

Approximate Distance Labeling Schemes

(Extended Abstract)

Cyril Gavoille¹, Michal Katz², Nir A. Katz²,
Christophe Paul¹, and David Peleg³

- ¹ LaBRI, Université Bordeaux I, 351, cours de la Libération, 33405 Talence Cedex, France. {gavoille,paul}@labri.fr.
² Dept. of Applied Mathematics, Bar Ilan University, Ramat Gan, 52900, Israel.
³ Dept. of Computer Science and Applied Mathematics, The Weizmann Institute of Science, Rehovot, 76100 Israel. peleg@wisdom.weizmann.ac.il.

Abstract. We consider the problem of labeling the nodes of an n -node graph G with short labels in such a way that the distance between any two nodes u, v of G can be approximated efficiently (in constant time) by merely inspecting the labels of u and v , without using any other information. We develop such constant approximate distance labeling schemes for the classes of trees, bounded treewidth graphs, planar graphs, k -chordal graphs, and graphs with a dominating pair (including for instance interval, permutation, and AT-free graphs). We also establish lower bounds, and prove that most of our schemes are optimal in terms of the length of the labels generated and the quality of the approximation.

Keywords: Local representations, approximate distance, labeling schemes, distributed data structures.

1 Introduction

1.1 Motivation

Common network representations are usually global in nature; in order to derive a useful piece of information, one must have access to a global data structure representing the entire network, even if the sought piece of information is local, pertaining to only few nodes.

In contrast, the notion of *labeling schemes* [2,1,12,17,11] [10] involves using a more *localized* representation scheme for the network. The idea is to label the nodes in a way that will allow one to infer information concerning any two nodes *directly* from their labels, without using *any* additional information sources.

Clearly, for such a labeling scheme to be useful, it should use relatively *short* labels (say, of length polylogarithmic in n), and yet allow efficient (say, polylogarithmic time) information deduction. Recently, this natural idea was studied with respect to capturing *distance* information. This has led to the notion of *distance labeling schemes*, which are schemes possessing the ability to determine the distance between two nodes efficiently (i.e., in polylogarithmic time) given

their labels [17]. The applicability of distance labeling schemes in the context of communication networks has been illustrated in [17], and various aspects of such schemes were further studied in [10].

Observe that efficient *exact* distance labeling schemes may not exist for *every* graph family. In particular, for a family of $\Omega(\exp(n^{1+\epsilon}))$ non-isomorphic n -vertex graphs, any distance labeling scheme must use labels whose total combined length is $\Omega(n^{1+\epsilon})$, hence at least one label must be of $\Omega(n^\epsilon)$ bits. Specifically, for the class of all unweighted graphs, any distance labeling scheme must label some n -vertex graphs with labels of size $\Omega(n)$ [10].

This raises the natural question of whether more efficient labeling schemes be constructed if we abandon the ambitious goal of capturing *exact* information, and settle for obtaining *approximate* estimates. This leads to the notion of *approximate distance* labeling schemes, which are the topic of the current paper.

1.2 Labeling Schemes for Approximate Distance

Let us define the notion of approximate distance labeling schemes more precisely. Given a connected undirected graph G and two nodes u and v , let $d_G(u, v)$ denote the distance between u and v in G . A *node-labeling* for the graph G is a non-negative integer function L that assigns a label $L(u, G)$ (in the form of a binary string) to each node u of G .

A *distance decoder* is an integer function f responsible for distance computation; given two labels λ_1, λ_2 (not knowing which graph they are taken from), it returns $f(\lambda_1, \lambda_2)$. We say that the pair $\langle L, f \rangle$ is a *distance labeling* for G if $f(L(u, G), L(v, G)) = d_G(u, v)$ for any pair of nodes $u, v \in V(G)$. We say that $\langle L, f \rangle$ is an (s, r) -*approximate* distance labeling for G if

$$d_G(u, v) \leq f(L(u, G), L(v, G)) \leq s \cdot d_G(u, v) + r \tag{1}$$

for any pair of nodes $u, v \in V(G)$. More generally, $\langle L, f \rangle$ is an (s, r) -approximate *distance labeling scheme* for the graph family \mathcal{G} if it is an (s, r) -approximate distance labeling for every graph $G \in \mathcal{G}$. This paper concerns the existence of approximate distance labeling schemes which use short labels. Let $|L(u, G)|$ denote the length of the binary label $L(u, G)$ associated with u in $G \in \mathcal{G}$, and denote

$$\ell_{\langle L, f \rangle}(\mathcal{G}) = \max_{G \in \mathcal{G}} \max_{u \in V(G)} |L(u, G)| \quad \text{and} \quad \ell_{(s, r)}(\mathcal{G}) = \min_{\langle L, f \rangle} \{ \ell_{\langle L, f \rangle}(\mathcal{G}) \}$$

where the minimum is taken over all (s, r) -approximate distance labeling schemes $\langle L, f \rangle$ for \mathcal{G} .

We focus on two important special cases of (s, r) -approximate distance labeling schemes. An $(s, 0)$ -approximate distance labeling scheme is referred to as a *s-multiplicative* distance labeling scheme. Analogously, a $(1, r)$ -approximate distance labeling scheme is referred to as an *r-additive* distance labeling scheme. A 0-additive (or 1-multiplicative) distance labeling is termed an *exact* distance labeling scheme.

One may consider two variations on the distance labeling definition: the *distinct* label model and the *non-distinct* label model. The latter allows two different vertices to have the same labels, and thus lets us build schemes with sub-logarithmic label size. In the extreme case, graphs of constant diameter D enjoy a D -additive distance labeling scheme with no labels at all in the non-distinct label model: it suffices to return the estimate D to satisfy Eq. (1). In contrast, $\log n$ is a lower bound on the size of labels in the distinct label model, even in the case of an n -node clique. We remark that any distance labeling scheme in the non-distinct label model can be transformed into a scheme in the distinct label model (with the same estimate quality) by adding a unique $\log n$ bit identifier to each of the nodes. The notations $\ell_{\langle L, f \rangle}(\mathcal{G})$ and $\ell_{\langle s, r \rangle}(\mathcal{G})$ are defined for the non-distinct label model, and thus differ from the distinct label model by an additive $\log n$ factor.

We are also interested in the *query time*, i.e., the worst-case time complexity of the distance decoder. We assume a ω -bit word RAM model of computation. Each memory word can contain an integer in the range $[0, 2^\omega)$. The instruction set available consists of standard arithmetic, bitwise logical and comparison operations on words, all assumed to require constant time. For the use of bucketing and perfect hashing functions, we assume that the arithmetic operation set contains integer multiplication and integer division. Denoting the size of the input graph by n , we assume that $\omega \geq \log n$.

1.3 Related Work

An exact distance labeling scheme for weighted n -node trees with integral weights from the range $[0, W)$ using $O(\log^2 n + \log n \log W)$ bit labels has been given in [17], and $O(\log^2 n)$ bit labeling schemes for interval graphs and permutation graphs were presented in [11], all with $O(\log n)$ query time. The *Squashed Cube Conjecture* of Graham and Pollak [8], proved by Winkler [19], implies an exact distance labeling scheme for arbitrary n -node graphs with $n \log_2(3) \approx 1.58n$ bit labels for general graphs, although with a prohibitive $\Theta(n)$ query time to decode the distance. With a different approach, [10] presented a distance labeling scheme with label size $11n + o(n)$ and with $O(\log \log n)$ query time. An $8k$ -multiplicative distance labeling scheme, for each integer $k \geq 1$, is built in [17], using $O(kn^{1/k} \log n \log W)$ bit labels with query time linear in the label size, where W stands for the integral weighted diameter. For unweighted n -node graphs, this yields a $O(\log n)$ -multiplicative distance labeling scheme using $O(\log^3 n)$ bit labels with polylog query time.

Some bounds on the size of the labels are established in [10] for several families of n -node graphs. In particular, it is shown that $\ell_{\langle s, 0 \rangle}(\mathcal{G}_n) = \Omega(n)$ for every $s < 3$, where \mathcal{G}_n denotes the family of all connected graphs. For the family $\mathcal{G}_{r(n), n}$ of graphs of treewidth bounded by $r(n)$, it is shown that $\ell_{\langle s, 0 \rangle}(\mathcal{G}_{r(n), n}) = \Omega(r(n))$, for all $s < 3$ and $r(n) < n/2$. On the other hand, an exact distance labeling scheme is proposed for $\mathcal{G}_{r(n), n}$, with $O(R(n) \log n)$ bit labels and $O(R(n))$ query time, where $R(n) = \sum_{i=1}^{\log n} r(n/2^i)$. (We have $R(n) \leq r(n) \log n$, and for

monotone $r(n) \geq n^\epsilon$ with constant $\epsilon > 0$, $R(n) = O(r(n))$.) It follows, for instance, that planar graphs support an exact distance labeling scheme with $O(\sqrt{n} \log n)$ bit labels, and that trees and bounded treewidth graphs support exact distance labeling schemes with $O(\log^2 n)$ bit labels.

It is also proved in [10] that $\ell_{(1,0)}(\mathcal{B}_n) = \Omega(\sqrt{n})$ for the family \mathcal{B}_n of bounded degree graphs, and that $\ell_{(1,0)}(\mathcal{P}_n) = \Omega(n^{1/3})$ for the family \mathcal{P}_n of bounded degree planar graphs. (For the family $\mathcal{P}_{W,n}$ of weighted bounded degree planar graphs with weights in the range $[0, \sqrt{n}]$, it is proved that $\ell_{(1,0)}(\mathcal{P}_{W,n}) = \Omega(\sqrt{n})$.) Finally, for the family \mathcal{T}_n of binary trees, it is shown that $\ell_{(1,0)}(\mathcal{T}_n) \geq \frac{1}{8} \log^2 n - O(\log n)$. (For the family $\mathcal{T}_{W,n}$ of weighted binary trees with weights in the range $[0, W]$, it is proved that $\ell_{(1,0)}(\mathcal{T}_{W,n}) \geq \frac{1}{2} \log n \log W - O(\log W)$.)

1.4 Our Contribution

Section 2 deals with bounds on s -multiplicative distance labeling schemes. It is first shown that planar graphs have a 3-multiplicative distance labeling scheme with $O(n^{1/3} \log n)$ bit labels. This should be contrasted with the upper and lower bounds of [10] mentioned above for *exact* distance labeling schemes for planar graphs.

Then, we propose a $(1 + 1/\log W)$ -multiplicative distance labeling scheme using $O(\log n \cdot \log \log W)$ bit labels for the family of weighted n -node trees with weighted diameter bounded by W . More generally, we show that the family of n -node graphs of treewidth at most $r(n)$ and weighted diameter bounded by W has a $(1 + 1/\log W)$ -multiplicative distance labeling scheme using $O(R(n) \log \log W)$ bit labels and with $O(r(n))$ query time, assuming that $W = n^{O(1)}$. Hence, unweighted trees (and bounded treewidth graphs) enjoy a $(1 + 1/\log n)$ -multiplicative distance labeling scheme using $O(\log n \cdot \log \log n)$ labels with constant query time.

We then turn to lower bounds on s -multiplicative distance labeling schemes. We establish, for every $s \geq 1$, a lower bound on such schemes on general n -node graphs. Specifically, it is shown that for general graphs, $\Omega(n^{1/(3s/4 - O(1))})$ bit labels are required in the worst-case. The current upper bound is $O(sn^{1/\lceil s/8 \rceil} \log^2 n)$, derived from the result of [17] mentioned previously. Also, for trees, we show a lower bound of $\Omega(\log n \cdot \log \log n)$ bit labels for every $(1 + 1/\log n)$ -multiplicative scheme, thus proving that the scheme establishing our upper bound is optimal. For the class of weighted trees, whose weights are in the range $[0, Z]$, we show for sufficiently large Z a lower bound of $\Omega(\log n \cdot \log \log Z)$ for $(1 + 1/\log Z)$ -multiplicative schemes (which is again an optimal bound). For lack of space, the lower bounds for trees are not presented in this abstract. Full proofs can be founded in [9].

In Section 3.1, we turn to r -additive distance labeling schemes. We consider the family of k -chordal n -node graphs of diameter D , and show that they enjoy a $\lfloor k/2 \rfloor$ -additive distance labeling scheme with labels of size $O(\log n \log D)$. In particular, the family of n -node chordal graphs has an 1-additive distance labeling scheme with $O(\log^2 n)$ bit labels. Moreover, these schemes are polynomial-time

constructible, and have $O(1)$ query time. We then consider the class of graphs of diameter D which have a dominating pair. This class includes, for instance, AT-free, permutation and interval graphs. We show that this class enjoys a 2-additive scheme, and even a 1-additive scheme for interval graphs, with $\log D + O(1)$ bit labels in the non-distinct label model, and $O(\log n)$ bit labels otherwise.

We then turn to lower bounds for r -additive distance labeling schemes. We show, for every $r \geq 0$, a lower bound of $\Omega(\sqrt{n/(r+1)})$ on the required label size for r -additive distance labeling schemes over general n -node graphs. The bound is $\Theta(n)$ if $r < 2$. We also show a $\Omega(\log^2(n/(r+1)))$ lower bound on the label size for r -additive schemes on k -chordal graphs, proving that our scheme for k -chordal graphs is optimal in the label size and in the quality of the approximation. We also notice that exact distance labeling schemes for AT-free graphs or k -chordal graphs require $\Omega(n)$ bit labels.

2 Multiplicative Approximate Schemes

2.1 A Scheme for Planar Graphs

This section presents a 3-multiplicative distance labeling scheme with $O(n^{1/3} \log n)$ bit labels for the family of planar graphs. Let us start with some background concerning partitions of planar graphs. A *region* is a connected subgraph of a planar graph. One can distinguish two kinds of nodes: internal (belonging to only one region) and boundary nodes (that belong to two or more regions). The following decomposition lemma has been established in [5] using the $O(\sqrt{n})$ -separator theorem.

Lemma 1. [5] *For every n -node planar graph G and integer $k > 0$, it is possible (in polynomial time) to partition the nodes of G into k regions, each of $O(n/k)$ nodes and with $O(\sqrt{n/k})$ boundary nodes, such that any path connecting an internal node of one region to an internal node of another must go through at least one boundary node of each region.*

Setting $k = \lceil n^{1/3} \rceil$, we get $O(n^{1/3})$ regions R_1, \dots, R_k , each with $O(n^{2/3})$ nodes and $O(n^{1/3})$ boundary nodes. Each region R_i is partitioned again into $O(n^{1/3})$ subregions S_j of $O(n^{1/3})$ (internal and boundary) nodes. Each node is given a unique integer identifier $I(v) \in \{1, \dots, n\}$, as well as a pair $r(v) = (i, j)$, indicating its region number i and its subregion number j . (For boundary nodes, choose a valid pair (i, j) arbitrarily.)

Consider a node u that belongs to a region R_i and to the subregion S_j within R_i . For every other region $R_{i'}$ of G , $i' \neq i$, let $p_u(R_{i'})$ denote the closest node to u in $R_{i'}$. Note that $p_u(R_{i'})$ is necessarily a boundary node in $R_{i'}$. Similarly, for every other subregion $S_{j'}$ of R_i , $j' \neq j$, let $p_u(S_{j'})$ denote the closest node to u in $S_{j'}$.

The label $L(u, G)$ assigned to the node u consists of the following fields:

- [a] its identifier $I(u)$ and pair $r(u) = (i, j)$;
- [b] the distance from u to all the nodes in S_j ;

- [c] the identifier of $p_u(R_{i'})$ and the distance $d_G(u, p_u(R_{i'}))$, for every region $R_{i'}$ of G , $i' \neq i$;
- [d] the identifier of $p_u(S_{j'})$ and the distance $d_G(u, p_u(S_{j'}))$, for every subregion $S_{j'}$ of R_i , $j' \neq j$;
- [e] the distance from u to all the boundary nodes in its region R_i .

The number of bits in the resulting label is bounded by $O(n^{1/3} \log n)$. The distance between u and v is computed from the labels $L(u, G)$ and $L(v, G)$ as follows:

1. Extract the region and subregion numbers of u and v from field [a] of the two labels.
2. If u and v belong to the same subregion S_j in the same region R_i , then do:
 - [a] extract $d_G(u, v)$ from field [b] of $L(v, G)$ using $I(v)$;
 - [b] set $\tilde{d}(u, v) = d_G(u, v)$.
3. If u and v are in the same region R_i but in different subregions, say $u \in S_j$ and $v \in S_{j'}$, then do:
 - [a] extract $z = p_u(S_{j'})$ and $d_G(u, z)$ from field [d] of $L(u, G)$;
 - [b] extract $d_G(z, v)$ from field [b] of $L(v, G)$;
 - [c] set $\tilde{d}(u, v) = d_G(u, z) + d_G(z, v)$.
4. If u and v belong to different regions, say $u \in R_i$ and $v \in R_{i'}$, then do:
 - [a] extract $z = p_u(R_{i'})$ and $d_G(u, z)$ from field [c] of $L(u, G)$;
 - [b] extract $d_G(z, v)$ from field [e] of $L(v, G)$;
 - [c] set $\tilde{d}(u, v) = d_G(u, z) + d_G(z, v)$.
5. Return $\tilde{d}(u, v)$.

Theorem 1. *There exists a 3-multiplicative distance labeling scheme using labels of size $O(n^{1/3} \log n)$ for the family of n -node planar graphs. Moreover, the labels are polynomial time constructible and the distance decoder is $O(1)$ -time complexity.*

(All proofs are omitted and can be founded in [9].)

2.2 A Scheme for Trees and Bounded Treewidth Graphs

A pair of integer functions $\langle \lambda, \phi \rangle$ is an (s, r) -estimator of $\{1, \dots, W\}$ if $\lambda : \{1, \dots, W\} \rightarrow \{1, \dots, 2^\alpha\}$ (where typically $2^\alpha \ll W$), $\phi : \{1, \dots, 2^\alpha\} \rightarrow \mathbb{N}$, and for every $x \in \{1, \dots, W\}$, $x \leq \phi(\lambda(x)) \leq s \cdot x + r$. Intuitively, we think of λ as a function “compacting” x , and of ϕ as a function attempting to reconstruct x from $\lambda(x)$. The size of the estimator is α , and its time complexity is the worst-case time complexity of the function ϕ .

Lemma 2. *For every $k \leq O(\omega)$, and for every $m \in \{0, \dots, k\}$, there exists a constant time $(1 + 2^{-m}, 0)$ -estimator of $\{1, \dots, 2^k\}$ of size $\alpha = m + \lceil \log(k - m + 1) \rceil$.*

Given a binary string S of length k , let $\text{rank}_S(i)$ denote the function that returns the number of 1’s up to and including position i , for $i \in \{1, \dots, k\}$, and let $\text{lsb}_k(S)$ denote the position p of the least significant bit in S set to 1, $p \in \{0, \dots, k\}$

($p = 0$ if S is null). The following two results require constant time integer multiplications and divisions on $O(\log n)$ -bit words, which are allowed in our computational model.

- Lemma 3.** 1. [16,15] For every integer $k \leq n$, and for every binary string S of length k , the operation $\text{rank}_S(\cdot)$ can be performed in $O(1)$ worst-case time with $o(k)$ pre-computed auxiliary bits, and $k + o(k)$ bits in total.
 2. [6] For every integer $k \leq O(\log n)$, the operation $\text{lsb}_k(\cdot)$ can be performed in $O(1)$ worst-case time with $O(k)$ pre-computed auxiliary bits.

We state the main result for weighted trees. Given an integer $W > 0$, a W -tree is a weighted tree whose weighted diameter is bounded by W . Note that an unweighted tree T of diameter D is a particular case of a W -tree, for $W \geq D$.

It is well-known that every n -node tree has a node, hereafter called a *separator*, which splits tree into connected components each of at most $n/2$ nodes.

Let T be any n -node W -tree, and $u \in V(T)$, and let $\langle \lambda, \phi \rangle$ be an (s, r) -estimator of $\{1, \dots, W\}$. Apply the above separator property to T , and let s_1 be the obtained separator. With each connected component F of $T \setminus \{s_1\}$, we associate a unique label $c(F) \in \{1, \dots, \text{deg}(s_1)\}$ such that for every two components A, B of $T \setminus \{s_1\}$, if $|V(A)| \geq |V(B)|$ then $c(A) \leq c(B)$. Let T_1 be the connected component of $T \setminus \{s_1\}$ containing u , and let $c_1 = c(T_1)$.

We recursively apply this decomposition scheme to $T_i, i \geq 1$, in order to obtain s_{i+1}, T_{i+1} and c_{i+1} such that $u \in V(T_{i+1})$ and $c_{i+1} = c(T_{i+1})$, until we have $u = s_{i+1}$. Let h be the index such that $u = s_{h+1}$. Note that $h + 1 \leq \log n$. We set the label of u in T to be

$$L(u, T) = \langle (c_1, \dots, c_h), (\lambda(d_T(u, s_1)), \dots, \lambda(d_T(u, s_h))) \rangle .$$

Given two nodes u, u' of labels $\langle (c_1, \dots, c_h), (\lambda_1, \dots, \lambda_h) \rangle$ and $\langle (c'_1, \dots, c'_h), (\lambda'_1, \dots, \lambda'_h) \rangle$ respectively, we compute the distance as follows:

1. compute the lowest index i_0 such that $c_{i_0} \neq c'_{i_0}$;
2. return $\phi(\lambda_{i_0}) + \phi(\lambda'_{i_0})$.

Theorem 2. Let α be the size of an (s, r) -estimator of $\{1, \dots, W\}$, and let t denote its time complexity. Then the above scheme is an $(s, 2r)$ -approximate distance labeling scheme using labels of size $\alpha \log n + O(\log n)$ for the family of n -node W -trees. Moreover, the distance decoder is $O(t)$ -time complexity.

In particular, choosing $k = \lceil \log W \rceil$ and $m = \lceil \log \log W \rceil$ in Lemma 2, we have the following (noting that if $W = n^{O(1)}$, then $k = O(\omega)$ and thus all the distances can be estimated in constant time).

- Corollary 1.** 1. There exists a $(1 + 1/\log W)$ -multiplicative distance labeling scheme using labels of size $O(\log n \cdot \log \log W)$ for the family of n -node W -trees. Moreover, the distance decoder is $O(1)$ -time complexity for $W = n^{O(1)}$.
 2. There exists a $(1 + 1/\log n)$ -multiplicative distance labeling scheme using labels of size $O(\log n \cdot \log \log n)$ for the family of n -node unweighted trees. Moreover, the distance decoder is $O(1)$ -time complexity.

The scheme for trees can be applied together with the tree-decomposition scheme for bounded treewidth graphs [18] (see also Section 3.1). In [10] it is shown that graphs of treewidth bounded by $r(n)$ have exact distance labeling scheme using $O(R(n) \log n)$ bit labels, for every monotone function $r(n)$, and with $O(R(n))$ query time, where $R(n) = \sum_{i=0}^{\log n} r(n/2^i)$. Using the above scheme for trees, this result can be extended as follows.

Corollary 2. *There exists a $(1+1/\log W)$ -multiplicative (resp., exact) distance labeling scheme using labels of size $O(R(n) \log \log W)$ (resp., $O(R(n) \log W)$) for the family of weighted n -node graphs of treewidth at most $r(n)$ and of weighted diameter W . Moreover, the distance decoder is $O(r(n))$ -time complexity for $W = n^{O(1)}$.*

It follows that n -node trees and bounded treewidth graphs enjoy exact distance labeling with $O(\log^2 n)$ bit labels with constant query time (improving the query time of the scheme of [17]).

2.3 A Lower Bound for Multiplicative Schemes

To prove a lower bound on s -multiplicative distance labeling schemes on general graphs, we need the following concept introduced in [10]. For every graph family \mathcal{F} under consideration, we assume that each n -node graph of \mathcal{F} is a labeled graph on the set of nodes $V_n = \{1, \dots, n\}$. Let $A \subseteq V_n$, and let $k > 1$ be a real number (k can be a function of n). \mathcal{F} is an (A, k) -family if for every two distinct graphs $G, H \in \mathcal{F}$ there exist $x, y \in A$ such that $d_G(x, y) \geq k \cdot d_H(x, y)$ or $d_H(x, y) \geq k \cdot d_G(x, y)$. The following lemma is shown in [10] and is useful to prove Lemma 5.

Lemma 4. [10] *Let \mathcal{F} be an (A, k) -family for $k > 1$. Then for any $s < k$, $\ell_{(s,0)}(\mathcal{F}) \geq (\log |\mathcal{F}|)/|A|$.*

Lemma 5. *Let G be any connected graph with n nodes, m edges and girth g . Then the family \mathcal{S}_G composed of all the n -node connected subgraphs of G satisfies, for every real number $1 \leq s < g - 1$, $\ell_{(s,0)}(\mathcal{S}_G) \geq m/n - O(1)$.*

In particular, Lemma 5 implies that, for every $s < g - 1$, $\ell_{(s,0)}(\mathcal{G}_n) \geq m(g, n)/n - O(1)$, where $m(g, n)$ denotes the maximum number of edges in an n -node graphs of girth g . In [4] it is conjectured that $m(2k + 2, n) = \Omega(n^{1+1/k})$, proving later for $k = 1, 2, 3$ and 5. In [14], it is shown that $m(4k + 2, n) = \Omega(n^{1+1/(3k-1)})$ and that $m(4k, n) = \Omega(n^{1+1/(3(k-1))})$. Therefore,

Theorem 3. *For every $s \geq 1$, $\ell_{(s,0)}(\mathcal{G}_n) = \Omega(n^{1/(3s/4 - O(1))})$.*

3 Additive Approximate Schemes

3.1 A Scheme for k -Chordal Graphs

A graph is k -chordal if it does not contain any chordless cycles of length larger than k (where a *chord* is an edge joining two non-neighbors of the cycle). *Chordal* graphs are exactly 3-chordal graphs.

We use the notion of tree-decomposition used by Roberston and Seymour in their work on graphs minors [18]. A *tree-decomposition* of a graph G is a tree T whose nodes are subsets of $V(G)$, such that

1. $\bigcup_{X \in V(T)} X = V(G)$;
2. for every $\{u, v\} \in E(G)$, there exists $X \in V(T)$ such that $u, v \in X$; and
3. for all $X, Y, Z \in V(T)$, if Y is on the path from X to Z in T , then $X \cap Z \subseteq Y$.

For every $S \subseteq V(G)$, denote $\text{diam}_G(S) = \max_{u, v \in S} d_G(u, v)$. We denote by $G \setminus S$ the subgraph of G induced by the set of nodes $V(G) \setminus S$.

S is a *separator* of G if $G \setminus S$ is composed of two or more connected components. Moreover, S is said to be *minimal* if every proper subset of S is not a separator of G . Given $x, y \in V(G)$, S is an *x, y -separator* if x and y belongs to two distinct connected components in $G \setminus S$. We have the following.

Lemma 6. *For every k -chordal graph G , there exists a tree-decomposition T such that $|V(T)| \leq |V(G)|$, and such that, for every $X \in V(T)$, $\text{diam}_G(X) \leq k/2$. Moreover, T is polynomial-time constructible.*

Using the scheme for trees, we have the following result:

Theorem 4. *There exists a $\lfloor k/2 \rfloor$ -additive distance labeling scheme with labels of size $O(\log n \log D)$ for the family of k -chordal n -node graphs with (weighted) diameter D . Moreover, the scheme is polynomial-time constructible, and the distance decoder is $O(1)$ -time complexity.*

Corollary 3. *There exists a 1-additive distance labeling scheme with labels of size $O(\log^2 n)$ for the family of n -node chordal graphs. Moreover, the scheme is polynomial-time constructible, and the distance decoder is $O(1)$ -time complexity.*

The previous bound is optimal with respect to both the approximation ratio and the label size. Indeed, as seen later in Section 3.3, every 1-additive distance labeling scheme on the family of trees (that are chordal) requires some labels of size $\Omega(\log^2 n)$. Moreover, every exact distance labeling scheme on chordal graphs of diameter 3 requires label of size $\Omega(n)$ (cf. Theorem 6).

3.2 Dominating Pair

Our next goal is to show that in certain cases (including for instance interval, permutation, and AT-free graphs), Theorem 4 can be improved upon. A pair of nodes $\{x, y\}$ in a graph G is called a *dominating pair* if for every path P in G between x and y , and for every $u \in V(G)$, there exists a node of P at distance at most 1 of u (we say that P dominates u).

Consider any graph G having a dominating pair $\{x, y\}$, and let $\delta = d_G(x, y)$. The distance labeling scheme for graphs having a dominating pair is based on the decomposition by distance layers from node x . Let us define $L_t = \{u \in V(G) \mid d_G(x, u) = t\}$, for every $t \geq 0$. Note that for every $0 < t < \delta$, L_t is a x, y -separator. Let S_t denote a minimal x, y -separator in L_t , for every $0 < t < \delta$. We set $S_0 = \{x\}$, $S_\delta = \{y\}$ and $S_{\delta+1} = \emptyset$. An *x, y -path* is a loop-free path between x and y . We establish a number of claims useful for our analysis.

Claim. Let C_x and C_y be the connected components of $G \setminus S_t$, for $0 < t < \delta$, containing respectively x and y . For any node $z \in S_t$, there exists an x, y -path contained in $C_x \cup C_y \cup \{z\}$, composed of a shortest x, z -path and of an induced z, y -path (i.e., a chordless z, y -path) called hereafter an x, z, y -path.

Claim. (1) Let t, t' such that $0 < t \leq t' < \delta$. Let P be a x, z, y -path with $z \in S_t$. Then, every $w' \in L_{t'} \setminus S_{t'}$ has no neighbors in $P \cap L_{t'+1}$.

(2) For every $t \geq 1$, and every $w \in L_t$, w has a neighbor in S_{t-1} .

Let us now consider two distinct nodes u and v , and let $t_u = d_G(x, u)$ and let where $t_v = d_G(x, v)$. W.l.o.g., assume that $t_u \leq t_v$. The next claim gives approximations of $d_G(u, v)$ depending on the respective positions of u and v in the distance layer decomposition. It allows us to prove Theorem 5.

Claim. (1) If $t_u = 0$, then $d_G(u, v) = t_v$.

(2) If $0 < t_u = t_v \leq \delta + 1$, then $1 \leq d_G(u, v) \leq 2$.

(3) If $0 < t_u < t_v \leq \delta + 1$, then $t_v - t_u \leq d_G(u, v) \leq t_v - t_u + 2$.

Theorem 5. *Let G be a graph of diameter D with a dominating pair. Then G has a 2-additive distance labeling with labels of size $\log D + O(1)$. Moreover, the distance decoder is $O(1)$ -time complexity.*

An *asteroidal triple* of a graph G is an independent triple of nodes, each two of which are joined by a path of G that avoids the neighborhood of the third node [13]. Graphs without asteroidal triple are termed *AT-free graphs*. This class includes, in particular, the classes of interval, permutation, bounded tolerance and co-comparability graphs (see [7]). It is well-known that AT-free graphs have a dominating pair [3] that can be founded in linear time. We thus have the following.

Corollary 4. *There exists a 2-additive distance labeling scheme with labels of size $\log D + O(1)$ for the family of AT-free graphs of diameter D . Moreover, the scheme is polynomial-time constructible, and the distance decoder is $O(1)$ -time complexity.*

To get a 1-additive scheme for a subclass of AT-free, we examine in more detail the situation described by the last claim in case (3), namely, $t_u < t_v$, i.e., $d_G(x, u) < d_G(x, v)$. Indeed, thanks to the case (1) of the same claim, we already have a 1-additive scheme for $t_u = t_v$. We can show (details can be founded in [9]):

Lemma 7. *Let G be a graph that has a dominating pair $\{x, y\}$. For any two nodes u and v of G , with $d_G(x, u) \leq d_G(x, v)$, if u does not belong to a chordless cycle C_k , $k \geq 4$, then $d_G(u, v) \leq \bar{d}(u, v) \leq d_G(u, v) + 1$.*

Since interval graphs are 3-chordal AT-free graphs, the following corollary holds:

Corollary 5. *There exists a 1-additive distance labeling scheme with labels of size $\log D + O(1)$ for interval graphs of diameter D . Moreover, the scheme is polynomial-time constructible, and the distance decoder is $O(1)$ -time complexity.*

3.3 Lower Bounds for Additive Schemes

In this section we establish some lower bounds on r -additive distance labeling schemes. For general graphs we show that labels of size $\Omega(\sqrt{n/r})$ are required in the worst-case (the lower bound can be improved to $\Theta(n)$ for $r < 2$), and that every r -additive distance labeling on k -chordal graphs must use labels of length $\Omega(\log^2 n)$, for every constant $r \geq 0$. Moreover, there is no exact distance labeling scheme with labels shorter than $\Omega(n)$ bits, proving that the scheme presented in Section 3.1 is optimal (cf. Theorem 4 and Corollary 3).

Let us now show that the labeling scheme of Section 3.1 for k -chordal graphs with bounded k is optimal in terms of the length of its labels. First, let us show that there is no exact distance labeling scheme for chordal graphs using “short” labels. Consider the family \mathcal{S}_n of connected *split* graphs, namely, all the n -node graphs composed of a clique C , and of an independent set I of $n - |C|$ nodes, such that each node of I is connected to at least one node of C . Let $\mathcal{C}_{n,k}$ be the class of connected n -node k -chordal graphs. Clearly, \mathcal{S}_n is a subclass of $\mathcal{C}_{n,3}$ (because every cycle of 4 or more nodes has a chord belonging to C) and \mathcal{S}_n is connected (in fact, it contains only graphs of diameter at most three).

Recall that $\ell_{(s,r)}(\cdot)$ is related to the non-distinct label model.

Theorem 6. *For every $k \geq 3$ and $s < 2$, $\ell_{(s,0)}(\mathcal{C}_{n,k}) \geq \ell_{(s,0)}(\mathcal{S}_n) \geq n/4 - O(1)$. On the other hand, the family \mathcal{S}_n supports a 1-additive distance labeling scheme with 1 bit labels (or $\lceil \log n \rceil + 1$ bit label, if we insist on distinct labels).*

Similarly, we have the following for AT-free graphs.

Theorem 7. *Let \mathcal{A}_n be the class of connected n -node AT-free graphs. For every $s < 2$, $\ell_{(s,0)}(\mathcal{A}_n) \geq n/4 - O(1)$.*

We now give a bound of $\Omega(\sqrt{n/r})$ on the label size for general graphs. We start with a simple observation: $\ell_{(s,r)}(\mathcal{F}) \geq \ell_{(s+r,0)}(\mathcal{F})$, for all $s \geq 1$, $r \geq 0$, and graph family \mathcal{F} . Thus, for general graphs, since $\ell_{(s,0)}(\mathcal{G}_n) = \Theta(n)$ for every $s < 3$ (cf. [10]), we have $\ell_{(1,r)}(\mathcal{G}_n) = \Theta(n)$ for every $r < 2$.

We complete Theorem 3, by showing that even for $r > \lfloor k/2 \rfloor$, every r -additive scheme for k -chordal graphs requires labels of length $\Omega(\log^2 n)$. Let \mathcal{T} be the family of trees, and let $\mathcal{T}_n \subset \mathcal{T}$ denote n -node trees. Since $\mathcal{T}_n \subseteq \mathcal{C}_{n,k}$, it suffices to show that $\ell_{(1,r)}(\mathcal{T}_n) = \Omega(\log^2 n)$ for every constant r . Using subdivision of edges, one can easily show:

Corollary 6. *For every $r \geq 0$, $\ell_{(1,r)}(\mathcal{T}_n) = \Omega(\log^2(n/(r + 1)))$, and thus, for every $k \geq 3$, $\ell_{(1,r)}(\mathcal{C}_{n,k}) = \Omega(\log^2(n/(r + 1)))$.*

We also improve the lower bound on r -additive schemes on general graphs.

Corollary 7. *$\ell_{(1,r)}(\mathcal{G}_n) = \Theta(n)$ if $r < 2$, and $\ell_{(1,r)}(\mathcal{G}_n) = \Omega(\sqrt{n/r})$, if $r \geq 2$.*

Acknowledgment. We would like to thank Stéphane Pérennes for his help in developing the upper bound for trees, and Douglas West for his bibliographic support.

References

1. M. A. Breuer and J. Folkman. An unexpected result on coding the vertices of a graph. *J. of Mathematical Analysis and Applications*, 20:583–600, 1967.
2. M. A. Breuer. Coding the vertexes of a graph. *IEEE Trans. on Information Theory*, IT-12:148–153, 1966.
3. D. G. Corneil, S. Olariu, and L. Stewart. Asteroidal triple-free graphs. *SIAM Journal on Discrete Mathematics*, 10(3):399–430, Aug. 1997.
4. P. Erdős, A. Rényi, and V. T. Sós. On a problem of graph theory. In *Studia Sci. Math. Hungar.*, vol. 1, pp. 215–235, 1966.
5. G. N. Frederickson and R. Janardan. Efficient message routing in planar networks. *SIAM Journal on Computing*, 18(4):843–857, Aug. 1989.
6. M. L. Fredman and D. E. Willard. Surpassing the information theoretic bound with fusion trees. *J. of Computer and System Sciences*, 47:424–436, 1993.
7. M. C. Golumbic. *Algorithmic Graph Theory and Perfect Graphs*. Academic Press, Harcourt Brace Jovanovich, Academic Press edition, 1980.
8. R. L. Graham and H. O. Pollak. On embedding graphs in squashed cubes. *Lecture Notes in Mathematics*, 303:99–110, 1972.
9. C. Gavoille, M. Katz, N. A. Katz, C. Paul, and D. Peleg. Approximate distance labeling schemes. TR RR-1250-00, LaBRI, University of Bordeaux, 351, cours de la Libération, 33405 Talence Cedex, France, Dec. 2000.
10. C. Gavoille, D. Peleg, S. Pérennes, and R. Raz. Distance labeling in graphs. In *12th Symposium on Discrete Algorithms (SODA)*, pp. 210–219. ACM-SIAM, Jan. 2001.
11. M. Katz, N. A. Katz, and D. Peleg. Distance labeling schemes for well-separated graph classes. In *17th Annual Symposium on Theoretical Aspects of Computer Science (STACS)*, vol. 1770 of LNCS, pp. 516–528. Springer, Feb. 2000.
12. S. Kannan, M. Naor, and S. Rudich. Implicit representation of graphs. In *20th Annual ACM Symposium on Theory of Computing (STOC)*, pp. 334–343, Chicago, IL, May 1988.
13. C. G. Lekkerkerker and J. Ch. Boland. Representation of a finite graph by a set of intervals on the real line. *Fund. Math.*, 51:45–64, 1962.
14. F. Lazebnik, V. A. Ustimenko, and A. J. Woldar. A new series of dense graphs of high girth. *Bulletin of American Mathematical Society (New Series)*, 32(1):73–79, 1995.
15. J. I. Munro and V. Raman. Succinct representation of balanced parentheses, static trees and planar graphs. In *38th Symposium on Foundations of Computer Science (FOCS)*, pp. 118–126. IEEE Comp. Society Press, 1997.
16. J. I. Munro. Tables. In *16th FST&TCS*, vol. 1180 of LNCS, pp. 37–42. Springer-Verlag, 1996.
17. D. Peleg. Proximity-preserving labeling schemes and their applications. In *25th International Workshop, Graph - Theoretic Concepts in Computer Science (WG)*, vol. 1665 of LNCS, pp. 30–41. Springer, June 1999.
18. N. Robertson and P. D. Seymour. Graph minors. II. Algorithmic aspects of tree-width. *Journal of Algorithms*, 7:309–322, 1986.
19. P. Winkler. Proof of the squashed cube conjecture. *Combinatorica*, 3(1):135–139, 1983.