$f(n) = g(n) + kn^3/3$$

requires the same growth order for f and g. Let us define

$$g(x) = \exp(h(x))$$

where

$$h(x) = c(\delta) \cdot x^{1-\delta} \cdot \ln x.$$

By the observations above, (3) and (4) will follow if we prove the inequalities

$$g(x) \geq g(x') + g(x - x') \tag{5}$$

$$g(x) \geq g(x-1) + g(x - x^\delta) \tag{6}$$

for every real x large enough and every x' with $x/2 \leq x' \leq x - 1$. We can immediately observe that (5) is a consequence of (6) as

$$g(x) - g(x') \geq g(x) - g(x-1) \geq g(x - x^\delta) \geq g(x/2) \geq g(x - x')$$

if x is large enough with respect to δ, because g is an increasing function and δ is a constant smaller than 1. Therefore only (6) remains to be proved.

We shall need the derivatives of g and h, which can be computed as

$$g'(x) = (\exp[h(x)])' = \exp[h(x)] \cdot h'(x) = g(x) \cdot h'(x)$$

and

$$\begin{aligned} h'(x) &= \left(c(\delta) \cdot x^{1-\delta} \cdot \ln x\right)' \\ &= c(\delta) \cdot x^{-\delta} \cdot [(1-\delta)\ln x + 1]. \end{aligned} \tag{7}$$

It is important to note for later use that

$$h'(x-1) = (1 + o(1)) \cdot h'(x)$$

as $x \to \infty$. Moreover, g and h are increasing, while h' is decreasing, except on a bounded part of the domain.

Next, we apply Cauchy's Mean value theorem in three steps, first for both g and h to estimate $g(x) - g(x-1)$, and second for h to estimate $g(x - x^\delta)$, as follows. For some ξ and ξ' with $x - 1 \leq \xi, \xi' \leq x$ we have

$$\begin{aligned} g(x) - g(x-1) &= g'(\xi) = \exp(h(\xi)) \cdot h'(\xi) \\ &\geq \exp[h(x-1)] \cdot h'(x) \\ &= \exp[h(x) - h'(\xi')] \cdot h'(x) \\ &\geq \exp[h(x) - h'(x-1)] \cdot h'(x) \\ &= \exp\left[h(x) - (1+o(1)) \cdot h'(x)\right] \cdot h'(x). \end{aligned} \tag{8}$$

On the other hand, for some ξ'' with $x - x^\delta \le \xi'' \le x$ we have $h(x - x^\delta) = h(x) - x^\delta \cdot h'(\xi'')$, therefore

$$\begin{aligned} g(x - x^\delta) &= \exp[h(x) - x^\delta \cdot h'(\xi'')] \\ &\le \exp[h(x) - x^\delta \cdot h'(x)]. \end{aligned} \tag{9}$$

Thus, to prove (6), it suffices to show that (8) is not smaller than (9). Taking logarithms this means

$$h(x) - (1 + o(1)) \cdot h'(x) + \ln h'(x) \quad \ge \quad h(x) - x^\delta \cdot h'(x). \tag{10}$$

Or equivalently

$$[x^\delta - 1 - o(1)] \cdot h'(x) \quad \ge \quad - \ln h'(x). \tag{11}$$

Using (7), we obtain that it is enough to prove

$$(c(\delta) + o(1)) \cdot (1 - \delta) \cdot \ln x \quad \ge \quad (\delta + o(1)) \cdot \ln x.$$

This is implied by the condition on $c(\delta)$ (even with strict inequality), completing the proof of the lemma. ◀

Theorem 1 follows immediately from putting together Lemmas 6 and 7.

4 Algorithm for Scattered Set on P_t-free graphs

The algorithm for Scattered Set for P_t-free graphs hinges on the following combinatorial bound.

▶ **Lemma 8.** *For every $t \ge 2$ and for every P_t-free graph with m edges, we have that G has treewidth at most $3m^{1-1/(t+2)}$.*

Proof. Let n be the number of vertices of G. We may ignore components of G that are trees or isolated vertices and hence we can assume that $n \le m$. We consider two cases. Suppose first that $m \ge n^{1+1/(t+1)}$. Then we have

$$m^{1-1/(t+2)} \ge n^{(1+1/(t+1))(1-1/(t+2))} = n.$$

Obviously, n is an upper bound on the treewidth of G, and hence the claim follows.

Suppose now that $m < n^{1+1/(t+1)}$. Let X be the subset of vertices of G with degree at least $n^{2/(t+1)}$. The degree sum of the vertices in X is at most $2m$, hence we have $|X| \le 2m/n^{2/(t+1)} < 2n^{1-1/(t+1)}$. By the definition of X, the graph $G - X$ has maximum degree less than $n^{2/(t+1)}$. Thus each component of X is a P_t-free graph with maximum degree less than $n^{2/(t+1)}$ and hence Lemma 6 implies that each component of $G - X$ has at most $n^{(2/(t+1))\lfloor t/2 \rfloor} \le n^{1-1/(t+1)}$ vertices. In particular, this implies that $G - X$ has treewidth at most $n^{1-1/(t+1)}$. As removing a vertex can decrease treewidth at most by one, it follows that G has treewidth at most $n^{1-1/(t+1)} + |X| = 3n^{1-1/(t+1)} < 3m^{1-1/(t+1)} \le 3m^{1-1/(t+2)}$. ◀

It is known that Scattered Set can be solved in time $d^{O(w)} \cdot n^{O(1)}$ on graphs of treewidth w using standard dynamic programming techniques (cf. [20, 14]). By Lemma 8, it follows that Scattered Set on P_t-free graphs can be solved in time

$$d^{3m^{1-1/(t+2)}} \cdot n^{O(1)} = 2^{O(m^{1-1/(t+2)} \log m)} = 2^{m^{1-1/(t+2)+o(1)}}$$

(taking into account that we may assume $n = O(m)$ and $d \leq n$). Observe that if every component of H is a path, then H is an induced subgraph of $P_{2|V(H)|}$, which implies that H-free graphs are $P_{2|V(H)|}$-free. Thus the algorithm described here for P_t-free graphs implies the first part of Theorem 3.

5 Lower bounds for Scattered Set

A standard consequence of ETH and the so-called Sparsification Lemma is that there is no subexponential-time algorithm for MIS even on graphs of bounded degree (see, e.g., [7]):

▶ **Theorem 9.** *Assuming ETH, there is no $2^{o(n)}$-time algorithm for* MIS *on n-vertex graphs of maximum degree 3.*

A very simple reduction can reduce MIS to 3-Scattered Set for P_5-free graphs, showing that, assuming ETH, there is no algorithm subexponential in the number of vertices for the latter problem. This proves Theorem 2 stated in the Introduction.

Proof (Theorem 2). Given an n-vertex m-edge graph G with maximum degree 3 and an integer k, we construct a graph G' with $n + m = O(n)$ vertices such that $\alpha(G) = \alpha_3(G')$. This reduction proves that a $2^{o(n)}$-time algorithm for 3-Scattered Set could be used to obtain a $2^{o(n)}$-time algorithm for MIS on graphs of maximum degree 3, and this would violate ETH by Theorem 9.

The graph G' contains one vertex for each vertex of G and additionally one vertex for each edge of G. The m vertices of G' representing the edges of G form a clique. Moreover, if the endpoints of an edge $e \in E(G)$ are $u, v \in V(G)$, then the vertex of G' representing e is connected with the vertices of G' representing u and v. This completes the construction of G'. It is easy to see that G' is P_5-free: an induced path of G' can contain at most two vertices of the clique corresponding to $E(G)$ and the vertices of G' corresponding to the vertices of G form an independent set.

If S is an independent set of G, then we claim that the corresponding vertices of G' are at distance at least 3 from each other. Indeed, no two such vertices have a common neighbor: if $u, v \in S$ and the corresponding two vertices in G' have a common neighbor, then this common neighbor represents an edge e of G whose endpoints are u and v, violating the assumption that S is independent. Conversely, suppose that $S' \subseteq V(G')$ is a set of k vertices with pairwise distance at least 3 in G'. If $k \geq 2$, then all these vertices represent vertices of G: observe that for every edge e of G, the vertex of G' representing e is at distance at most 2 from every other non-isolated vertex of G'. We claim that S' corresponds to an independent set of G. Indeed, if $u, v \in S'$ and there is an edge e in G' with endpoints u and v, then the vertex of G' representing e is a common neighbor of u and v, a contradiction. ◀

Next we give negative results on the existence of algorithms for Scattered Set that have running time subexponential in the number of edges. To rule out such algorithms, we construct instances that have bounded degree: then being subexponential in the number of vertices or the number of edges are the same. We consider first claw-free graphs. The key insight here is that Scattered Set with $d = 3$ in line graphs (which are claw-free) is essentially the Induced Matching problem, for which it is easy to prove hardness results.

▶ **Theorem 10.** *Assuming ETH, d-*Scattered Set *does not have a $2^{o(n)}$ algorithm on n-vertex claw-free graphs of maximum degree 4 for any fixed $d \geq 3$.*

Proof. Given an n-vertex graph G with maximum degree 3, we construct a claw-free graph G' with $O(dn)$ vertices and maximum degree 4 such that $\alpha_d(G') = \alpha(G)$. Then by Theorem 9, a $2^{o(n)}$-time algorithm for d-Scattered Set for n-vertex claw-free graphs of maximum degree 4 would violate ETH.

The construction is slightly different based on the parity of d; let us first consider the case when d is odd. Let us construct the graph G^+ by attaching a path Q_v of $\ell = (d-1)/2$ edges to each vertex $v \in V(G)$; let us denote by $e_{v,1}, \ldots, e_{v,\ell}$ the edges of this path such that $e_{v,1}$ is incident with v. The graph G' is defined as the line graph of G^+, that is, each vertex of G' represents an edge of G^+ and two vertices of G' are adjacent if the corresponding two vertices share an endpoint. It is well known that line graphs are claw-free. As G^+ has $O(dn)$ edges and maximum degree 4 (recall that G has maximum degree 3), the line graph G' has $O(dn)$ vertices an edges. Thus an algorithm for Scattered Set with running time $2^{o(n)}$ on n-vertex claw-free graphs of maximum degree 3 could be used to solve MIS on n-vertex graphs with maximum degree 3 in time $2^{o(n)}$, contradicting ETH.

If there is an independent set S of size k in G, then we claim that the set $S' = \{e_{v,\ell} \mid v \in S\}$ is a d- scattered set of size k in G'. To see this, suppose for a contradiction that there are two vertices $u, v \in S$ such that the vertices of G' representing $e_{u,\ell}$ and $e_{v,\ell}$ are at distance at most $d-1$ from each other. This implies that there is a path in G^+ that has at most d edges and whose first and last edges are $e_{u,\ell}$ and $e_{v,\ell}$, respectively. However, such a path would need to contain all the ℓ edges of path Q_u and all the ℓ edges of Q_v, hence it can contain at most $d - 2\ell = 1$ edges outside these two paths. But u and v are not adjacent in G^+ by assumption, hence more than one edge is needed to complete Q_u and Q_v to a path, a contradiction.

Conversely, let S' be a distance-d scattered set in G', which corresponds to a set S^+ of edges in G^+. Observe that for any $v \in V(G)$, at most one edge of S^+ can be incident to the vertices of Q_v: otherwise, the corresponding two vertices in the line graph G' would have distance at most $\ell < d$. It is easy to see that if S^+ contains an edge incident to a vertex of Q_v, then we can always replace this edge with $e_{v,\ell}$, as this can only move it farther away from the other edges of S^+. Thus we may assume that every edge of S^+ is of the form $e_{v,\ell}$. Let us construct the set $S = \{v \mid e_{v,\ell} \in S^+\}$, which has size exactly k. Then S is independent in G: if $u, v \in S$ are adjacent in G, then there is a path of $2\ell + 1 = d$ edges in G^+ whose first an last edges are $e_{v,\ell}$ and $e_{u,\ell}$, respectively, hence the vertices of G' corresponding to them have distance at most $d-1$.

If $d \geq 4$ is even, then the proof is similar, but we obtain the graph G^+ by first subdividing each edge and attaching paths of length $\ell = d/2 - 1$ to each original vertex. The proof proceeds in a similar way: if u and v are adjacent in G, then G^+ has a path of $2\ell + 2 = d$ edges whose first and last edges are $e_{v,\ell}$ and $e_{u,\ell}$, respectively, hence the vertices of G' corresponding to them have distance at most $d-1$. ◀

There is a well-known and easy way of proving hardness of MIS on graphs with large girth: subdivide edges increases girth and the size of the largest independent set changes in a controlled way.

▶ **Lemma 11.** *If there is an $2^{o(n)}$-time algorithm for* MIS *on n-vertex graphs of maximum degree 3 and girth more than g for any fixed $g > 0$, then ETH fails.*

Proof. Let g be a fixed constant and let G be a simple graph with n vertices, m edges, and maximum degree 3 (hence $m = O(n)$). We construct a graph G' by subdividing each edge with $2g$ new vertices. We have that G' has $n' = O(n + gm) = O(n)$ vertices, maximum degree 3, and girth at least $3(2g+1)$. It is known and easy to show that subdividing the edges this way increases the size of the maximum independent set exactly by gm. Thus a $2^{o(n')}$- time algorithm for n'-vertex graphs of maximum degree 3 and girth at least g could be used to give a $2^{o(n)}$-time algorithm for n-vertex graphs of maximum degree g, hence ETH would fail by Theorem 9. ◀

We use the lower bound of Lemma 11 to prove lower bounds for Scattered Set on C_t-free graphs.

▶ **Theorem 12.** *Assuming ETH, d-*Scattered Set *does not have a $2^{o(n)}$ algorithm on n-vertex C_t-free graphs with maximum degree 3 for any fixed $t \geq 3$ and $d \geq 2$.*

Proof. Let G be an n-vertex m-edge graph of maximum degree 3 and girth more than t. We construct a graph G' the following way: we subdivide each edge of G with $d-2$ new vertices to create a path of length $d-1$, and attach a path of length $d-1$ to each of the $(d-2)m = O(dn)$ new vertices created. The resulting graph has maximum degree 3, $O(d^2n)$ vertices and edges, and girth more than $(d-1)t$ (hence it is C_t-free). We claim that $\alpha_d(G') = \alpha(G) + m(d-2)$ holds. This means that an $2^{o(n')}$-time algorithm for Scattered Set n'-vertex C_t-free graphs with maximum degree 3 would give a $2^{o(n)}$-time algorithm for n-vertex graphs of maximum degree 3 and girth more than t and this would violate ETH by Lemma 11.

To see that $\alpha_d(G') = \alpha(G) + m(d-2)$ holds, consider first an independent set S of G. When constructing G', we attached $m(d-2)$ paths of length $d-1$. Let S' contain the degree-1 endpoints of these $m(d-2)$ paths, plus the vertices of G' corresponding to the vertices of S. It is easy to see that any two vertices of S' has distance at least d from each other: S is an independent set in G, hence the corresponding vertices in G' are at distance at least $2(d-1)$ from each other, while the degree-1 endpoints of the paths of length $d-1$ are at distance at least d from every other vertex that can potentially be in S'. This shows $\alpha_d(G') \geq \alpha(G) + m(d-2)$ Conversely, let S' be a set of vertices in G' that are at distance at least d from each other. The set S' contains two types of vertices: let S'_1 be the vertices that correspond to the original vertices of G and let S'_2 be the $m(d-2)d$ new vertices introduced in the construction of G'. Observe that S'_2 can be covered by $m(d-2)$ paths of length $d-1$ and each such path can contain at most one vertex of S', hence at most $m(d-2)$ vertices of S' can be in S'_2. We claim that S'_1 can contain at most $\alpha(G)$ vertices, as $S' \cap S'_1$ corresponds to an independent set of G. Indeed, if u and v are adjacent vertices of G, then the corresponding two vertices of G' are at distance $d-1$, hence they cannot be both present in S'. This shows $\alpha_d(G') \leq \alpha(G) + m(d-2)$, completing the proof of the correctness of the reduction. ◀

As the following corollary shows, putting together Theorems 10 and 12 implies Theorem 3(2).

▶ **Corollary 13.** *If H is a graph having a component that is not a path, then, assuming ETH, d-*Scattered Set *has no $2^{o(n+m)}$-time algorithm on n-vertex m-edge H-free graphs for any fixed $d \geq 3$.*

Proof. Suppose first that H is not a forest and hence some cycle C_t for $t \geq 3$ appears as an induced subgraph in H. Then the class of H-free graphs is a superset of C_t-free graphs, which means that statement follows from Theorem 12 (which gives a lower bound for a more restricted class of graphs).

Assume therefore that H is a forest. Then it has to have a component that is a tree, but not a path, hence it has a vertex v of degree at least 3. The neighbors of v are independent in the forest H, which means that the claw $K_{1,3}$ appears in H as an induced subgraph. Then the class of H-free graphs is a superset of claw-free graphs, which means that statement follows from Theorem 10 (which gives a lower bound for a more restricted class of graphs). ◀

6 Conclusion

In spite of our results, it remains an open problem for an infinite class of graphs H, whether a subexponential or even a polynomial algorithm exists for MIS on H-free graphs. Namely, as indicated in the Introduction, among connected graphs these are the ones in which the triple of lengths of paths starting from the unique vertex of degree three is (i, j, k) with $i \leq j \leq k$ and with $(i, j, k) \neq (1, 1, 1), (1, 1, 2)$. Moreover, for paths, it is an unsolved question whether the problem is polynomial-time solvable for $H = P_t$, $t \geq 6$.

Our subexponential algorithm uses simple branching which clearly works for Weighted Independent Set as well.

For Scattered Set, we have seen that on P_t-free graphs there are algorithms subexponential in the number of edges, and Theorem 2 shows that polynomial-time algorithms are unlikely. But can one give a tight lower bound on the subexponential running time, perhaps showing that $1 - O(1/t)$ in the exponent of the exponent is in some sense best possible?

After the acceptance of this manuscript we learned that independently and simultaneously Brause (Ch. Brause, "A subexponential-time algorithm for the Maximum Independent Set in P_t-free graphs", *Discrete Applied Mathematics*, `DOI:10.1016/j.dam.2016.06.016`) also proved the subexponentiality of MIS on P_t-free graphs. (His time bound is weaker than the one in this paper.) Moreover, an unpublished result of Lokshtanov, Pilipczuk, and van Leuwen yields an algorithm with much better bound on the running time.

References

1. Geir Agnarsson, Peter Damaschke, and Magnús M. Halldórsson. Powers of geometric intersection graphs and dispersion algorithms. *Discrete Applied Mathematics*, 132(1-3):3–16, 2003. `doi:10.1016/S0166-218X(03)00386-X`.
2. V. E. Alekseev. The effect of local constraints on the complexity of determination of the graph independence number. In *Combinatorial-algebraic methods in applied mathematics*, pages 3–13. Gor′kov. Gos. Univ., Gorki, 1982.
3. Vladimir E. Alekseev. Polynomial algorithm for finding the largest independent sets in graphs without forks. *Discrete Applied Mathematics*, 135(1-3):3–16, 2004. `doi:10.1016/S0166-218X(02)00290-1`.
4. Gábor Bacsó and Zsolt Tuza. A characterization of graphs without long induced paths. *J. Graph Theory*, 14(4):455–464, 1990. `doi:10.1002/jgt.3190140409`.
5. Gábor Bacsó and Zsolt Tuza. Dominating cliques in P_5-free graphs. *Period. Math. Hungar.*, 21(4):303–308, 1990. `doi:10.1007/BF02352694`.
6. Binay K. Bhattacharya and Michael E. Houle. Generalized maximum independent sets for trees in subquadratic time. In *Algorithms and Computation, 10th International Symposium, ISAAC'99, Chennai, India, December 16-18, 1999, Proceedings*, pages 435–445, 1999. `doi:10.1007/3-540-46632-0_44`.
7. Marek Cygan, Fedor V. Fomin, Lukasz Kowalik, Daniel Lokshtanov, Dániel Marx, Marcin Pilipczuk, Michal Pilipczuk, and Saket Saurabh. *Parameterized Algorithms*. Springer, 2015. `doi:10.1007/978-3-319-21275-3`.
8. Paul Erdős, Michael E. Saks, and Vera T. Sós. Maximum induced trees in graphs. *J. Comb. Theory, Ser. B*, 41(1):61–79, 1986. `doi:10.1016/0095-8956(86)90028-6`.

9 Hiroshi Eto, Fengrui Guo, and Eiji Miyano. Distance-d independent set problems for bipartite and chordal graphs. *J. Comb. Optim.*, 27(1):88–99, 2014. `doi:10.1007/s10878-012-9594-4`.

10 Russell Impagliazzo, Ramamohan Paturi, and Francis Zane. Which problems have strongly exponential complexity? *J. Comput. Syst. Sci.*, 63(4):512–530, 2001. `doi:10.1006/jcss.2001.1774`.

11 Daniel Lokshantov, Martin Vatshelle, and Yngve Villanger. Independent set in P_5-free graphs in polynomial time. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014*, pages 570–581, 2014. `doi:10.1137/1.9781611973402.43`.

12 Daniel Lokshtanov, Dániel Marx, and Saket Saurabh. Lower bounds based on the exponential time hypothesis. *Bulletin of the EATCS*, 105:41–72, 2011. URL: `http://albcom.lsi.upc.edu/ojs/index.php/beatcs/article/view/96`.

13 Daniel Lokshtanov, Marcin Pilipczuk, and Erik Jan van Leeuwen. Independence and efficient domination on P_6-free graphs. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016*, pages 1784–1803, 2016. `doi:10.1137/1.9781611974331.ch124`.

14 Dániel Marx and Michal Pilipczuk. Optimal parameterized algorithms for planar facility location problems using Voronoi diagrams. In *Algorithms – ESA 2015 – 23rd Annual European Symposium, Patras, Greece, September 14-16, 2015, Proceedings*, pages 865–877, 2015. `doi:10.1007/978-3-662-48350-3_72`.

15 M. Miller and J. Širáň. Moore graphs and beyond: A survey of the degree/diameter problem. *Electronic J. Combinatorics*, 20:1–92, 2013.

16 George J. Minty. On maximal independent sets of vertices in claw-free graphs. *J. Combin. Theory Ser. B*, 28(3):284–304, 1980. `doi:10.1016/0095-8956(80)90074-X`.

17 Bert Randerath and Ingo Schiermeyer. On maximum independent sets in P_5-free graphs. *Discrete Applied Mathematics*, 158(9):1041–1044, 2010. `doi:10.1016/j.dam.2010.01.007`.

18 Daniel J. Rosenkrantz, Giri Kumar Tayi, and S. S. Ravi. Facility dispersion problems under capacity and cost constraints. *J. Comb. Optim.*, 4(1):7–33, 2000. `doi:10.1023/A:1009802105661`.

19 Najiba Sbihi. Algorithme de recherche d'un stable de cardinalité maximum dans un graphe sans étoile. *Discrete Math.*, 29(1):53–76, 1980. `doi:10.1016/0012-365X(90)90287-R`.

20 Dimitrios M. Thilikos. Fast sub-exponential algorithms and compactness in planar graphs. In *Algorithms – ESA 2011 – 19th Annual European Symposium, Saarbrücken, Germany, September 5-9, 2011. Proceedings*, pages 358–369, 2011. `doi:10.1007/978-3-642-23719-5_31`.

21 David Zuckerman. Linear degree extractors and the inapproximability of Max Clique and Chromatic Number. *Theory of Computing*, 3(1):103–128, 2007. `doi:10.4086/toc.2007.v003a006`.