Adjacency Labeling Schemes for Small Classes

Édouard Bonnet[*]   Julien Duron[†]   John Sylvester[‡]   Viktor Zamaraev[§]

Abstract

A graph class admits an *implicit representation*[1] if, for every positive integer n, its n-vertex graphs have a $b(n)$-bit (adjacency) labeling scheme with $b(n) = O(\log n)$, i.e., their vertices can be labeled by binary strings of length $b(n)$ such that the presence of an edge between any pair of vertices u, v can be deduced solely from the labels of u and v. The famous Implicit Graph Conjecture posited that every *hereditary* (i.e., closed under taking induced subgraphs) *factorial* (i.e., containing $2^{O(n \log n)}$ n-vertex graphs) class admits an implicit representation. The conjecture was finally refuted [Hatami and Hatami, FOCS '22], and does not even hold among *monotone* (i.e., closed under taking subgraphs) factorial classes [Bonnet et al., ICALP '24]. However, monotone *small* (i.e., containing at most $n!c^n$ many n-vertex graphs for some constant c) classes do admit implicit representations.

This motivates the *Small Implicit Graph Conjecture*: Every hereditary small class admits an $O(\log n)$-bit labeling scheme. We provide evidence supporting the Small Implicit Graph Conjecture. First, we show that every small *weakly sparse* (i.e., excluding some fixed bipartite complete graph as a subgraph) class has an implicit representation. This is a consequence of the following fact of independent interest proven in the paper: Every weakly sparse small class has bounded expansion (hence, in particular, bounded degeneracy). The latter generalizes and strengthens the previous results that every monotone small class has bounded degeneracy [Bonnet et al., ICALP '24], and that every weakly sparse class of bounded twin-width has bounded expansion [Bonnet et al., Combinatorial Theory '22]. Second, we show that every hereditary small class admits an $O(\log^3 n)$-bit labeling scheme, which provides a substantial improvement over the best-known polynomial upper bound of $n^{1-\varepsilon}$ on the size of adjacency labeling schemes for such classes. To do so, we establish that every small class has *neighborhood complexity* $O(n \log n)$, also of independent interest. We then apply a classic result, due to Welzl [SoCG '88], on efficiently ordering the universe of a set system of low Vapnik-Chervonenkis dimension such that every set can be described as the union of a limited number of intervals along this order.

---

[*]Univ. Lyon, ENS de Lyon, UCBL, CNRS, LIP, France, `edouard.bonnet@ens-lyon.fr`

[†]Univ. Lyon, ENS de Lyon, UCBL, CNRS, LIP, France, `julien.duron@ens-lyon.fr`

[‡]Department of Computer Science, University of Liverpool, UK, `john.sylvester@liverpool.ac.uk`

[§]Department of Computer Science, University of Liverpool, UK, `viktor.zamaraev@liverpool.ac.uk`

[1]To simplify and lighten the abstract, we give the definition of *implicit representations* for *factorial* classes. The reader will find the general definition in the introduction.

1 Introduction

A *class* of graphs is a set of graphs which is closed under isomorphism. For a class of graphs $\mathcal{X}$ we denote by $\mathcal{X}_n$ the set of graphs in $\mathcal{X}$ with vertex set $[n]$. Let $\mathcal{X}$ be a class of graphs and $b : \mathbb{N} \to \mathbb{N}$ be a function. A $b(n)$-*bit adjacency labeling scheme* (or simply $b(n)$-*bit labeling scheme*) for $\mathcal{X}$ is a pair (encoder, decoder) of algorithms where for any n-vertex graph $G \in \mathcal{X}_n$ the encoder assigns binary strings, called *labels*, of length $b(n)$ to the vertices of G such that the adjacency between any pair of vertices can be inferred by the decoder only from their labels. We note that the decoder depends on the class $\mathcal{X}$, but not on the graph G. The function $b(\cdot)$ is the *size* of the labeling scheme. Adjacency labeling schemes were introduced by Kannan, Naor, and Rudich [KNR88, KNR92], and independently by Muller [Mul88] in the late 1980s and have been actively studied since then.

The binary word, obtained by concatenating labels of the vertices of a graph $G \in \mathcal{X}_n$ assigned by an adjacency labeling scheme, uniquely determines graph G. Thus, a $b(n)$-bit labeling scheme cannot represent more than $2^{nb(n)}$ graphs on n vertices, and therefore, if $\mathcal{X}$ admits a $b(n)$-bit labeling scheme, then $|\mathcal{X}_n| \leqslant 2^{nb(n)}$. This implies a lower bound of $\frac{\log |\mathcal{X}_n|}{n}$ on the size $b(n)$ of any adjacency labeling scheme for $\mathcal{X}$. We say that a graph class $\mathcal{X}$ admits an *implicit representation*, if it admits an information-theoretic *order optimal* adjacency labeling scheme, i.e., if $\mathcal{X}$ has a $b(n)$-bit labeling scheme, where $b(n) = O(\log |\mathcal{X}_n|/n)$. A natural and important question is: which classes admit an adjacency labeling scheme of a size that matches this information-theoretic lower bound?

Of particular interest is the case of adjacency labeling schemes of size $O(\log n)$. This is because, under the natural assumption that vertices get assigned pairwise distinct labels, $\lceil \log n \rceil$ is a lower bound on the size of any labeling scheme. Thus, understanding the expressive power of the labeling schemes of size of order $\log n$ is a natural question.

By the above discussion adjacency labeling schemes of size $O(\log n)$ can only exist for (at most) *factorial* graph classes, i.e., classes $\mathcal{X}$ in which the number $|\mathcal{X}_n|$ of n-vertex graphs grows not faster than $2^{\Theta(n \log n)}$. It is known that the latter condition alone is not enough to guarantee adjacency labels of size $O(\log n)$ [Mul88]. Kannan, Naor, and Rudich [KNR88] asked if the extra restriction on $\mathcal{X}$ of being *hereditary*, i.e., closed under taking induced subgraphs, would be enough for the existence of $O(\log n)$-bit adjacency labeling scheme. This question was later stated by Spinrad [Spi03] in the form of a conjecture, reiterated by Scheinerman [GHL16, Chapter 6], that became known as the *Implicit Graph Conjecture*.

(*IGC*): Any hereditary factorial graph class admits an $O(\log n)$-bit labeling scheme.

The conjecture remained open for three decades until it was recently refuted in a strong form by Hatami and Hatami, who showed that for any δ there exists a hereditary factorial class that does not admit an $(n^{1/2-\delta})$-bit labeling scheme [HH22]. This refutation leaves wide open the question of characterizing hereditary graph classes that admit $O(\log n)$-bit labeling schemes and no plausible conjectural dichotomy is currently available.

It was recently shown that IGC does not hold even in the family of *monotone* graph classes, the hereditary classes which are closed under taking (not necessarily induced) subgraphs [BDS$^+$24b]. On the positive side, it was shown in the same work that IGC holds for every monotone *small* class, i.e., a monotone class $\mathcal{X}$ with $|\mathcal{X}_n| \leqslant n! \cdot c^n$ for some constant $c > 0$.

The hereditary small classes form a subfamily of hereditary factorial classes that contains many classes of practical or theoretical importance. For instance, forests [Ott48], planar graphs [Tut62], classes of bounded treewidth [BP69], proper minor-closes classes [NSTW06], unit interval graphs [Han82], classes of bounded clique-width [ALR09], and more generally, classes of bounded twin-width [BGK$^+$22] are all small. All of these classes are known to admit order-optimal, and, sometimes, even

asymptotically optimal adjacency labeling schemes. For example, forests admit a $(\log n + O(1))$-bit labeling scheme [ADK17], planar graphs admit a $(1 + o(1)) \log n$-bit labeling scheme [DEG$^+$21], classes of bounded treewidth admit a $(1 + o(1)) \log n$-bit labeling scheme [GL07], proper minor-closed classes admit a $(2 + o(1)) \log n$-bit labeling scheme [GL07], graphs of clique-width at most k admit a $O(k \log k \cdot \log n)$-bit labeling scheme [Spi03, Ban22], and graphs of twin-width at most k admit a $2^{2^{O(k)}} \cdot \log n$-bit labeling scheme [BGK$^+$22]. It is known that not every hereditary (and even monotone) small class admits an asymptotically optimal labeling scheme of size $(1 + o(1)) \log n$ [BDS$^+$24a]. However, this does not rule out the existence of order optimal labeling schemes of size $O(\log n)$ for all hereditary small classes. Alon [Alo23] recently showed that every hereditary class $\mathcal{X}$ with $|\mathcal{X}_n| = 2^{o(n^2)}$ (which include all hereditary small classes) admits an adjacency labeling scheme of size $n^{1-\varepsilon}$ for some $\varepsilon > 0$ depending on the class. To our knowledge, no better upper bound was established for hereditary small classes.

The importance of hereditary small classes and the existence of $O(\log n)$-bit labeling schemes for all monotone small classes motivated the formulation of the *Small Implicit Graph Conjecture.*

(*Small IGC*) [BDS$^+$24b]: Any hereditary small graph class admits an $O(\log n)$-bit labeling scheme.

1.1 Our contribution

In this paper we provide evidence toward the Small Implicit Graph Conjecture in two independent directions. First, we show that the Small IGC holds in the important special case of weakly sparse graph classes. Secondly, we obtain an adjacency labeling scheme of polylogarithmic size for any hereditary small class substantially improving upon the best-known polynomial-size upper bound on adjacency labeling schemes for such classes.

Our first result is a consequence of a structural property of small classes of graphs. To state this property, we need to introduce some terms. A hereditary class $\mathcal{X}$ is said to be *weakly sparse* if there exists an integer t such that the complete bipartite graph $K_{t,t}$ is not a subgraph of any graph of $\mathcal{X}$. The *degeneracy* of a graph G is the minimum number k such that every induced subgraph of G contains a vertex of degree at most k. We defer a formal definition of *bounded expansion* until Section 2 as it is somewhat technical. Instead, we note that *bounded expansion* generalizes *bounded degeneracy*, thus Theorem 1.1 a fortiori holds when the former is replaced by the latter.

Theorem 1.1. *Every weakly sparse small class has bounded expansion.*

This structural insight into small graph classes is of independent interest due to the significance of weakly sparse graph classes. The notion of weakly sparseness is the broadest form of sparsity. It generalizes the properties of bounded degree, bounded degeneracy, and nowhere denseness [NO12]. By the celebrated Kővári–Sós–Turán theorem [KST54], among hereditary classes, weakly sparse classes are precisely the classes of graphs with a truly subquadratic number of edges. It is often observed that the extra restriction of being weakly sparse significantly simplifies the structure of graphs possessing a particular property. For example, weakly sparse graph classes of bounded clique-width have bounded treewidth [GW00]; weakly sparse graph classes of bounded twin-width have bounded expansion, and thus bounded degeneracy [BGK$^+$22] (a fact that Theorem 1.1 generalizes, as it was proven in the same paper that every class of bounded twin-width is small); weakly sparse string graphs have bounded degeneracy [FP14], and even polynomial expansion [DN16]; weakly sparse graph classes that are well-quasi-ordered by the induced-subgraph relation have bounded pathwidth [ALR19], etc.

A conjecture from 2016 [War16], driving an ongoing program in algorithmic graph theory, suggests that, under some widely-believed complexity-theoretic assumption, a hereditary graph class

admits a fixed-parameter tractable (FPT) first-order (FO) model-checking algorithm if and only if it is monadically dependent.[2] It was recently proven that every hereditary small class is monadically dependent [DMT24] (and, in the same paper, the *only if* part of the conjecture was established). Thus, if the conjecture holds, it should in particular be true that every hereditary small class admits an FPT FO model-checking algorithm. Theorem 1.1 confirms that this is indeed the case within weakly sparse classes, as such an algorithm is known to exist for classes of bounded expansion [DKT13].

Dvořák and Norine proved that if the expansion of a class is upper bounded by a sufficiently slowly-growing subexponential function, then the class is small [DN10]. From the definition of *expansion* given in the next section, it should be clear that the converse cannot hold since the small classes consisting of all complete graphs or of all bipartite complete graphs have unbounded expansion. A minimum requirement to establish a partial converse is thus that the class is weakly sparse. Theorem 1.1 confirms that under this condition the expansion of any small class is indeed bounded (albeit not necessarily by a subexponential nor a single-exponential function).

We apply Theorem 1.1 to obtain our first main result that the Small IGC holds for weakly sparse graph classes. This follows from a classical labeling scheme for classes of bounded degeneracy [KNR88] (see also [BDS+24b, Lemma 3.5]).

Theorem 1.2. *Every weakly sparse small class admits an $O(\log n)$-bit adjacency labeling scheme.*

Our second main result holds more generally for any hereditary small class and goes some way toward the full strength of the Small IGC.

Theorem 1.3. *Every hereditary small class admits an $O(\log^3 n)$-bit adjacency labeling scheme.*

This result is based on our second structural result of independent interest that hereditary small graph classes have low neighborhood complexity.

Theorem 1.4. *The neighborhood complexity of any hereditary small graph class is $O(n \log n)$.*

The neighborhood complexity is a measure of structural complexity of graphs. Informally (a formal definition is given in Section 2), it reflects the diversity of the set system of the vertex neighborhoods in the graph. Low neighborhood complexity of graph classes implies nice structural and algorithmic properties. For example, in the universe of monotone graph classes, linear neighborhood complexity characterizes classes of bounded expansion [RVS19], and almost linear neighborhood complexity characterizes nowhere dense classes [EGK+17]. Low neighborhood complexity is also used to obtain FPT algorithms for various graph problems (see for instance [EGK+17, BKR+22]), and was recently used as a central ingredient of an FPT algorithm for the FO model-checking problem on monadically stable graph classes [DEM+23]. It is known that every hereditary graph class of bounded twin-width has linear neighborhood complexity [BKR+22] (see also [BFLP24] for improved bounds). We omit formal definitions and result statements, and refer the reader to the respective work.

We obtain our Theorem 1.3 by combining Theorem 1.4 with some known result from the theory of sets systems about paths with low crossing number. In a standard form the latter result is not suitable for our need, and therefore we reformulate it and provide a self-contained proof. We believe that the result stated in this form can be of use for other needs.

---

[2]As with the other side remarks, we do not define the technical terms in this sentence since they will not reappear.

1.2 Organization

The paper is organized as follows. In Section 3, we prove that weakly sparse small classes have bounded expansion (and thus bounded degeneracy); from this, we derive an $O(\log n)$-bit adjacency labeling scheme for such classes. In Section 4.1, we show that every hereditary small class has neighborhood complexity $O(n \log n)$. In Section 4.2, we use this result together with a classical result from the theory of set systems stated in a suitable form, to obtain $O(\log^3 n)$-bit labeling schemes for all hereditary small classes. In Appendix A, we provide a self-contained proof of the required form of the result from the theory of set systems. In the next section, we introduced necessary notation and state auxiliary facts.

2 Preliminaries

Graphs. We denote by $V(G)$ and $E(G)$ the vertex set and edge set of a graph G, respectively. When we refer to an n-vertex graph G as *labeled*, we mean that the vertex set of G is $[n]$, and we distinguish two different labeled graphs even if they are isomorphic. For a set of graphs X, we denote by $\mathrm{Lab}(X)$ the set of all labeled graphs isomorphic to a graph in X. A graph H is an *induced subgraph* of G (resp. *subgraph* of G), if it can be obtained by removing vertices of G (resp. by removing vertices and edges of G). A subgraph H of G is *spanning* if $V(H) = V(G)$. We denote by $G[S]$ the subgraph of G induced by S, i.e., obtained by removing every vertex of $V(G) \setminus S$.

The *degree* of a vertex v in G, denoted by $\deg(v)$, is the number of vertices in G adjacent to v. The *minimum* (respectively, *maximum*) *degree of G* is the minimum (respectively, maximum) degree of a vertex in G. The *average degree of G* is the ratio $\sum_{v \in V(G)} \deg(v)/|V(G)|$. Recall that the degeneracy of G is the minimum number k such that every induced subgraph of G contains a vertex of degree at most k. The following are immediate consequences of these definitions: (1) if the minimum degree of G is at least d, then the average degree of G and the degeneracy of G is at least d; (2) if the degeneracy of G is at least d, then G contains an induced subgraph of minimum degree at least d. One more relation between these three parameters is the following folklore result, which is not hard to derive by iteratively removing vertices of degree less than $d/2$.

Observation 2.1. *If the average degree of a graph G is at least d, then G contains an induced subgraph of minimum degree, and thus average degree, at least $\lceil d/2 \rceil$.*

We use $G - S$ as a shorthand for $G[V(G) \setminus S]$, and $G - v$, for $G - \{v\}$. We denote by $\mathrm{Aut}(G)$ the automorphism group of a graph G, and we set $\mathrm{aut}(G) := |\mathrm{Aut}(G)|$. If X is a set of pairwise non-isomorphic n-vertex graphs, then the number of labeled graphs isomorphic to a graph in X is exactly

$$|\mathrm{Lab}(X)| = \sum_{G \in X} \frac{n!}{\mathrm{aut}(G)}. \tag{1}$$

For any non-negative integer r, the *r-subdivision* of a graph G, denoted by $r\text{-}\mathrm{subd}(G)$, is the graph obtained from G by adding one path P_{uv} on r vertices for each edge $uv \in E(G)$, and by replacing each edge $uv \in E(G)$ by a $(r+1)$-edge path $uP_{uv}v$. In the special case were $r = 0$, $r\text{-}\mathrm{subd}(G)$ is the graph G. We will refer to a vertex from one of the paths P_{uv} as a *subdivision vertex* and to other vertices as *branching vertices*.

For completeness, we include a definition[3] of classes of *bounded expansion*. A class $\mathcal{C}$ has *bounded expansion* if there is a function $f : \mathbb{N} \to \mathbb{R}$ such that for every integer $r \geqslant 0$, no r-subdivision of a graph of average degree larger than $f(r)$ is a subgraph of a graph in $\mathcal{C}$. We will only apply existing

---

[3]Actually, we rather give an established characterization [Dvo07] as it is more compact.

results on classes of bounded expansion, and we will do so in a black-box fashion. Thus the reader should only observe that these classes are weakly sparse and of bounded average degree (by virtue of satisfying the given property for $r = 0$).

Graph Classes. A class of graphs is *hereditary* if it is closed under taking induced subgraphs, and it is *monotone* if it closed under taking subgraphs. A hereditary graph class is *small* if there exists a constant c such that the number of n-vertex *labeled* graphs in the class is at most $n! \cdot c^n$ for every $n \in \mathbb{N}$.

Graph contiguity. The *contiguity* of a graph G, denoted $\mathrm{ctg}(G)$, is the minimum integer k such that there exists a linear order of its vertices in which the neighborhood of each vertex of G can be partitioned into at most k disjoint intervals. For a function $k : \mathbb{N} \to \mathbb{N}$ and a class of graphs $\mathcal{C}$, we say that $\mathcal{C}$ has contiguity at most k, if for every n-vertex graph $G \in \mathcal{C}$ the contiguity of G is at most $k(n)$.

The idea of contiguity was introduced in the context of range searching under the name of *crossing number of spanning paths* [Wel88] (see also Appendix A). Independently, it was introduced in the context of DNA reconstruction under the name of *k-consecutive ones property* [GGKS95], which generalizes the notion of *consecutive ones property* for matrices [BL76, FG65]. More recently, the concept was utilized as part of an FPT algorithm for the FO model-checking problem on monadically stable graph classes [DEM+23], and has appeared in the context of implicit graph representations [Alo23, AAA+23, BDSZ24]. In particular, Alon used it to show that every hereditary graph class $\mathcal{X}$ with $|\mathcal{X}_n| = 2^{o(n^2)}$ admits a $n^{1-\varepsilon}$-bit labeling scheme for some $\varepsilon = \varepsilon(\mathcal{X}) > 0$ [Alo23]. This was done via the following simple connection between contiguity and adjacency labeling schemes.

Proposition 2.2. *Let $k : \mathbb{N} \to \mathbb{N}$ be a function, and $\mathcal{C}$ be a class of graphs of contiguity at most k. Then $\mathcal{C}$ admits an adjacency labeling scheme of size at most $(2k(n) + 1) \log n$.*

Proof. Let G be a n-vertex graph in $\mathcal{C}$. To construct adjacency labels for G of the target size, we fix a linear ordering σ of the vertices of G witnessing contiguity at most $k(n)$. The label of a vertex v of G then consists of its position in σ, followed by the positions of the endpoints of the at most $k(n)$ intervals describing the neighborhood of v. The decoder infers adjacency between two vertices u and v by testing if the position of u in σ belongs to one of the intervals describing the neighborhood of v. $\square$

Set systems, shatter functions, and neighborhood complexity. A pair $(X, \mathcal{S})$, where X is a finite set and $\mathcal{S}$ is a family of subsets of X is called a set system. The *incidence matrix* M of the set system $(X, \mathcal{S})$ is the matrix with $|X|$ columns and $|\mathcal{S}|$ rows indexed by the elements of X and $\mathcal{S}$, respectively, where M_{ij} is 1 if the i-th set of $\mathcal{S}$ contains the j-th element of X, and 0 otherwise.

The *primal shatter function* of a set system $(X, \mathcal{S})$ is the function $\pi_{\mathcal{S}}$ given by

$$\pi_{\mathcal{S}}(m) = \max_{A \subseteq X, |A|=m} |\{Y \cap A \ : \ Y \in \mathcal{S}\}|.$$

The set system $(X^*, \mathcal{S}^*)$, where $X^* = \mathcal{S}$, $\mathcal{S}^* = \{S_x \ : \ x \in X\}$, and $S_x = \{S \in \mathcal{S} \ : \ x \in X\}$, is called the dual set system of $(X, \mathcal{S})$. The shatter function of $(X^*, \mathcal{S}^*)$ is called the *dual shatter function* of $(X, \mathcal{S})$. Note that if M is the incidence matrix of (X, S), then the transpose of M is the incidence matrix of $(X^*, \mathcal{S}^*)$. A subset $A \subseteq X$ is *shattered* by $\mathcal{S}$ if every subset B of A can be obtained as the intersection $B = A \cap Y$ for some $Y \in \mathcal{S}$, i.e., $\{Y \cap A \mid Y \in \mathcal{S}\} = 2^A$. The

Vapnik–Chervonenkis dimension (or *VC dimension* for short) of $\mathcal{S}$ is the maximum of the sizes of all shattered subsets of X.

Given a vertex v of a graph G, we use standard notation $N_G(v)$ to denote the set $\{u : vu \in E(G)\}$ of neighbors of v in G. The *neighborhood set system* of G is the set system $(V(G), \mathcal{N}_G)$ where

$$\mathcal{N}_G = \{N_G(v) : v \in V(G)\}.$$

Observe that the adjacency matrix of G is the incidence matrix of $(V(G), \mathcal{N}_G)$. Thus, since the adjacency matrix of an undirected graph is symmetric, the primal and the dual shatter functions of a neighborhood set system coincide.

Observation 2.3. *For neighborhood set systems, the primal and dual shatter functions coincide. That is, for any n-vertex graph G,*

$$\pi_{\mathcal{N}_G}(m) = \pi^*_{\mathcal{N}_G}(m) \qquad \text{for all } m \in [n].$$

We define the *neighborhood complexity* of a graph G, denoted by ν_G, as the primal (equivalently, dual) shatter function of its neighborhood set system, i.e., $\nu_G(m) = \pi_{\mathcal{N}_G}(m) = \pi^*_{\mathcal{N}_G}(m)$ for all $m \in [|V(G)|]$.

Given a graph class $\mathcal{C}$, the *neighborhood complexity* of $\mathcal{C}$ is the function $\nu_{\mathcal{C}} : \mathbb{N} \to \mathbb{N}$ defined by

$$\nu_{\mathcal{C}}(n) := \sup_{G \in \mathcal{C}, \, A \subseteq V(G), \, |A| = n} |\{N(v) \cap A \; : \; v \in V(G)\}| = \sup_{G \in \mathcal{C}} \pi_{\mathcal{N}(G)}(n) = \sup_{G \in \mathcal{C}} \nu_G(n).$$

3   Implicit representation for weakly sparse small classes

In this section we prove Theorem 1.1, which states that any weakly sparse small class has bounded expansion, and hence, in particular, bounded degeneracy. Together with the known labeling scheme for bounded degeneracy graphs (see e.g. [BDS$^+$24b, Lemma 3.5]), this immediately implies that any weakly sparse small class admits an $O(\log n)$-bit adjacency labeling scheme (Theorem 1.2).

The proof of Theorem 1.1 relies on the fact (Corollary 3.3) that for any weakly sparse class $\mathcal{X}$ of unbounded expansion there exists an integer $r \geqslant 1$ such that $\mathcal{X}$ contains r-subdivisions of graphs of arbitrarily large average degree. This is a consequence of the forward direction of a characterization of classes of (un)bounded expansion by Dvořák [Dvo18] and a classical result due to Kühn and Osthus [KO04], both stated below.

Theorem 3.1 ([Dvo18, Theorem 5, $(a) \Leftrightarrow (c)$]). *A hereditary graph class $\mathcal{X}$ has unbounded expansion if and only if there is an integer $r \geqslant 0$ such that for any integer d, $\mathcal{X}$ contains the r-subdivision of a graph of average degree at least d.*

Theorem 3.2 ([KO04, Theorem 2]). *For all $d, t \in \mathbb{N}$, there exists an integer $D := D(d, t)$ such that every $K_{t,t}$-free graph G of average degree at least D contains $1\text{-}\mathrm{subd}(F)$ as an induced subgraph for some graph F of average degree at least d.*

We now extract the announced fact, of which we will only need the *only if* direction.

Corollary 3.3. *A weakly sparse graph class $\mathcal{X}$ is of unbounded expansion if and only if there is an integer $r \geqslant 1$ such that for any integer d, $\mathcal{X}$ contains the r-subdivision of a graph of average degree at least d.*

Proof. The *if* part follows directly from Theorem 3.1. For the *only if* part, we first apply the corresponding implication in Theorem 3.1. If $r \geqslant 1$, we are done. If instead $r = 0$, then $\mathcal{X}$ contains graphs of arbitrarily large average degree. Therefore, Theorem 3.2 implies that $\mathcal{X}$ contains 1-subdivisions of graphs of arbitrarily large average degree, i.e., the sought statement holds for $r = 1$. □

Equipped with this tool, the proof of Theorem 1.1 shows the contrapositive as follows. Assuming a weakly sparse class $\mathcal{X}$ has unbounded expansion, we fix a positive integer r such that $\mathcal{X}$ contains r-subdivisions of graphs of arbitrarily large average degree, which exists by Corollary 3.3. Let F be a graph with large average degree such that the r-subdivision of F is in $\mathcal{X}$. First, we extract from F a subgraph H with large minimum degree that contains a bounded-degree spanning tree. This is possible due to Observation 2.1 and the following lemma.

Lemma 3.4 ([BDS$^+$24b, Lemma 3.3]). *Let F be a graph of minimum degree d. Then F has an induced subgraph H of minimum degree at least d with a spanning tree of* maximum degree *at most d.*

Next, we use the properties of H to show that it contains many (not necessarily induced) subgraphs each with a small number of automorphisms, which are translated to many *induced* subgraphs with a small number of automorphisms in the r-subdivision of H and thus in the r-subdivision of the original graph F. This allows us to lower bound the number of labeled graphs on the same number of vertices in $\mathcal{X}$ by counting subgraphs of H, and finally conclude that $\mathcal{X}$ is not small. The quantitative connection is made possible by the following observations.

Observation 3.5. *For any integer $r \geqslant 0$, two graphs G and H are isomorphic if and only if r-subd(G) and r-subd(H) are isomorphic.*

Observation 3.6. *Let F be a graph. For any integer $r > 0$, for every subgraph H of F, the graph r-subd(H) is isomorphic to an induced subgraph of r-subd(F).*

Observation 3.7. *Let G be a connected graph containing a vertex of degree at least 3, then*

$$\mathrm{aut}(r\text{-}\mathrm{subd}(G)) = \mathrm{aut}(G).$$

Proof. We start by showing that any automorphism φ of r-subd(G) maps branching vertices to branching vertices and subdivision vertices to subdivision vertices. First, note that as G has a vertex of degree at least 3, the graph r-subd(G) also has a vertex of degree at least 3 (hence a branching vertex) say u, which must be mapped by φ to a vertex of degree at least 3 (hence, to a branching vertex), say v. Now, since G is connected, the branching vertices are exactly the ones that are at distances $0 \pmod{r+1}$ from u, and they are mapped to vertices that are at distances $0 \pmod{r+1}$ from v, and thus, to branching vertices. Hence, the restriction $\tilde{\varphi}$ of φ to the branching vertices of r-subd(G) is a permutation of the vertices of G. We claim that $\tilde{\varphi}$ is an automorphism of G. Indeed, two vertices a and b are adjacent in G if and only if a and b are at distance $r+1$ in r-subd(G) if and only if $\varphi(a)$ and $\varphi(b)$ are at distance $r+1$ in r-subd(G) if and only if $\tilde{\varphi}(a) = \varphi(a)$ and $\tilde{\varphi}(b) = \varphi(b)$ are adjacent in G.

From this, we deduce that $\varphi \mapsto \tilde{\varphi}$ is a bijection between the automorphisms of r-subd(G) and the automorphisms of G: it is injective since an automorphism of r-subd(G) is uniquely determined by the images of the branching vertices, and surjective since there are at least as many automorphisms of r-subd(G) as there are of G. Hence, $\mathrm{aut}(r\text{-}\mathrm{subd}(G)) = \mathrm{aut}(G)$. □

Before proceeding to our main result, we provide some basic facts about automorphisms, which we will rely on. Given two graphs F and G, we denote by $\#\mathrm{Sub}(F \to G)$ the number of subgraphs

8

of G isomorphic to F, and by $\#\mathrm{Emb}(F \to G)$ the number of embeddings of F into G, i.e., the number of injective homomorphisms from F to G. Thus

$$\#\mathrm{Emb}(F \to G) = \#\mathrm{Sub}(F \to G) \cdot \mathrm{aut}(F).$$

For instance, if G is the cycle on an even number n of vertices, and F is the disjoint union of $n/2$ edges, then $\#\mathrm{Sub}(F \to G) = 2$ and $\mathrm{aut}(F) = (n/2)! \cdot 2^{\frac{n}{2}}$, so $\#\mathrm{Emb}(F \to G) = (n/2)! \cdot 2^{\frac{n}{2}+1}$.

We will use the following two well known facts (see [KLT02]).

Lemma 3.8. *Let F be a spanning subgraph of a graph G. Then*

$$\mathrm{aut}(G) \leqslant \#\mathrm{Emb}(F \to G) = \#\mathrm{Sub}(F \to G) \cdot \mathrm{aut}(F).$$

Lemma 3.9. *Let G be a connected graph of maximum degree Δ. Then*

$$\mathrm{aut}(G) \leqslant n \cdot \Delta! \cdot (\Delta - 1)^{n-\Delta-1} \leqslant n\Delta^n.$$

We apply the above two facts to upper bound the number of automorphisms of a graph containing a bounded-degree spanning tree in terms its average degree.

Lemma 3.10. *Let $d \geqslant 1$ and G be a connected n-vertex graph with at most dn edges, which has a spanning tree T of maximum degree τ. Then $\mathrm{aut}(G) \leqslant (6\tau d)^n$.*

Proof. By Lemma 3.9, $\mathrm{aut}(T) \leqslant n\tau^n \leqslant (2\tau)^n$. Moreover, $\#\mathrm{Sub}(T \to G)$ is at most the number of ways of choosing $n-1$ edges among $E(G)$. Hence

$$\#\mathrm{Sub}(T \to G) \leqslant \binom{dn}{n-1} \leqslant \left(\frac{edn}{n-1}\right)^{n-1} = \left(1 + \frac{1}{n-1}\right)^{n-1} \cdot (ed)^{n-1} \leqslant e(ed)^{n-1} \leqslant (ed)^n,$$

where in the penultimate inequality we used the fact that $1 + x \leqslant e^x$ for any real x. Hence, by Lemma 3.8, we have $\mathrm{aut}(G) \leqslant \#\mathrm{Sub}(T \to G) \cdot \mathrm{aut}(T) \leqslant (ed)^n \cdot (2\tau)^n \leqslant (6\tau d)^n$. $\qquad\square$

We are now ready to prove the main result of this section.

Theorem 1.1. *Every weakly sparse small class has bounded expansion.*

Proof. For the sake of contradiction, suppose that a weakly sparse graph class $\mathcal{X}$ has unbounded expansion, but there exists a constant c such that for every $n \in \mathbb{N}$ the number of labeled n-vertex graphs in $\mathcal{X}$ is at most $n! \cdot c^n$. Let $r \geqslant 1$ be an integer such that $\mathcal{X}$ contains r-subdivisions of graphs of arbitrary large average degree, which exists by Corollary 3.3.

Let $\delta := 24$, $d := \lceil \max(\delta^2, \delta c^{2r}) \rceil$. Let F be a graph of average degree at least $4d$, such that $r\text{-}\mathrm{subd}(F)$ is in $\mathcal{X}$, and let F' be a subgraph of F of minimum degree at least $2d$, which exists by Observation 2.1. By Lemma 3.4, F' contains a subgraph H' of minimum degree at least $2d$ with a spanning tree T of maximum degree $2d$. If H' has more than $d|V(H')|$ edges, we remove all but $d|V(H')| - (|V(H')| - 1)$ of them from the set $E(H') \setminus E(T)$ to obtain a spanning subgraph with exactly $d|V(H')|$ edges that still contains T as a spanning tree. We denote this subgraph by H and let $k := |V(H)|$, and thus $|E(H)| = dk$.

Let $\mathcal{H}$ be the set of spanning subgraphs of H that contain T as a subgraph and have exactly δk edges. Then, since $\delta \leqslant d$, we have

$$|\mathcal{H}| = \binom{dk - (k-1)}{\delta k - (k-1)} \geqslant \binom{dk - k}{\delta k - k} \geqslant \left(\frac{d-1}{\delta - 1}\right)^{(\delta - 1)k} \geqslant \left(\frac{d}{\delta}\right)^{(\delta - 1)k}.$$

9

Since every graph in $\mathcal{H}$ has $\delta k \leqslant dk$ edges and contains a spanning tree of maximum degree $2d$, from Lemma 3.10,

$$\max_{A \in \mathcal{H}} \operatorname{aut}(A) \leqslant (6 \cdot 2d \cdot d)^k \leqslant (4d)^{2k}. \tag{2}$$

Thus, denoting by $\mathcal{U}$ the set of pairwise non-isomorphic graphs in $\mathcal{H}$, we obtain

$$|\mathcal{U}| \geqslant \frac{|\mathcal{H}|}{\max\limits_{A \in \mathcal{H}} \operatorname{aut}(A)} \geqslant \left(\frac{d}{\delta}\right)^{(\delta-1)k} \cdot \frac{1}{(4d)^{2k}} = \left(\frac{d}{\delta}\right)^{(\delta+1)k} \cdot \frac{\delta^{2k}}{4^{2k} \cdot d^{4k}} = \left(\frac{d}{\delta}\right)^{(\delta+1)k} \cdot \frac{6^{2k}}{d^{4k}} \tag{3}$$

Let $r\text{-}\operatorname{subd}(\mathcal{U})$ be the set of r-subdivisions of graphs in $\mathcal{U}$, and observe that every graph in $r\text{-}\operatorname{subd}(\mathcal{U})$ has exactly $N := (r\delta+1)k$ vertices. Since every graph in $\mathcal{U}$ is a subgraph of F, and $r\text{-}\operatorname{subd}(F)$ is in $\mathcal{X}$, Observation 3.6 implies that every graph in $r\text{-}\operatorname{subd}(\mathcal{U})$ is in $\mathcal{X}$. Thus, to contradict the assumption that, for every $n \in \mathbb{N}$, class $\mathcal{X}$ has at most $n! \cdot c^n$ labeled n-vertex graphs, it is enough to show that $|\operatorname{Lab}(r\text{-}\operatorname{subd}(\mathcal{U}))| > c^N \cdot N!$. This is what we do in the remainder of the proof.

It follows from Observation 3.5 that $r\text{-}\operatorname{subd}(\mathcal{U})$ is a set of pairwise non-isomorphic graphs and $|r\text{-}\operatorname{subd}(\mathcal{U})| = |\mathcal{U}|$. Furthermore, since every graph A in $\mathcal{H}$ is connected and has k vertices and $\delta k \geqslant k+1$ edges, it contains a vertex of degree at least 3 and thus, by Observation 3.7, $\operatorname{aut}(r\text{-}\operatorname{subd}(A)) = \operatorname{aut}(A)$. Therefore, from (1), we have

$$|\operatorname{Lab}(r\text{-}\operatorname{subd}(\mathcal{U}))| \geqslant |r\text{-}\operatorname{subd}(\mathcal{U})| \cdot \frac{N!}{\max\limits_{A \in \mathcal{H}} \operatorname{aut}(r\text{-}\operatorname{subd}(A))} = |\mathcal{U}| \cdot \frac{N!}{\max\limits_{A \in \mathcal{H}} \operatorname{aut}(A)},$$

and thus, from (2) and (3), and the fact that $N/r = (\delta + 1/r)k \leqslant (\delta+1)k$, we obtain

$$|\operatorname{Lab}(r\text{-}\operatorname{subd}(\mathcal{U}))| \geqslant \left(\frac{d}{\delta}\right)^{N/r} \cdot \frac{6^{2k}}{d^{4k}} \cdot \frac{N!}{(4d)^{2k}} > \left(\frac{d}{\delta}\right)^{N/r} \cdot \frac{N!}{d^{6k}},$$

Finally, since $\frac{N}{4r} = \frac{(r\delta+1)k}{4r} > \frac{\delta k}{4} = 6k$, and d satisfies both $\sqrt{d} \geqslant \delta$ and $(d/\delta)^{1/r} \geqslant c^2$, we have

$$|\operatorname{Lab}(r\text{-}\operatorname{subd}(\mathcal{U}))| > \left(\frac{d}{\delta}\right)^{\frac{N}{2r}} \cdot \left(\frac{d}{\delta}\right)^{\frac{N}{2r}} \cdot \frac{N!}{d^{6k}} \geqslant \left(\frac{d}{\delta}\right)^{\frac{N}{2r}} \cdot d^{\frac{N}{4r}} \cdot \frac{N!}{d^{6k}} > \left(\left(\frac{d}{\delta}\right)^{1/r}\right)^{\frac{N}{2}} \cdot N! \geqslant c^N \cdot N!,$$

which completes the proof. $\qquad\square$

4 Short adjacency labels for small classes

In this section we prove:

Theorem 1.3. *Every hereditary small class admits an $O(\log^3 n)$-bit adjacency labeling scheme.*

To do so, we first show a result of independent interest (in Section 4.1), that the neighborhood complexity of any small class is at most $O(n \log n)$. Then, we prove in Section 4.2 that such neighborhood complexity implies $O(\log^2 n)$ contiguity. The latter together with Proposition 2.2 implies Theorem 1.3. The core part of this proof strategy is to show how the bound on the neighborhood complexity can be translated to the bound on contiguity. This is done via a classical result from the theory of set systems of bounded Vapnik-Chervonenkis dimension on so-called *paths with low crossing number* (Theorem 4.1) applied to neighborhood set systems. While the result about paths

with low crossing number is a known fact, we need it in somewhat unusual form that we could not find elsewhere in the literature. Therefore, in Appendix A we present a self-contained proof of this result in a suitable form.

4.1 Neighborhood complexity of small classes

In this section we show that small classes have low neighborhood complexity.

Theorem 1.4. *The neighborhood complexity of any hereditary small graph class is $O(n \log n)$.*

Proof. Let $\mathcal{C}$ be a hereditary small class of graphs and $c \geqslant 1$ be a constant such that $|\mathcal{C}_n| \leqslant n! \cdot c^n$ holds for every $n \in \mathbb{N}$. To prove the claim, we will show that for every graph H in $\mathcal{C}$ it holds that

$$\nu_H(m) < 9c^2 m \log(m+1) \qquad \text{for all } m \in [|V(H)|].$$

Suppose this is not the case. Then, by definition, there exists a graph $H \in \mathcal{C}$ and an n-element set $A \subseteq V(H)$ such that $|\{N_H(v) \cap A \ : \ v \in V(H)\}| \geqslant 9c^2 n \log(n+1)$. Note that the latter inequality implies $n \geqslant 4$. Let $B' \subseteq V(H)$ be any maximum-size set of vertices whose neighborhoods have pairwise distinct intersections with A in the graph H, and let $B := B' \setminus A$. Denote by G the subgraph of H induced by $A \cup B$. By the assumption, we have

1. $|A| = n$,

2. $|B| \geqslant |B'| - |A| \geqslant 9c^2 n \log(n+1) - n > 8c^2 n \log n$,

3. $A \cap B = \emptyset$, and

4. all vertices in B have pairwise distinct neighborhoods in A in graph G.

Let $n_1 := n \lfloor \log n \rfloor$ and $N := n_1 + n$. We will show that the hereditary closure of G contains more than $N! c^N$ labeled graphs on N vertices, i.e., that $|\text{Lab}(\text{IndSub}_N(G))| > N! c^N$, where $\text{IndSub}_N(G)$ is the set of N-vertex induced subgraphs of G. This would contradict our assumption on the number of graphs in $\mathcal{C}$. To do this, let us first fix a labeling of A using the integers from 1 to n. We claim that any two distinct labeled subsets $X_1, X_2 \subseteq B$ of size n_1, labeled with the integers from $n+1$ to N, induce two distinct labeled graphs $G_1 = G[A \cup X_1]$ and $G_2 = G[A \cup X_2]$ contained in $\mathcal{C}$. Indeed, since all vertices in B have pairwise distinct neighborhoods in A, G_1 and G_2 contain a vertex with the same label in $[n+1, N]$ but different neighborhoods in A (i.e. different neighborhoods among vertices with labels from $[n]$).

Hence, the number of N-vertex labeled graphs in $\mathcal{C}$ that are induced subgraphs of G is lower bounded by the number of ways to choose a labeled subset of size n_1 from B, which is

$$n_1! \cdot \binom{|B|}{n_1} \geqslant n_1! \cdot \left(\frac{|B|}{n_1}\right)^{n_1} > n_1! \cdot (8c^2)^{n_1}. \tag{4}$$

We want to show that this is more than the total number of N-vertex labeled graphs in $\mathcal{C}$, which, by assumption, does not exceed

$$N! \cdot c^N = (n_1 + n)! \cdot c^{n_1+n} \leqslant n_1! \cdot (n_1 + n)^n \cdot c^{2n_1}. \tag{5}$$

From (4) and (5), we should then show that

$$n_1! \cdot (8c^2)^{n_1} \geqslant n_1! \cdot (n_1 + n)^n \cdot c^{2n_1},$$

11

which is equivalent to establishing that

$$8^{n_1} = 2^{3n_1} \geqslant (n_1 + n)^n = 2^{n \log(n_1+n)}.$$

The latter indeed holds since

$$
\begin{aligned}
n \log(n_1 + n) &\leqslant n \log(2n \lfloor \log n \rfloor) \\
&\leqslant n \log 2 + n \log n + n \log \lfloor \log n \rfloor \\
&\leqslant n(1 + \log 2) + n \lfloor \log n \rfloor + n \log \lfloor \log n \rfloor \\
&\leqslant n \lfloor \log n \rfloor + n \lfloor \log n \rfloor + n \log \lfloor \log n \rfloor \\
&\leqslant n_1 + n_1 + n_1 = 3n_1,
\end{aligned}
$$

where the penultimate inequality holds because $n \geqslant 4$. $\qquad\square$

4.2 From neighborhood complexity to contiguity

In this section, we first show how to translate a bound on neighborhood complexity to a bound on contiguity (Theorem 4.3), and then apply it to obtain bounds on contiguity for hereditary small classes (Theorem 4.4). A crucial technical tool that we use to establish this translation is a suitable form of a known result about paths with low crossing number (Theorem 4.1). To state this result we must first introduce the required notions.

Let $(X, \mathcal{S})$ be a set system. A set $S \in \mathcal{S}$ *crosses* a 2-element set $\{x, y\} \subset X$ if exactly one of x and y belongs to S. Let $\mathcal{F}$ be a multiset of 2-element subsets in X. The *crossing number of $\mathcal{F}$ with respect to a set $S \in \mathcal{S}$* is the number of elements in $\mathcal{F}$ that are crossed by S. The *crossing number of $\mathcal{F}$ with respect to $\mathcal{S}$* is the maximum crossing number of $\mathcal{F}$ with respect to a set $S \in \mathcal{S}$. If $(X, \mathcal{F})$, considered as a graph, is a path spanning all elements of X, then we say that $(X, \mathcal{F})$ is a path on X. In this case, the crossing number of $\mathcal{F}$ is referred to as the crossing number of the corresponding path. We are now ready to state the main technical tool whose self-contained proof can be found in Appendix A.

Theorem 4.1. *Let X be an n-element set, $f : \mathbb{R}_{\geqslant 0} \to \mathbb{R}_{\geqslant 0}$ be a strictly increasing function, and d be a natural number. Let $(X, \mathcal{S})$ be a set system of VC dimension d that satisfies $\pi_{\mathcal{S}}^*(m) \leqslant f(m)$ for all $m \in [n]$. Then there exists a path on X with crossing number at most*

$$2 \log |\mathcal{S}| + 10d \cdot \sum_{j=1}^{n} \frac{1}{f^{-1}(j/2)}.$$

We will apply Theorem 4.1 to neighborhood set systems. Its importance in our context is due to (a) the fact that the dual shatter function of the neighborhood set system of a graph coincides with its primal dual function (Observation 2.3), i.e., the neighborhood complexity of the graph; and (b) the following observation about close relationship between the crossing number of a path for the neighborhood set system and the contiguity of the graph.

Observation 4.2. *Let G be an n-vertex graph. If there exists a path on $V(G)$ that has crossing number with respect to $\mathcal{N}_G$ at most k, then the contiguity of G is at most $k/2 + 1$.*

Proof. Let $\{\{v_1, v_2\}, \{v_2, v_3\}, \ldots, \{v_{n-1}, v_n\}\}$ be a path on $V(G)$ with crossing number with respect to $\mathcal{N}_G$ at most k. We claim that the linear ordering $\sigma = v_1, v_2, \ldots, v_n$ of $V(G)$ witnesses the target bound on the contiguity of G.

Let v be an arbitrary vertex in G and let N be its neighborhood. For convenience, mark a vertex of G by 1 if it belongs to N and 0 otherwise. This marking together with the ordering σ correspond to the binary word $w = w_1 w_2 \ldots w_n$, where w_i is the mark of v_i for every $i \in [n]$. A relation between pairs crossed and neighborhoods partitioned now follows from two observations:

1. in the word w, the maximal intervals of consecutive 1s correspond to intervals of consecutive vertices in σ that partition the neighborhood N of v;

2. each pair of consecutive letters $w_i w_{i+1}$ in w with $w_i \neq w_{i+1}$ corresponds to a crossing of $\{v_i, v_{i+1}\}$ by N.

To conclude, each interval in the partition of N corresponds to two crossings apart from the at most two intervals at each end of the ordering which each corresponds to one crossing. Thus the number of intervals is at most $2 + \lfloor \frac{k-2}{2} \rfloor \leqslant k/2 + 1$, as claimed. $\qquad\square$

The next theorem translates a bound on neighborhood complexity to a bound on contiguity. It follows directly from Theorem 4.1, Observation 4.2, Observation 2.3, and the definition of neighborhood complexity.

Theorem 4.3. *Let G be an n-vertex graph, $f : \mathbb{R}_{\geqslant 0} \to \mathbb{R}_{\geqslant 0}$ be a strictly increasing function, and d be a natural number. If $(V, \mathcal{N}_G)$ has VC dimension d and $\nu_G(m) \leqslant f(m)$ for all $m \in [n]$, then*

$$\mathrm{ctg}(G) \leqslant 1 + \log n + 5d \cdot \sum_{j=1}^{n} \frac{1}{f^{-1}(j/2)}.$$

We now apply Theorem 4.3 to hereditary small classes, which have $O(n \log n)$ neighborhood complexity by Theorem 1.4, to obtain a bound on their contiguity in closed form.

Theorem 4.4. *Let $\mathcal{X}$ be a hereditary small class, then the contiguity of $\mathcal{X}$ is $O(\log^2 n)$.*

Proof. Since $\mathcal{X}$ is a small class, by Theorem 1.4 there exists some constant C such that $\nu_{\mathcal{X}} \leqslant f(n)$ where $f(n) := Cn \log n$. Consequently, for every $n \in \mathbb{N}$ and every n-vertex graph $G \in \mathcal{X}$, we have $\pi_{\mathcal{N}_G}(m) \leqslant f(m)$ for all $m \in [n]$. Thus, by definition, the VC dimension d of $(V(G), \mathcal{N}_G)$ satisfies $2^d \leqslant Cd \log d$, which implies

$$d \leqslant \log C + \log d + \log \log d \leqslant \log C + 4d/5,$$

and therefore $d \leqslant 4 \log C$.

Now, by Theorem 4.3, we have

$$\mathrm{ctg}(G) \leqslant 1 + \log n + 20 \cdot \log C \cdot \sum_{j=1}^{n} \frac{1}{f^{-1}(j)}. \tag{6}$$

Since f is increasing on $[1, \infty)$ with image $[0, \infty)$, its inverse f^{-1} exists on $[0, \infty)$, and is also increasing. The exact inverse f^{-1} is somewhat complicated, in particular $f^{-1}(1) = e^{W(\ln(2)/C)}$, where $W(\cdot)$ is the Lambert function. We consider $f^{-1}(1) > 0$ just as a constant depending on C, and show that

$$f^{-1}(x) \geqslant \frac{x}{C \log x} \qquad \text{for all } x \geqslant 2. \tag{7}$$

To show (7), since $\log x \geqslant 1$ for all $x \geqslant 2$ and we can assume $C \geqslant 1$, we have

$$f\left(\frac{x}{C \log x}\right) = C \cdot \frac{x}{C \log x} \cdot \log\left(\frac{x}{C \log x}\right) \leqslant x. \tag{8}$$

Thus, applying f^{-1} to both sides of (8) gives $f^{-1}(x) \geqslant \frac{x}{C \log x}$, establishing (7).

Then, by (6),

$$\mathrm{ctg}(G) \leqslant 1 + \log n + 20 \cdot \log C \cdot \sum_{j=1}^{n} \frac{1}{f^{-1}(j)}$$

$$\leqslant 1 + \log n + 20 \cdot \log C \cdot e^{-W(\ln(2)/C)} + 20 \cdot \log C \cdot \sum_{j=2}^{n} \frac{C \log j}{j}$$

$$= \Theta(\log^2 n),$$

as claimed. □

The main Theorem 1.3 of this section is a consequence of Theorems 1.4 and 4.4 and Proposition 2.2.

5 Conclusion

In this paper we obtained strong evidence in support of the Small Implicit Graph Conjecture that posits that every hereditary small class admits an $O(\log n)$-bit adjacency labeling scheme. Specifically, we showed that (1) every weakly sparse small class admits such a labeling scheme; (2) every hereditary small class admits an $O(\log^3 n)$-bit adjacency labeling scheme. To obtain these results we established two structural properties of hereditary small classes that are of independent interest. The first property is that every weakly sparse small class has bounded degeneracy, and even, bounded expansion. The second property is that every hereditary small class has neighborhood complexity $O(n \log n)$.

The latter property leaves a tantalizing open question of whether the bound on the neighborhood complexity can be further improved. All hereditary small classes known to us have linear neighborhood complexity (e.g., classes of bounded twin-width [BKR+22, BFLP24]), which motivates the following:

Conjecture 5.1. *Every hereditary small class of graphs has neighborhood complexity $O(n)$.*

If this conjecture is true, our approach would imply $O(\log^2 n)$-bit labeling schemes for all hereditary small classes.

References

[AAA+23] Bogdan Alecu, Vladimir E. Alekseev, Aistis Atminas, Vadim Lozin, and Viktor Zamaraev. Graph parameters, implicit representations and factorial properties. *Discrete Mathematics*, 346(10):113573, 2023.

[ADK17]   Stephen Alstrup, Søren Dahlgaard, and Mathias Bæk Tejs Knudsen. Optimal induced universal graphs and adjacency labeling for trees. *Journal of the ACM (JACM)*, 64(4):1–22, 2017.

[Alo23]   Noga Alon. Implicit representation of sparse hereditary families. *Discret. Comput. Geom.*, to appear, 2023.

[ALR09]   Peter Allen, Vadim V. Lozin, and Michaël Rao. Clique-width and the speed of hereditary properties. *Electron. J. Comb.*, 16(1), 2009.

[ALR19]   Aistis Atminas, Vadim V Lozin, and Igor Razgon. Graphs without large bicliques and well-quasi-orderability by the induced subgraph relation. *Journal of Combinatorics*, 10(2):327–337, 2019.

[Ban22]   Avah Banerjee. An adjacency labeling scheme based on a decomposition of trees into caterpillars. In *Combinatorial Algorithms: 33rd International Workshop, (IWOCA 2022), proceedings*, pages 114–127. Springer, 2022.

[BDS⁺24a] Édouard Bonnet, Julien Duron, John Sylvester, Viktor Zamaraev, and Maksim Zhukovskii. Small but unwieldy: A lower bound on adjacency labels for small classes. In *Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2024)*, pages 1147 – 1165, 2024.

[BDS⁺24b] Édouard Bonnet, Julien Duron, John Sylvester, Viktor Zamaraev, and Maksim Zhukovskii. Tight bounds on adjacency labels for monotone graph classes. In *51st International Colloquium on Automata, Languages, and Programming, ICALP 2024*, volume 297 of *LIPIcs*, pages 31:1–20, 2024.

[BDSZ24]   Édouard Bonnet, Julien Duron, John Sylvester, and Viktor Zamaraev. Symmetric-Difference (Degeneracy) and Signed Tree Models. In *49th International Symposium on Mathematical Foundations of Computer Science (MFCS 2024)*, volume 306 of *LIPIcs*, pages 32:1–32:16, 2024.

[BFLP24]   Édouard Bonnet, Florent Foucaud, Tuomo Lehtilä, and Aline Parreau. Neighbourhood complexity of graphs of bounded twin-width. *European Journal of Combinatorics*, 115:103772, 2024.

[BGK⁺22]   Édouard Bonnet, Colin Geniet, Eun Jung Kim, Stéphan Thomassé, and Rémi Watrigant. Twin-width II: small classes. *Combinatorial Theory, 2 (2)*, 2022.

[BKR⁺22]   Édouard Bonnet, Eun Jung Kim, Amadeus Reinald, Stéphan Thomassé, and Rémi Watrigant. Twin-width and polynomial kernels. *Algorithmica*, 84(11):3300–3337, 2022.

[BL76]   Kellogg S Booth and George S Lueker. Testing for the consecutive ones property, interval graphs, and graph planarity using pq-tree algorithms. *Journal of computer and system sciences*, 13(3):335–379, 1976.

[BP69]   Lowell W Beineke and Raymond E Pippert. The number of labeled k-dimensional trees. *Journal of Combinatorial Theory*, 6(2):200–205, 1969.

[DEG⁺21]   Vida Dujmović, Louis Esperet, Cyril Gavoille, Gwenaël Joret, Piotr Micek, and Pat Morin. Adjacency labelling for planar graphs (and beyond). *Journal of the ACM (JACM)*, 68(6):1–33, 2021.

[DEM+23] Jan Dreier, Ioannis Eleftheriadis, Nikolas Mählmann, Rose McCarty, Michał Pilipczuk, and Szymon Toruńczyk. First-order model checking on monadically stable graph classes. *arXiv preprint arXiv:2311.18740*, 2023.

[DKT13] Zdenek Dvorák, Daniel Král, and Robin Thomas. Testing first-order properties for subclasses of sparse graphs. *J. ACM*, 60(5):36:1–36:24, 2013.

[DMT24] Jan Dreier, Nikolas Mählmann, and Szymon Toruńczyk. Flip-breakability: A combinatorial dichotomy for monadically dependent graph classes. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, STOC 2024, page 1550–1560, New York, NY, USA, 2024. Association for Computing Machinery.

[DN10] Zdeněk Dvořák and Serguei Norine. Small graph classes and bounded expansion. *J. Comb. Theory, Ser. B*, 100(2):171–175, 2010.

[DN16] Zdenek Dvorák and Sergey Norin. Strongly sublinear separators and polynomial expansion. *SIAM J. Discret. Math.*, 30(2):1095–1101, 2016.

[Dud78] R. M. Dudley. Central limit theorems for empirical measures. *Ann. Probab.*, 6(6):899–929, 1978.

[Dvo07] Zdeněk Dvořák. Asymptotical structure of combinatorial objects. *PhD thesis*, 2007.

[Dvo18] Zdeněk Dvořák. Induced subdivisions and bounded expansion. *Eur. J. Comb.*, 69:143–148, 2018.

[EGK+17] Kord Eickmeyer, Archontia C. Giannopoulou, Stephan Kreutzer, O-joung Kwon, Michal Pilipczuk, Roman Rabinovich, and Sebastian Siebertz. Neighborhood complexity and kernelization for nowhere dense classes of graphs. In *44th International Colloquium on Automata, Languages, and Programming, ICALP 2017*, volume 80 of *LIPIcs*, pages 63:1–63:14, 2017.

[FG65] Delbert Fulkerson and Oliver Gross. Incidence matrices and interval graphs. *Pacific journal of mathematics*, 15(3):835–855, 1965.

[FP14] Jacob Fox and János Pach. Applications of a new separator theorem for string graphs. *Comb. Probab. Comput.*, 23(1):66–74, 2014.

[GGKS95] Paul W Goldberg, Martin C Golumbic, Haim Kaplan, and Ron Shamir. Four strikes against physical mapping of DNA. *Journal of Computational Biology*, 2(1):139–152, 1995.

[GHL16] Ralucca Gera, Stephen Hedetniemi, and Craig Larson, editors. *Graph theory—favorite conjectures and open problems. 1.* Problem Books in Mathematics. Springer, 2016.

[GL07] Cyril Gavoille and Arnaud Labourel. Shorter implicit representation for planar graphs and bounded treewidth graphs. In *15th Annual European Symposium on Algorithms, ESA 2007*, volume 4698 of *LNCS*, pages 582–593. Springer, 2007.

[GW00] Frank Gurski and Egon Wanke. The tree-width of clique-width bounded graphs without $K_{n,n}$. In *Graph-Theoretic Concepts in Computer Science, 26th International Workshop, WG 2000*, volume 1928 of *LNCS*, pages 196–205. Springer, 2000.

[Han82]   Phil Hanlon. Counting interval graphs. *Transactions of the American Mathematical Society*, 272(2):383–426, 1982.

[Hau95]   David Haussler. Sphere Packing Numbers for Subsets of the Boolean n-Cube with Bounded Vapnik-Chervonenkis Dimension. *J. Comb. Theory, Ser. A*, 69(2):217–232, 1995.

[HH22]   Hamed Hatami and Pooya Hatami. The implicit graph conjecture is false. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS 2022)*, pages 1134–1137. IEEE, 2022.

[KLT02]   Ilia Krasikov, Arieh Lev, and Bhalchandra D Thatte. Upper bounds on the automorphism group of a graph. *Discrete Mathematics*, 256(1-2):489–493, 2002.

[KNR88]   Sampath Kannan, Moni Naor, and Steven Rudich. Implicit representation of graphs. In *Proceedings of the twentieth annual ACM symposium on Theory of computing (STOC 1988)*, pages 334–343, 1988.

[KNR92]   Sampath Kannan, Moni Naor, and Steven Rudich. Implicit representation of graphs. *SIAM Journal on Discrete Mathematics*, 5(4):596–603, 1992.

[KO04]   Daniela Kühn and Deryk Osthus. Induced subdivisions in $K_{s,s}$-free graphs of large average degree. *Comb.*, 24(2):287–304, 2004.

[KST54]   Tamás Kővári, Vera T. Sós, and Pál Turán. On a problem of Zarankiewicz. In *Colloquium Mathematicum*, volume 3, pages 50–57. Polska Akademia Nauk, 1954.

[Mat10]   Jiří Matoušek. *Geometric discrepancy*, volume 18 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, 2010. An illustrated guide, Revised paperback reprint of the 1999 original.

[Mul88]   John H. Muller. *Local Structure in Graph Classes*. PhD thesis, Georgia Institute of Technology, 1988.

[NO12]   Jaroslav Nešetřil and Patrice Ossona de Mendez. *Sparsity - Graphs, Structures, and Algorithms*, volume 28 of *Algorithms and combinatorics*. Springer, 2012.

[NSTW06]   Serguei Norine, Paul Seymour, Robin Thomas, and Paul Wollan. Proper minor-closed families are small. *Journal of Combinatorial Theory, Series B*, 96(5):754–757, 2006.

[Ott48]   Richard Otter. The number of trees. *Ann. of Math. (2)*, 49:583–599, 1948.

[RVS19]   Felix Reidl, Fernando Sánchez Villaamil, and Konstantinos Stavropoulos. Characterising bounded expansion by neighbourhood complexity. *European Journal of Combinatorics*, 75:152–168, 2019.

[Spi03]   Jeremy P Spinrad. *Efficient Graph Representations*, volume 19. Fields Institute monographs. American Mathematical Soc., 2003.

[Tut62]   William Thomas Tutte. A census of planar triangulations. *Canadian Journal of Mathematics*, 14:21–38, 1962.

[War16]   Algorithms, Logic and Structure Workshop in Warwick – Open Problem Session, 2016.

[Wel88]   Emo Welzl. Partition trees for triangle counting and other range searching problems. In Herbert Edelsbrunner, editor, *Proceedings of the Fourth Annual Symposium on Computational Geometry (SoCG 1988)*, pages 23–33. ACM, 1988.

A   Paths with low crossing number

The main goal of this section is to provide a self-contained proof of Theorem 4.1. To state the theorem, we recall definitions of some crucial notions. Given a set system $(X, \mathcal{S})$, we say that a set $S \in \mathcal{S}$ *crosses* a 2-element set $\{x, y\} \subset X$ if exactly one of x and y belongs to S. Let $\mathcal{F}$ be a multiset of 2-element subsets in X. The crossing number of $\mathcal{F}$ with respect to a set $S \in \mathcal{S}$ is the number of elements in $\mathcal{F}$ that are crossed by S. The crossing number of $\mathcal{F}$ with respect to $\mathcal{S}$ is the maximum crossing number of $\mathcal{F}$ with respect to a set $S \in \mathcal{S}$. If $(X, \mathcal{F})$, considered as a graph, is a tree or a path spanning all elements of X, then we say that $(X, \mathcal{F})$ is a tree or a path on X, respectively. In this case, the crossing number of $\mathcal{F}$ is referred to as the crossing number of the corresponding tree or path, respectively.

Theorem 4.1. *Let X be an n-element set, $f : \mathbb{R}_{\geqslant 0} \to \mathbb{R}_{\geqslant 0}$ be a strictly increasing function, and d be a natural number. Let $(X, \mathcal{S})$ be a set system of VC dimension d that satisfies $\pi_{\mathcal{S}}^*(m) \leqslant f(m)$ for all $m \in [n]$. Then there exists a path on X with crossing number at most*

$$2 \log |\mathcal{S}| + 10d \cdot \sum_{j=1}^{n} \frac{1}{f^{-1}(j/2)}.$$

This theorem is a known result stated in an unusual form, but one that is useful for us. The main feature of the above formulation is that the bound on the dual shatter function is given in a general form (expressed as a function f) rather than polynomial; in turn, the bound on the crossing number is expressed via the inverse of f. More specifically, the above formulation has two main differences compared to the standard one (for instance [Mat10, Wel88]). First, in a standard formulation, the bound on the dual shatter function is $\pi_{\mathcal{S}}^*(m) \leqslant m^r$ for some constant $r > 1$. Second, the resulting bound on the crossing number of a path on X is in the form of $O(n^{1-1/r})$. Note that if we take $f(m)$ to be m^r in Theorem 4.1, then $f^{-1}(m) = m^{1/r}$ and

$$2 \log |\mathcal{S}| + 10d \cdot \sum_{j=1}^{n} \frac{1}{(j/2)^{1/r}}$$

can be upper bounded by $O(n^{1-1/r})$, which recovers the result in a standard form.

The benefit of the present formulation is that one can apply Theorem 4.1 when a preferable upper bound on $\pi_{\mathcal{S}}^*(m)$ is non-polynomial. For example, if $f(m) = m \cdot g(m)$ for some non-decreasing function $g(m)$, then Theorem 4.1 gives a bound of $O(g(n) \cdot \log n)$ on the crossing number. In particular, this becomes useful when $g(m) = m^{o(1)}$; e.g., if $g(m) = \log(m)$, the bound on the crossing number becomes $O(\log^2 n)$.

To provide a self-contained proof of Theorem 4.1, we need to state a number of auxiliary known results in a suitable (non-standard) form and repeat their proofs with some adjustments. Before each statement we explain how its formulation and proof differs from standard ones. We hope this exposition might be of use in some other work.

Packing lemma.   We begin with Haussler's Packing Lemma [Hau95] (Lemma A.1), which is an improvement over a quantitatively weaker Packing Lemma obtained by Dudley [Dud78] and re-

discovered by Welzl [Wel88]. The proof below is a simplification by Chazelle of Haussler's proof [Hau95], which is taken from Matoušek's book [Mat10].

A set system $(X, \mathcal{S})$ is δ-*separated* if for any two distinct sets $S_1, S_2 \in \mathcal{S}$ their symmetric difference contains at least δ elements, i.e., $|(S_1 \setminus S_2) \cup (S_2 \setminus S_1)| \geqslant \delta$. In a standard formulation, the lemma assumes that the shatter function $\pi_{\mathcal{S}}(m)$ is bounded by $O(m^r)$ for some r, and the bound on the cardinality of a δ-separated set $\mathcal{P}$ is $O((n/\delta)^r)$. For our purposes, in Lemma A.1, we state the bound on $|\mathcal{P}|$ directly in terms of the shatter function. Note that the standard form of the lemma follows from ours if one uses the bound $\pi_{\mathcal{S}}(m) = O(m^r)$. The difference of the proof below from the one in [Mat10] is that in (13) we bound t in terms of the shatter function directly, rather than in terms of the bound on the shatter function.

Lemma A.1 (Packing Lemma, see [Mat10, Lemma 5.14]). *Let $(X, \mathcal{S})$ be a set system on an n-element set of VC dimension d. Let $\delta \in [n]$ be an integer, and let $\mathcal{P} \subseteq \mathcal{S}$ be δ-separated. Then,*

$$|\mathcal{P}| \leqslant 2 \cdot \pi_{\mathcal{S}} \left(\left\lceil \frac{4dn}{\delta} \right\rceil \right).$$

Before we prove this lemma we must introduce an auxiliary graph construction. For a set system $(X, \mathcal{S})$, we define the *unit distance graph* $\mathrm{UD}(\mathcal{S})$ to be the graph with

$$V(\mathrm{UD}(\mathcal{S})) = \mathcal{S} \qquad \text{and} \qquad E(\mathrm{UD}(\mathcal{S})) = \{\{S, S'\} : |S \Delta S'| = 1\}.$$

We will also need the following bound on the number of its edges.

Lemma A.2 ([Mat10, Lemma 5.15]). *If $(X, \mathcal{S})$ is a set system of VC dimension d on a finite set X then the unit distance graph $\mathrm{UD}(\mathcal{S})$ has at most $d|\mathcal{S}|$ edges.*

With this, we are now ready to prove the Packing Lemma.

Proof of Lemma A.1. Choose a random s-element subset $A \subseteq X$, where s is given by

$$s := \left\lceil \frac{4dn}{\delta} \right\rceil. \tag{9}$$

Set $\mathcal{Q} := \{S \cap A : S \in \mathcal{P}\}$, and for each set $Q \in \mathcal{Q}$ define its weight $w(Q)$ as the number of sets $S \in \mathcal{P}$ with $S \cap A = Q$. Note that

$$\sum_{Q \in \mathcal{Q}} w(Q) = |\mathcal{P}|.$$

Let $E := E(\mathrm{UD}(\mathcal{Q}))$ be the edge set of the unit distance graph $\mathrm{UD}(\mathcal{Q})$, and define an edge weighting of $\mathrm{UD}(\mathcal{Q})$ by giving each edge $e = \{Q, Q'\} \in E$ weight $w(e) := \min(w(Q), w(Q'))$. Let

$$W = \sum_{e \in E} w(e),$$

where we note that this random variable depends on the random choice of A. The desired bound on $|\mathcal{P}|$ in the Packing Lemma is obtained by estimating the expectation of W in two different ways.

First, we claim that for any set $A \subseteq X$, we have the following (deterministic) bound on W

$$W \leqslant 2d \cdot \sum_{Q \in \mathcal{Q}} w(Q) = 2d \cdot |\mathcal{P}|. \tag{10}$$

To see this, note first that the VC dimension of $\mathcal{P}$ is at most d, and thus, by Lemma A.2, the unit distance graph $\mathrm{UD}(\mathcal{Q})$ has some vertex Q_0 of degree at most $2d$. By removing Q_0, the total edge weight drops by at most $2d \cdot w(Q_0)$, and we are left with a unit distance graph on $(X, \mathcal{Q}')$, where $\mathcal{Q}' = \mathcal{Q} \setminus Q_0$ and $\mathcal{Q}'$ has VC dimension at most d. We can thus repeat this process until no vertices are left, and we see that the sum of edge weights removed is at most $2d$ times the sum of vertex weights, as claimed.

Next, we bound the expectation $\mathbb{E}[W]$ from below by considering the following random experiment. First, we choose a random $(s-1)$-element set $A \subseteq X$, and then we choose a random element $a \in X \setminus A'$. The set $A = A' \cup \{a\}$ is a uniform random s-element subset of X, we consider $\mathrm{UD}(\mathcal{Q})$ with the same edge weighting as above. Each edge of $\mathrm{UD}(\mathcal{Q})$ is a pair of sets of $\mathcal{Q}$ differing in exactly one element of A. We let $E_1 \subseteq E$ be the edges for which the difference element is a, and let W_1 be the sum of their weights. By symmetry, we have

$$\mathbb{E}[W] = s \cdot \mathbb{E}[W_1]. \tag{11}$$

We are going to bound $\mathbb{E}[W_1]$ from below. Let $A' \subset X$ be an arbitrary but fixed $(s-1)$-element set. We estimate the conditional expectation $\mathbb{E}[W_1 \mid A']$; that is, the expected value of W_1 when conditioned on a fixed set A', and $a \in X \setminus A'$ is selected uniformly at random. Divide the sets of $\mathcal{P}$ into equivalence classes $\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_t$ according to their intersection with the set A'. By the definition of $\pi_{\mathcal{S}}(x)$ and since it is non-decreasing, we have

$$t \leqslant \pi_{\mathcal{S}}(s - 1) \leqslant \pi_{\mathcal{S}}(s). \tag{12}$$

Let $\mathcal{P}_i$ be one of the equivalence classes, and $b := |\mathcal{P}_i|$. Suppose that an element $a \in X \setminus A'$ has been chosen. If b_1 sets of $\mathcal{P}_i$ contain a and $b_2 = b - b_1$ sets do not contain a, then the class $\mathcal{P}_i$ gives rise to an edge of E_1 of weight $\min(b_1, b_2)$. Note that b_1 and b_2 depend on the choice of a while b does not.

For any non-negative real numbers b_1, b_2 with $b_1 + b_2 = b$, we have $\min(b_1, b_2) \geqslant b_1 b_2 / b$. The value $b_1 b_2$ is the number of ordered pairs of sets (S_1, S_2) with S_1, S_2 being two sets from the class $\mathcal{P}_i$ where one contains a and the other one does not. Now, since $\mathcal{P}$ is δ-separated by hypothesis, if $S_1, S_2 \in \mathcal{P}_i$ are two distinct sets, then they differ in at least δ elements. Therefore, the probability that S_1 and S_2 differ in a random element $a \in X \setminus A'$ is at least $\frac{\delta}{n-s+1} \geqslant \delta/n$. Hence the expected contribution of each pair (S_1, S_2) of distinct sets of $\mathcal{P}_i$ to the quantity $b_1 b_2$ is at least δ/n, thus

$$\mathbb{E}\left[b_1 b_2 \mid A'\right] \geqslant b(b-1) \cdot \frac{\delta}{n}.$$

Consequently, the expected contribution $\mathbb{E}\left[\min(b_1, b_2) \mid A'\right]$ of the equivalence class $\mathcal{P}_i$ to the sum of edge weights W_1 is at least $\mathbb{E}\left[b_1 b_2 / b \mid A'\right] \geqslant (b-1)\delta/n$. Summing up over all equivalence classes, and using the bound on t from (12),

$$\mathbb{E}[W_1] \geqslant \mathbb{E}\left[W_1 \mid A'\right] \geqslant \frac{\delta}{n} \sum_{i=1}^{t} (|\mathcal{P}_i| - 1) \geqslant \frac{\delta}{n}(|\mathcal{P}| - t) \geqslant \frac{\delta}{n}(|\mathcal{P}| - \pi_{\mathcal{S}}(s)). \tag{13}$$

Combining this with the estimate (10), the equality (11), and the definition (9) of s, leads to

$$2d \cdot |\mathcal{P}| \geqslant s \cdot \frac{\delta}{n} \cdot (|\mathcal{P}| - \pi_{\mathcal{S}}(s)) \geqslant 4d \cdot (|\mathcal{P}| - \pi_{\mathcal{S}}(s)).$$

Rearranging gives $|\mathcal{P}| \leqslant 2 \cdot \pi_{\mathcal{S}}(s)$, so the statement follows from (9). $\qquad\square$

Short Edge Lemma. A standard form of the Short Edge Lemma assumes a polynomial bound on the dual shatter function, i.e., $\pi_{\mathcal{S}}^*(m) = O(m^r)$ for some constant $r > 1$, and the resulting bound is expressed via the bound $O(m^{1/r})$ on the inverse of $\pi_{\mathcal{S}}^*(m)$. In our presentation of the Short Edge Lemma (Lemma A.3) below we use a bound on the dual shatter function in a general from expressed as a function f, and the resulting bound is stated in terms of f^{-1}, the inverse of f. As before, by plugging in the bound $\pi_{\mathcal{S}}^*(m) = O(m^r)$ in Lemma A.3 one would recover the standard form of the lemma. The proof is taken from [Mat10] with suitable adjustments.

Lemma A.3 (Short Edge Lemma, see [Mat10, Lemma 5.18]). *Let X be an n-element set, $f : \mathbb{R}_{\geqslant 0} \to \mathbb{R}_{\geqslant 0}$ be a strictly increasing function, and d be a natural number. Let $(X, \mathcal{S})$ be a set system of VC dimension d with $\pi_{\mathcal{S}}^*(m) \leqslant f(m)$ for all $m \in [n]$. Then, for any multiset $\mathcal{Q}$ with elements from $\mathcal{S}$, there exist elements $x, y \in X$ such that the edge $\{x, y\}$ is crossed by at most*

$$5d \cdot \frac{|\mathcal{Q}|}{f^{-1}(n/2)}$$

sets of $\mathcal{Q}$.

Proof. We form a set system $\mathcal{D}$ which is dual to $\mathcal{Q}$. That is, we consider the multiset $\mathcal{Q}$ as the ground set (if some set appears in $\mathcal{Q}$ several times, then it is considered with the appropriate multiplicity). For each $x \in X$, we let D_x be the set of all sets of $\mathcal{Q}$ containing x, and we set

$$\mathcal{D} = \{D_x : x \in X\}.$$

The symmetric difference $D_x \Delta D_y$ of two sets from $\mathcal{D}$ consists of the sets in $\mathcal{Q}$ crossing the pair $\{x, y\}$. We thus want to show that $D_x \Delta D_y$ is small for some $x \neq y$. We may assume $D_x \neq D_y$ for $x \neq y$, as otherwise we are done, and hence $|\mathcal{D}| = |X| = n$.

The primal shatter function $\pi_{\mathcal{D}}$ is certainly no larger than the dual shatter function $\pi_{\mathcal{S}}^*$, and hence $\pi_{\mathcal{D}}(m) \leqslant f(m)$ by the assumption on $\pi_{\mathcal{S}}^*$. Suppose that any two sets $D_x, D_y \in \mathcal{D}$ have symmetric difference at least δ, then the packing lemma (Lemma A.1) implies

$$n = |\mathcal{D}| \leqslant 2 \cdot \pi_{\mathcal{S}}^* \left(\left\lceil \frac{4d|\mathcal{Q}|}{\delta} \right\rceil \right) \leqslant 2f \left(\left\lceil \frac{4d|\mathcal{Q}|}{\delta} \right\rceil \right).$$

Since f is strictly increasing, f^{-1} is also strictly increasing. Therefore, we have

$$f^{-1}(n/2) \leqslant \left\lceil \frac{4d|\mathcal{Q}|}{\delta} \right\rceil \leqslant \frac{5d|\mathcal{Q}|}{\delta},$$

and therefore

$$\delta \leqslant 5d \cdot \frac{|\mathcal{Q}|}{f^{-1}(n/2)},$$

as claimed. $\qquad\square$

Trees with low crossing number. As with the Short Edge Lemma, the following result (Theorem A.4) is stated with a general bound f on the dual shatter function and the resulting bound on the crossing number of a tree is expressed in terms of f^{-1}, the inverse of f. The proof is taken from [Mat10], which is very similar to Welzl's original proof [Wel88]. The statement in [Mat10] is for crossing number of matchings, and therefore it is slightly modified to work for trees as in [Wel88].

Theorem A.4 (see [Mat10, Lemma 5.17]). *Let X be an n-element set, $f : \mathbb{R}_{\geqslant 0} \to \mathbb{R}_{\geqslant 0}$ be a strictly increasing function, and d be a natural number. Let $(X, \mathcal{S})$ be a set system of VC dimension d with $\pi_{\mathcal{S}}^*(m) \leqslant f(m)$ for all $m \in [n]$. Then there exists a tree T on X with crossing number at most*

$$\log |\mathcal{S}| + 5d \cdot \sum_{j=1}^{n} \frac{1}{f^{-1}(j/2)}.$$

Proof. We build T by selecting edges one by one according to the following strategy.

1. Suppose $\{u_1, v_1\}, \ldots, \{u_i, v_i\}$ have already been selected.

2. Define the weight $w_i(S)$ of a set $S \in \mathcal{S}$ as $2^{k_i(S)}$, where $k_i(S)$ is the number of edges among $\{u_1, v_1\}, \ldots, \{u_i, v_i\}$ crossed by S. In particular, $w_0(S) = 1$ for all $S \in \mathcal{S}$.

3. We select the next edge $\{u_{i+1}, v_{i+1}\}$ from $X_i = X \setminus \{u_1, u_2, \ldots, u_i\}$ as a pair of points with the total weight of sets crossing $\{u_{i+1}, v_{i+1}\}$ being the smallest possible.

4. We continue in this manner until $n - 1$ edges have been selected.

Let $E := \{\{u_1, v_1\}, \{u_2, v_2\} \ldots, \{u_{n-1}, v_{n-1}\}\}$. Then $T = (X, E)$ is a tree. Next, we will bound the crossing number k of the resulting tree T. To this end, we estimate the final total weight of all sets of $\mathcal{S}$, i.e.,

$$w_{n-1}(\mathcal{S}) = \sum_{S \in \mathcal{S}} w_{n-1}(S).$$

By definition of w_{n-1} we have $k \leqslant \log w_{n-1}(\mathcal{S})$.

Let us investigate how $w_{i+1}(\mathcal{S})$ increases compared to $w_i(\mathcal{S})$. Let $\mathcal{S}_{i+1}$ denote the collection of the sets of $\mathcal{S}$ crossing $\{u_{i+1}, v_{i+1}\}$. For each set in $\mathcal{S}_{i+1}$ the weight increases by a multiplicative factor of two, and for the other sets it remains unchanged. From this we get

$$w_{i+1}(\mathcal{S}) = w_i(\mathcal{S}) - w_i(\mathcal{S}_{i+1}) + 2w_i(\mathcal{S}_{i+1}) = w_i(\mathcal{S}) \left[1 + \frac{w_i(\mathcal{S}_{i+1})}{w_i(\mathcal{S})} \right].$$

We now estimate $\frac{w_i(\mathcal{S}_{i+1})}{w_i(\mathcal{S})}$ using the Short Edge Lemma (Lemma A.3). To begin we define a multiset $\mathcal{Q}_i$ of sets by adding $S \cap X_i$ to $\mathcal{Q}_i$ with multiplicity $w_i(S)$ for every $S \in \mathcal{S}$. Thus

$$|\mathcal{Q}_i| = \sum_{S \in \mathcal{S}} w_i(S) = w_i(\mathcal{S}).$$

Note that, by the construction of $\mathcal{Q}_i$, the number of sets in $\mathcal{Q}_i$ that cross $\{u_{i+1}, v_{i+1}\}$ is $w_i(\mathcal{S}_{i+1})$. Furthermore, by our choice of edge in (2) and (3), the edge $\{u_{i+1}, v_{i+1}\}$ is crossed by the minimum number of sets in $\mathcal{Q}_i$. Thus, by applying the Short Edge Lemma on the set system $(X_i, \{S \cap X_i \ : \ S \in \mathcal{S}\})$ to the multiset $\mathcal{Q}_i$, and defining $n_i := |X_i| = n - i$, we obtain:

$$w_i(\mathcal{S}_{i+1}) \leqslant 5d \cdot \frac{w_i(\mathcal{S})}{f^{-1}(n_i/2)}.$$

Hence, we have

$$w_{i+1}(\mathcal{S}) \leqslant w_i(\mathcal{S}) \left[1 + \frac{5d}{f^{-1}((n-i)/2)} \right].$$

Therefore

$$w_{n-1} \leqslant w_0(\mathcal{S}) \cdot \prod_{i=0}^{n-1} \left[1 + \frac{5d}{f^{-1}((n-i)/2)} \right] = |\mathcal{S}| \cdot \prod_{j=1}^{n} \left[1 + \frac{5d}{f^{-1}(j/2)} \right].$$

Taking logarithms of both sides and using the inequality $\ln(1 + x) \leqslant x$ we obtain

$$k \leqslant \log w_{n-1}(\mathcal{S}) \leqslant \log |\mathcal{S}| + 5d \cdot \sum_{j=1}^{n} \frac{1}{f^{-1}(j/2)},$$

as claimed. $\qquad\qquad\square$

From a tree to a path.   Finally, Theorem 4.1 is obtained from Theorem A.4 by using the following Lemma A.5 to "convert" a tree with crossing number k to a path with crossing number at most $2k$. The argument is taken from [Wel88] and we provided it here for completeness.

Lemma A.5 ([Wel88, Lemma 3.3]). *Let $(X, \mathcal{S})$ be a set system, and T be a tree on X with crossing number at most k with respect to $\mathcal{S}$. Then, there exists a path on X with crossing number at most $2k$ with respect to $\mathcal{S}$.*

Proof. We begin with the following claim.

Claim A.6. Given any family $\mathcal{F}$ of two element sets on X, replacing any pair $\{x, y\}, \{y, z\} \in \mathcal{F}$ by the set $\{x, z\}$ will not increase the crossing number with respect to $\mathcal{S}$.

Proof of Claim. If any $S \in \mathcal{S}$ crosses $\{x, z\}$, then exactly one of x or y is not contained in S. Thus, S must cross one of $\{x, y\}$ or $\{y, z\}$. $\qquad\qquad\diamond$

Let $\mathcal{D} = \{\{u_1, u_2\}, \{u_2, u_3\}, \ldots, \{u_{\ell-1}, u_\ell\}\}$ be a multiset consisting of edges from T in the order they are traversed in a depth-first search (DFS) tour[4] of T starting from an arbitrary vertex. Additionally, let $x_1, \ldots, x_n$ be a labeling of X in the order they are discovered by the DFS tour $\mathcal{D}$.

Note that each edge of T appears exactly twice in $\mathcal{D}$, thus the crossing number of $\mathcal{D}$ is at most $2k$. We can now apply Claim A.6 iteratively to reduce $\mathcal{D}$ to $\{\{x_1, x_2\}, \{x_2, x_3\}, \ldots, \{x_{n-1}, x_n\}\}$ without increasing the crossing number with respect to $\mathcal{S}$. $\qquad\qquad\square$

---

[4]A DFS tour visits all vertices of the tree starting from a given vertex and going as far as possible down a given branch, then backtracking until it finds an unexplored path.