

# Distance Labeling in Graphs\* (Extended abstract)

Cyril Gavoille †

David Peleg †

Stéphane Pérennes §

Ran Raz ‡

## Abstract

We consider the problem of labeling the nodes of a graph in a way that will allow one to compute the distance between any two nodes directly from their labels (without using any additional information). Our main interest is in the minimal length of labels needed in different cases. We obtain upper bounds and (most importantly) lower bounds for several interesting families of graphs. In particular, our main results are the following:

1. For general graphs, the length needed is  $\Theta(n)$ .
2. For trees, the length needed is  $\Theta(\log^2 n)$ .
3. For planar graphs, we show an upper bound of  $O(\sqrt{n} \log n)$  and a lower bound of  $\Omega(n^{1/3})$ .
4. For bounded degree graphs, we show a lower bound of  $\Omega(\sqrt{n})$ .

The upper bounds for planar graphs and for trees follow by a more general upper bound for graphs with a  $\tau(n)$ -separator. The two lower bounds, however, are obtained by two different arguments that may be interesting in their own right.

We also show some lower bounds on the length of the labels, even if it is only required that distances be approximated to a multiplicative factor  $s$ . For example, we show that for general graphs the required length is  $\Omega(n)$ . We also consider the problem of the time complexity of the distance function once the labels are computed. We show that there are graphs with optimal labels of length  $3 \log n$ , such that if we use any labels with fewer than  $n$  bits per label, computing the distance function requires exponential time. A similar result is obtained for planar and bounded degree graphs.

## 1 Introduction.

### 1.1 Motivation.

Most common network representations are global in nature, and require users to have access to data on the entire network structure in order to derive useful information, even if the sought piece of information is very local, and pertains to only few nodes.

In contrast, the notion of *adjacency labeling schemes*, introduced by Breuer and Folkman [2, 3], involves using more *localized* labeling schemes for networks. The idea is to label the nodes in a way that

will allow one to infer the adjacency of two nodes *directly* from their labels, without using *any* additional information sources.

Obviously, labels of unrestricted size can be used to encode any desired information. However, for such a labeling scheme to be useful, it should strive to use relatively *short* labels (say, of length polylogarithmic in  $n$ ), and yet allow efficient (say, polylogarithmic time) information deduction. The feasibility of such *efficient* adjacency labeling schemes was explored over a decade ago by Kannan, Naor and Rudich [7].

Interest in this natural idea was recently revived by the observation that in addition to *adjacency* labeling schemes, it may be possible to devise similar schemes for 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 (say, in polylogarithmic time again) given their labels [9].

The relevance of distance labeling schemes in the context of communication networks has been pointed out in [9], and illustrated by presenting an application of such labeling schemes to distributed connection setup procedures in circuit-switched networks. It seems very plausible that distance labeling schemes may be useful also in the design of “memory-free” routing schemes, which are routing schemes geared towards supporting architectures based on very fast and simple switches, allowed to store very little data locally. Some other problems where distance labeling schemes may play an active role are bounded (“time-to-live”) broadcast protocols and topology update mechanisms.

### 1.2 Distance labeling.

Let us define the notion of distance labeling schemes more precisely. Given a 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  $L(u, G)$  to each node  $u$  of  $G$ . A *distance decoder* is a function  $f$  responsible for distance computation; given two labels  $\lambda_1, \lambda_2$  (not knowing from which graph they are taken), it returns an integer  $f(\lambda_1, \lambda_2)$ . We say that  $\langle L, f \rangle$  is a *distance labeling* for  $G$  if  $f(L(u, G), L(v, G)) = d_G(u, v)$  for any pair of nodes  $u, v \in V(G)$ . More generally,  $\langle L, f \rangle$  is a

\*This work has been supported by AFIRST.

†LaBRI, Université Bordeaux I, 351, cours de la Libération, 33405 Talence Cedex, France. gavoille@labri.u-bordeaux.fr.

‡Department of Computer Science and Applied Mathematics, The Weizmann Institute of Science, Rehovot, 76100 Israel. {peleg, ranraz}@wisdom.weizmann.ac.il.

§SLOOP I3S-CNRS/INRIA, Université de Nice-Sophia Antipolis, 2004 Route des Lucioles, BP 93, F-06902 Sophia Antipolis Cedex, France. Stephane.Perennes@sophia.inria.fr.

bounded treewidth  $\ell(\mathcal{G}_{BTW}, n) = O(\log^2 n)$ .

Our main results concern establishing some lower bounds on the size of the labels. (Some of these bounds hold even if it is only required that the distances are *approximated* to a multiplicative stretch factor  $s$ ). In particular, we prove the following:

1. For the family  $\mathcal{G}$  of general graphs, we prove  $\ell^s(\mathcal{G}, n) \geq n/2 - O(1)$  and  $\bar{\ell}^s(\mathcal{G}, n) \geq n^2/2 - O(n \log n)$ , for any  $s < 2$ .
2. For the family  $\mathcal{B}_k$  of bipartite graphs whose smaller part is of size  $k$ , we prove  $\bar{\ell}^s(\mathcal{B}_k, n) \geq k(n - k) - O(n \log n)$ , for any  $s < 3$ , and thus that  $\bar{\ell}^s(\mathcal{G}, n) \geq n^2/4 - O(n \log n)$ , for any  $s < 3$ .
3. For the family  $\mathcal{D}_3$  of graphs of maximum degree 3, we prove  $\bar{\ell}(\mathcal{D}_3, n) = \Omega(n^{3/2})$ .
4. For the family  $\mathcal{P}$  of bounded degree planar graphs, we prove  $\bar{\ell}(\mathcal{P}, n) = \Omega(n^{4/3})$ . (This answers negatively a question of [9], but leaves an intriguing gap between our upper and lower bounds.)
5. For the family  $\mathcal{T}$  of unweighted binary trees, we prove  $\ell(\mathcal{T}, n) \geq \log^2(n)/8 - O(\log n)$ . (For the family  $\mathcal{T}$  of binary trees with weights from the range  $[0, M - 1]$ , we prove  $\ell(\mathcal{T}, n) \geq \log(n) \log(M)/2 - O(\log M)$ .)

Finally, we consider the time complexity of the distance function once the labels are computed. In the full paper (see [4]) we show that there are graphs with optimal labels of size  $3 \log n$  such that, if one uses labels with fewer than  $n$  bits, it requires an exponential time to compute the distance function. A similar result is obtained for planar graphs, and bounded degree graphs.

## 2 Upper Bounds.

### 2.1 General Graphs.

Consider the family  $\mathcal{G}$  of all graphs. A trivial scheme can be based on labeling each node with its vector of distances to all other nodes, establishing  $\ell(\mathcal{G}, n) = O(n \log n)$  with  $O(1)$  time to decode the distance. Let us now show that  $\ell(\mathcal{G}, n) = O(n)$  with  $O(\log \log n)$  time to decode the distance. We start with some preliminary claims regarding dominating sets. Consider a graph  $G$  and let  $\rho$  be a positive integer. A  $\rho$ -dominating set for  $G$  is a set  $S \subseteq V(G)$  satisfying that for every node  $v \in V(G)$  there is a node  $w \in S$  at distance at most  $\rho$  from it. It is well-known (cf. [11]) that for every  $n$ -node connected graph  $G$  and integer  $\rho \geq 1$ , there exists a  $\rho$ -dominating set of cardinality at most  $\max\{1, \lfloor \frac{n}{\rho+1} \rfloor\}$ .

A collection  $\mathcal{S} = \{(S_i, \rho_i)\}_{i \in I}$ ,  $I = \{0, 1, \dots, k\}$ , such that  $S_i$  is a  $\rho_i$ -dominating set for  $G$  and  $S_k = V(G)$ , is called a *dominating collection* for  $G$ . The above

discussion implies the following trivial fact, needed for later use. (Throughout this extended abstract, some proofs are omitted or deferred to the full paper [4].)

**FACT 2.1.** *For every connected  $n$ -node graph  $G$  and  $k = \lceil \log \log n \rceil$ , there exists a dominating collection  $\mathcal{S} = \{(S_i, \rho_i)\}_{i \in I}$ ,  $I = \{0, \dots, k\}$  for  $G$ , such that  $\rho_i = 2^{k-i}$  and  $|S_i| \leq \frac{n}{2^{k-i}}$  for every  $i \in I$ .*

Let  $S$  be a  $\rho$ -dominating set for  $G$ . For every  $x \in V(G)$ , let  $\text{dom}_S(x)$  denote the *dominator* of  $x$  in  $S$ , namely, an arbitrary node  $x' \in S$  minimizing  $d_G(x, x')$ .

**LEMMA 2.1.** *For every two nodes  $x, y \in V(G)$ ,*

1.  $d_G(\text{dom}_S(x), \text{dom}_S(y)) - 2\rho \leq d_G(x, y) \leq d_G(\text{dom}_S(x), \text{dom}_S(y)) + 2\rho$ .
2. *Knowing  $d_G(\text{dom}_S(x), \text{dom}_S(y))$  and  $d_G(x, y) \bmod (4\rho)$ , one can compute  $d_G(x, y)$ .*

Our main lemma, based on a recursive construction using a dominating collection, is the following.

**LEMMA 2.2.** *There exists a distance labeling scheme  $\langle L, f \rangle$  such that for any  $n$ -node graph  $G$ , and any dominating collection  $\mathcal{S} = \{(S_i, \rho_i)\}_{i \in \{0, \dots, k\}}$ ,  $\ell_{\langle L, f \rangle}(G) \leq \sum_{i=0}^{k-1} |S_{i+1}| \log(4\rho_i) + |S_0| \log n + O(k \log n)$ . Moreover,  $f$  can be computed in time  $O(k)$  and each label can be computed in time  $O(\sum_{i=0}^k |S_i|)$ .*

**Proof.** We define a sequence of functions  $\{f_i\}_{i \in I}$  and of labelings  $\{L^i\}_{i \in I}$  such that for  $u, v \in S_i$ ,  $f_i(L^i(u, G), L^i(v, G)) = d_G(u, v)$ . The pair  $\langle L^i, f_i \rangle$  is then said to be  *$i$ -valid*. We denote by  $t(i)$  the maximum time needed to compute  $f_i$ , and let  $a(i) = \max_{u \in S_i} |L^i(u, G)|$ .

The proof is by induction. Starting with  $i = 0$  we define an ordering of the nodes of  $S_0$ . The label  $L^0(u, G)$  of a node  $u$  in  $S_0$  is made of two fields :

- [a] its rank *order*( $u$ ) in the ordering of  $S_0$ ;
- [b] the list  $\{d_G(u, v)\}_{v \in S_0}$ , given in the ordering chosen.

The distance decoder  $f_0$  is as follows: Given two labels  $L^0(x, G), L^0(y, G)$ , we use field [a] of  $L^0(y, G)$  in order to find *order*( $y$ ). Then we use field [b] of  $L^0(x, G)$ , containing the list  $\{d_G(x, v)\}_{v \in S_0}$ , select the *order*( $y$ )'th item in this list and output this result. Clearly, the pair  $\langle L^0, f_0 \rangle$  is 0-valid, Also note that  $a(0) = |S_0| \log n + \log n$ , and the operation requires constant time<sup>1</sup>, i.e.,  $t(0) = O(1)$ .

Now we proceed inductively, assuming that  $\langle L^i, f_i \rangle$  is  $i$ -valid and defining  $L^{i+1}$ . For every node  $u \in S_{i+1}$ , we compute its dominator in  $S_i$ ,  $u' = \text{dom}_{S_i}(u)$ , and we

<sup>1</sup>Constant meaning involving operations on  $\log n$  bit words on a RAM.

also choose some arbitrary *ordering* of the elements of  $S_{i+1}$ . Then we assign to  $u$  a label  $L^{i+1}(u, G)$  composed of the following fields:

- [a] the label  $L^i(u', G)$  assigned to  $u'$  ( $\leq a(i)$  bits).
- [b] the rank  $order(u)$  of  $u$  in  $S_{i+1}$  ( $O(\log n)$  bits);
- [c] the list of values  $\{d_G(u, v) \bmod 4\rho_i\}_{v \in S_{i+1}}$ , given according to the ordering chosen for  $S_{i+1}$  ( $|S_{i+1}| \log(4\rho_i)$  bits).

To compute  $d_G(x, y)$  for  $x, y \in S_{i+1}$  from the labels  $L^{i+1}(x, G), L^{i+1}(y, G)$  of  $x$  and  $y$ , we proceed as follows defining  $f_{i+1}$ :

1. For  $x' = \text{dom}_{S_i}(x)$  and  $y' = \text{dom}_{S_i}(y)$ , obtain  $L^i(x', G)$  and  $L^i(y', G)$  from field [a] of  $L^{i+1}(x, G)$  and  $L^{i+1}(y, G)$  respectively (constant time).
2. Determine  $d_G(x', y')$  by computing  $f_i(L^i(x', G), L^i(y', G))$  (time  $t(i)$ ).
3. Obtain the rank  $order(y)$  of  $y$  in  $S_{i+1}$  from field [b] of  $L^{i+1}(y, G)$  (constant time).
4. Obtain  $d_G(x, y) \bmod (4\rho_i)$ , which is the  $order(y)$ 'th entry in field [c] of  $L^{i+1}(x, G)$  (constant time as the list is sorted).
5. Compute  $d_G(x, y)$  as in Lemma 2.1, relying on the fact that  $S_i$  is a  $\rho_i$ -dominating set (constant time).

It is easy to verify that  $\langle L^{i+1}, f_{i+1} \rangle$  is  $(i+1)$ -valid. Concerning the resulting label sizes and computation times, we have  $t(i+1) \leq t(i) + O(1)$  and  $a(i+1) \leq a(i) + |S_{i+1}| \log(4\rho_i) + O(\log n)$ . As  $S_k = V(G)$ , these recurrences imply the lemma. ■

**THEOREM 2.1.** *For the class  $\mathcal{G}$  of general graphs,  $\ell(\mathcal{G}, n) \leq 9n + O(\log n \log \log n)$ . Moreover, the distance can be computed in (sublinear) time  $O(\log \log n)$  and the set of labels can be computed in time  $O(n^2)$ .*

**Proof.** The theorem is proved by first constructing a dominating collection  $\mathcal{S} = \{(S_0, \rho_0), \dots, (S_k, \rho_k)\}$ , for  $k = \lceil \log \log n \rceil$ , as in Fact 2.1, and then applying Lemma 2.2.

Let us now calculate the size of the resulting labels.

$$\ell_{\langle L, f \rangle}(G) \leq \sum_{i=0}^{k-1} |S_{i+1}| \log(4\rho_i) + |S_0| \log n + O(k \log n).$$

Recalling that  $\rho_i = 2^{k-i}$ ,  $|S_i| \leq n/2^{k-i}$  and  $k = \lceil \log \log n \rceil$ , we get that the second term is bounded by  $\frac{n}{\log n} \cdot \log n = n$ , the third term is bounded by  $O(\log n \log \log n)$ , and the first term is bounded by

$$\begin{aligned} \sum_{i=0}^{k-1} \frac{n}{2^{k-(i+1)}} \cdot \log(4 \cdot 2^{k-i}) &\leq n \sum_{i=0}^{k-1} \frac{k-i+2}{2^{k-(i+1)}} \\ &\leq n \sum_{i=0}^{\infty} \frac{i+3}{2^i} \leq 8n. \end{aligned}$$

Hence overall,  $\ell_{\langle L, f \rangle}(G) \leq 9n + O(\log n \log \log n)$ .

Considering the time complexity, in order to obtain the labeling one needs to compute the dominating collection and then some dominating sets with geometric sizes. This takes at most  $\Theta(n^2)$  steps.

The time analysis for computing  $d_G(x, y)$  from the labels  $L(x, G), L(y, G)$  using the distance decoder  $f(L(x, G), L(y, G))$ , follows directly from Lemma 2.2, and the fact that here  $k = \lceil \log \log n \rceil$ . ■

## 2.2 Distance Labeling and Separators.

It is known [7] that planar graphs support a  $4 \log n$  bit adjacency labeling scheme. In contrast, we show later on (in Section 3.3) that one cannot solve the general distance labeling problem for planar graphs using labels shorter than  $\Omega(n^{1/3})$  bits. Conversely, we now show that using the recursive  $O(\sqrt{n})$ -separator property, the problem can be solved using  $O(\sqrt{n} \log n)$  bit labels.

More generally, in this section we deal with recursive  $r(n)$ -separators. For an  $n$ -node graph  $G$ , a subset of nodes  $S$  is a *separator* if its deletion splits  $G$  into (zero or more) connected components of size at most  $\alpha n$ , for a given constant  $\alpha < 1$ . (For concreteness, we hereafter assume  $\alpha = 2/3$ .)

Given a class  $\mathcal{G}$  of graphs and an integer function  $r(n)$ , we say that  $\mathcal{G}$  has a *recursive  $r(n)$ -separator* (or simply  *$r(n)$ -separator*) if for every connected  $n$ -node graph  $G \in \mathcal{G}$  there exists a separator  $S$  of size at most  $r(n)$  such that every connected component of the graph  $G \setminus S$ , obtained from  $G$  by removing all the nodes of  $S$ , belongs to  $\mathcal{G}$  (thus in particular it has a separator of size at most  $r(\alpha n)$ ).

It is well-known that planar graphs have an  $O(\sqrt{n})$ -separator. More generally, graphs of genus  $\gamma$  have an  $O(\sqrt{\gamma n})$ -separator [5], and graphs with  $K_k$  minors excluded have an  $O(k\sqrt{n} \log n)$ -separator [12] or a  $O(\sqrt{k^3 n})$ -separator [1], and are conjectured to have an  $O(k\sqrt{n})$ -separator. Trees, series-parallel graphs, and bounded treewidth graphs, all have an  $O(1)$ -separator.

Let  $R(n) = \sum_{i \leq \log_{1.5} n} r\left(\left(\frac{2}{3}\right)^i \cdot n\right)$ . Note that for monotone  $r(n)$ ,  $R(n) \leq r(n) \log n$ , and  $R(n) = O(r(n))$  whenever  $r(n) = n^\epsilon$  for constant  $\epsilon > 0$ . The following is a generalization of the result of [9] for trees.

**THEOREM 2.2.** *For a family  $\mathcal{G}$  of graphs with a  $r(n)$ -separator,  $\ell(\mathcal{G}, n) \leq O(R(n) \log n + \log^2 n)$ . Moreover the distance can be computed in  $O(\log n)$  time.*

**COROLLARY 2.1.** 1. *For the family  $\mathcal{G}_P$  of planar graphs,  $\ell(\mathcal{G}_P, n) \leq O(\sqrt{n} \log n)$ .*

2. *For the family  $\mathcal{G}_{BTW}$  of graphs with bounded treewidth,  $\ell(\mathcal{G}_{BTW}, n) \leq O(\log^2 n)$ .*

### 3 Lower Bounds.

For any family of  $2^k$  labeled graphs on the set of nodes  $V_n = \{1, \dots, n\}$ , and any distance labeling scheme  $\langle L, f \rangle$ , some graph  $G$  must satisfy  $\bar{\ell}_{\langle L, f \rangle}(G) + O(n \log n) \geq k$ , since for every graph  $G$ , the tuple of labels  $\langle L(1, G), \dots, L(n, G) \rangle$  suffices to reconstruct all the edges of the graph, as it suffices to test for every pair of nodes if their distance is 1 or not. (Clearly, delimiting the fields of the tuple costs at most  $O(n \log n)$ .) In particular, there exists some  $n$ -node graph  $G$  with  $\ell_{\langle L, f \rangle}(G) \geq n/2 - O(\log n)$ , since there are  $2^{\binom{n}{2}}$  graphs labeled on  $V_n$ .

In this section we present lower bounds for the following graph classes: (1) general graphs with small stretched distance labeling; (2) graphs with a  $\tau(n)$ -separator and small stretched distance labeling; (3) sparse and bounded degree graphs; (4) planar graphs; (5) trees.

The first four lower bounds use the same technique, which is formalized in the next subsection.

#### 3.1 The Main Lower-Bound Theorem.

Let  $A \subseteq V_n = \{1, \dots, n\}$ , and let  $k > 1$  be a real number ( $k$  can be a function of  $n$ ). Consider a family  $\mathcal{F}$  of labeled graphs on the set of nodes  $V_n$ . Two graphs  $G, H \in \mathcal{F}$  are said to *exhibit a  $k$ -gap over  $A$*  if 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 graph family  $\mathcal{F}$  is an  $(A, k)$ -family if every two distinct graphs  $G, H \in \mathcal{F}$  exhibit a  $k$ -gap over  $A$ . The family  $\mathcal{F}$  is an  $A$ -family if there exists a real  $k > 1$  such that  $\mathcal{F}$  is a  $(A, k)$ -family. For such a family, we define

$$\begin{aligned} \bar{\ell}_{\langle L, f \rangle}(A, \mathcal{F}) &= \max\{L_{sum}(A, G) \mid G \in \mathcal{F}\}, \\ \bar{\ell}(A, \mathcal{F}) &= \min\{\bar{\ell}_{\langle L, f \rangle}(A, \mathcal{F}) \mid \langle L, f \rangle \text{ is a} \\ &\quad \text{distance labeling scheme for } \mathcal{F}\}, \end{aligned}$$

and similarly for  $s$ -stretched distance labeling schemes.

**THEOREM 3.1.** *Let  $\mathcal{F}$  be an  $(A, k)$ -family, for  $k > 1$ . Then for any stretch  $s < k$ ,*

1.  $\ell^s(\mathcal{F}) \geq \frac{1}{|A|} \cdot \log |\mathcal{F}|$ ;
2.  $\bar{\ell}^s(A, \mathcal{F}) \geq \log |\mathcal{F}| - |A| \log \log |\mathcal{F}|$ .

**Proof.** Let  $\langle L, f \rangle$  be any ( $s$ -stretched) distance labeling scheme on  $\mathcal{F}$  with  $s < k$ . Assume that  $A = \{a_1, \dots, a_\alpha\}$ . For every graph  $G \in \mathcal{F}$ , let  $L(G) = \langle L(a_1, G), \dots, L(a_\alpha, G) \rangle$ , and let  $\mathcal{L} = \{L(G) \mid G \in \mathcal{F}\}$ . First, let us show that for every two distinct  $G, H \in \mathcal{F}$ ,  $L(G) \neq L(H)$ , i.e.,  $|\mathcal{L}| = |\mathcal{F}|$ .

Assume, by way of contradiction, that  $L(G) = L(H)$  for some  $G, H \in \mathcal{F}$ , namely,  $L(a_i, G) = L(a_i, H)$  for every  $a_i \in A$ . By definition of  $\mathcal{F}$ , there exists a pair  $x, y \in A$  such that  $d_G(x, y) \geq k \cdot d_H(x, y)$  or

$d_H(x, y) \geq k \cdot d_G(x, y)$ . Without loss of generality assume the former. Hence as  $s < k$ , we have

$$(3.1) \quad d_G(x, y) > s \cdot d_H(x, y).$$

Since  $\langle L, f \rangle$  is  $s$ -stretched, we have  $d_G(x, y) \leq f(L(x, G), L(y, G))$  and  $f(L(x, H), L(y, H)) \leq s \cdot d_H(x, y)$ . However, since  $L(a_i, G) = L(a_i, H)$  for every  $a_i \in A$ , we have in particular that

$$f(L(x, G), L(y, G)) = f(L(x, H), L(y, H)).$$

Hence  $d_G(x, y) \leq s \cdot d_H(x, y)$ , contradicting Ineq. (3.1).

Now we simply evaluate the cardinality of  $\mathcal{L}$  according to a given restriction on the label length. For the first claim of the theorem, let  $\ell^s(\mathcal{F}) = l$ . Then there exists an  $s$ -stretched distance labeling scheme  $\langle L, f \rangle$  for  $\mathcal{F}$ , such that  $L_{max}^s(G) \leq l$  for every  $G \in \mathcal{F}$ . This implies that  $\mathcal{L} \subseteq [0, 2^l - 1]^\alpha$  and hence  $|\mathcal{F}| = |\mathcal{L}| \leq 2^{l\alpha}$ . The claim now follows as  $\log |\mathcal{F}| \leq l\alpha$ .

A slightly more complex argument implies the second claim as well. Suppose that there exists an  $s$ -stretched distance labeling scheme  $\langle L, f \rangle$  for  $\mathcal{F}$ , s.t.  $L_{sum}^s(A, G) \leq l$  for every  $G \in \mathcal{F}$ . This implies that

$$(3.2) \quad |\mathcal{F}| = |\mathcal{L}| \leq \binom{\alpha + l}{\alpha} \cdot 2^l \leq (\alpha + l)^\alpha \cdot 2^l.$$

If  $\alpha + l \geq \log |\mathcal{F}|$  then  $l \geq \log |\mathcal{F}| - \alpha \geq \log |\mathcal{F}| - \alpha \log \log |\mathcal{F}|$ , and the claim follows. So suppose  $\alpha + l < \log |\mathcal{F}|$ . Then Ineq. (3.2) implies that  $|\mathcal{F}| < (\log |\mathcal{F}|)^\alpha \cdot 2^l$ , hence  $\log |\mathcal{F}| < \alpha \log \log |\mathcal{F}| + l$ . ■

Let us remark that the theorem applies, in particular, to exact (nonapproximate) schemes. This requires us to interpret such a scheme over a class of  $n$ -node graphs  $\mathcal{G}_n$  as an  $s$ -stretched scheme with  $s = 1$ , and take  $k = 1 + 1/n$ .

In some cases it is possible to amplify the lower bound of the above theorem, by enlarging the graph in a suitable way. A  $(B, r)$ -graph is a graph  $T$  in which  $B = \{b_1, \dots, b_\beta\}$  is a subset of nodes, and  $r$  a particular node of  $T$ ,  $r \notin B$ . Given an  $A$ -family  $\mathcal{F}$ ,  $A = \{a_1, \dots, a_\alpha\}$ , and a  $(B, r)$ -graph  $T$ , we create for each  $G \in \mathcal{F}$  a graph  $\Psi_T(G)$  as follows. To each  $1 \leq i \leq \alpha$  we associate a distinct copy  $T_i$  of  $T$ , which is a  $(B^i, r^i)$ -graph, with  $B^i = \{b_1^i, \dots, b_\beta^i\}$ . We then set  $\Psi_T(G) = G \cup \bigcup_{i=1}^\alpha T_i$ , where the node  $r^i$  is identified with the node  $a_i$  in  $G$  for every  $1 \leq i \leq \alpha$ . Denote by  $\mathcal{F} \circ T$  the family of graphs  $\{\Psi_T(G) \mid G \in \mathcal{F}\}$ .

We note that all the graphs in  $\mathcal{F} \circ T$  share the same set of nodes,  $V_n \cup \bigcup_{i=1}^\alpha V(T_i)$ .

**LEMMA 3.1.** *Let  $\mathcal{F}$  be an  $A$ -family for  $A = \{a_1, \dots, a_\alpha\}$ , let  $T$  be a  $(B, r)$ -graph for  $B = \{b_1, \dots, b_\beta\}$ , let  $\delta = \max_j \{d_T(b_j, r)\}$ , and let  $\langle L, f \rangle$  be any distance labeling scheme on the family  $\mathcal{F} \circ T$ . Then*

$$\bar{\ell}_{\langle L, f \rangle}(\mathcal{F} \circ T) \geq \beta \cdot \bar{\ell}(A, \mathcal{F}) - \alpha \beta \lceil \log(\delta + 1) \rceil.$$

**Proof.** Partition  $Z = \bigcup_{i=1}^{\alpha} B^i$  into  $\beta$  disjoint sets  $A_t = \{b_t^1, \dots, b_t^\alpha\}$ ,  $1 \leq t \leq \beta$ , each of cardinality  $\alpha$ .

For every  $1 \leq t \leq \beta$ , define a distance labeling scheme  $\langle L^t, f^* \rangle$  on  $\mathcal{F}$  as follows. For  $G \in \mathcal{F}$ ,  $u \in V(G)$ ,

$$L^t(u, G) = \begin{cases} \langle L(u, \Psi_T(G)), 0 \rangle, & \text{if } u \notin A \\ \langle L(b_t^i, \Psi_T(G)), d_T(b_t, r) \rangle, & \text{if } u = a_i \in A \end{cases}$$

and  $f^*(\langle \lambda_1, d_1 \rangle, \langle \lambda_2, d_2 \rangle) = f(\lambda_1, \lambda_2) - (d_1 + d_2)$ .

Clearly,  $f^*$  returns the correct distance between any two nodes of  $V(G) \setminus A$  (as the fields  $d_1$  and  $d_2$  are null, and  $V(G) \subset V(\Psi_T(G))$ ). Note that each  $a_i \in A$  is a cut-vertex in  $\Psi_T(G)$ . Thus,  $d_G(a_i, u) = d_{\Psi_T(G)}(b_t^i, u) - d_{\Psi_T(G)}(b_t^i, a_i)$ , for every  $u \in V(G)$ . Moreover,  $d_{\Psi_T(G)}(b_t^i, a_i) = d_T(b_t, r)$ . So,  $f^*$  is a distance decoder for  $L^t$ . Note that  $f^*$  does not depend on  $t$ .

For every  $t$  and  $i$  and every  $G$ ,  $|L^t(a_i, G)| \leq |L(b_t^i, \Psi_T(G))| + \lceil \log(\delta + 1) \rceil$ , because the second field of  $L^t$  labels has  $\delta + 1$  possible values (namely, the code 0, and  $d_T(b_t, r) \in [1, \delta]$ ).

Define a distance labeling scheme  $\langle L^*, f^* \rangle$  on  $\mathcal{F}$  as follows. For every  $G \in \mathcal{F}$ , we have  $L_{sum}^t(A, G) = \sum_{i=1}^{\alpha} |L^t(a_i, G)|$ . Define  $t(G)$  to be the index  $t$  that minimize  $L_{sum}^t(A, G)$ . Now define  $L^*(u, G)$  for every  $G \in \mathcal{F}$  and  $u \in V(G)$  by  $L^*(u, G) = L^{t(G)}(u, G)$ . Since  $f^*$  is a distance decoder for every  $L^t$ , we conclude that  $f^*$  is a distance decoder for  $L^*$ .

Since  $\langle L^*, f^* \rangle$  is a distance labeling scheme on  $\mathcal{F}$  there exists  $G_0 \in \mathcal{F}$  such that  $L_{sum}^*(A, G_0) \geq \bar{\ell}(A, \mathcal{F})$  and hence for every  $1 \leq t \leq \beta$ , we have  $L_{sum}^t(A, G_0) \geq \bar{\ell}(A, \mathcal{F})$ . Denote  $H_0 = \Psi_T(G_0)$ . It follows that for every  $1 \leq t \leq \beta$ ,  $\sum_{i=1}^{\alpha} (|L(b_t^i, H_0)| + \lceil \log(\delta + 1) \rceil) \geq \bar{\ell}(A, \mathcal{F})$ . Since  $\sum_{i=1}^{\alpha} |L(b_t^i, H_0)| = L_{sum}(A_t, H_0)$ , the above inequality can be rewritten as  $L_{sum}(A_t, H_0) \geq \bar{\ell}(A, \mathcal{F}) - \alpha \lceil \log(\delta + 1) \rceil$ . Since the sets  $A_t$  are pairwise disjoint,  $L_{sum}(Z, H_0) = \sum_{t=1}^{\beta} L_{sum}(A_t, H_0) \geq \beta \cdot \bar{\ell}(A, \mathcal{F}) - \alpha \beta \lceil \log(\delta + 1) \rceil$ . We complete the proof by noting that  $\bar{\ell}_{(L, f)}(\mathcal{F} \circ T) \geq L_{sum}(Z, H_0)$ . ■

Hence we can hope to amplify the lower bound of Theorem 3.1 by a multiplicative factor of  $\beta = |B|$ . The family  $\mathcal{F} \circ T$  contains graphs of size larger than  $n$  but smaller than  $n + \alpha|V(T)|$ . This remains  $O(n)$  for  $(B, r)$ -graphs  $T$  with  $|V(T)| = O(n/\alpha)$  and then we can also have  $|B| = O(n/\alpha)$ . Thus, for suitable families  $\mathcal{F}$ ,  $\mathcal{G}$ , and a graph  $T$  so that  $\mathcal{F} \subseteq \mathcal{G}$ , and  $\mathcal{F} \circ T \subseteq \mathcal{G}$ , we can have

$$\bar{\ell}(\mathcal{G}, n) \geq \Omega(n \cdot \ell(\mathcal{G}, n)) - O(n \log n).$$

Lemma 3.1 is used to prove Theorem 3.4 and Theorem 3.5.

### 3.2 Applications of the Lower Bound.

Our first application of the main lower-bound theorem is the following.

**THEOREM 3.2.** *Let  $\mathcal{G}$  be the family of general graphs, and let  $s < 2$ . Then  $\ell^s(\mathcal{G}, n) \geq (n-1)/2 - o(1)$  and  $\bar{\ell}^s(\mathcal{G}, n) \geq n^2/2 - O(n \log n)$ .*

**Proof.** Let  $\mathcal{F}$  be the family of all labeled graphs of diameter 2 on  $V_n = \{1, \dots, n\}$ .  $\mathcal{F}$  is a  $(V_n, 2)$ -family, because for any two distinct graphs  $G, H$  of  $\mathcal{F}$  there always exists a pair  $(x, y)$  of  $V_n$  for which either  $d_G(x, y) = 1$  and  $d_H(x, y) = 2$ , or  $d_G(x, y) = 2$  and  $d_H(x, y) = 1$ .

To apply the main lower bound theorem we need to estimate  $|\mathcal{F}|$ . Let  $\mathcal{G}$  be the set of all (connected or disconnected) graphs on  $V_n$ . Clearly  $\mathcal{F} \subset \mathcal{G}$  and  $|\mathcal{G}| = 2^{\binom{n}{2}}$ . Let us bound the probability that a graph  $G$  taken uniformly at random from  $\mathcal{G}$  is in  $\mathcal{F}$ . One possible way for taking a graph  $G$  randomly and uniformly from  $\mathcal{G}$  consists of setting all the possible edges with probability  $p = 1/2$ . Note that  $G \notin \mathcal{F}$  if and only if there is a pair  $x, y \in V_n$  such that  $x$  and  $y$  are not adjacent, and such that there is no  $z \in V_n \setminus \{x, y\}$  adjacent to both  $x$  and  $y$ . This occurs, for a given pair  $\{x, y\}$ , with probability  $p(1-p^2)^{n-2} = 1/2 \cdot (3/4)^{n-2}$ . Hence, it occurs for at least one pair with probability of at most  $\binom{n}{2} \cdot 1/2 \cdot (3/4)^{n-2} < 1/2$  for every sufficiently large  $n$ . Therefore,  $\log |\mathcal{F}| \geq n(n-1)/2 - 1$ . Both claims of the theorem now follow by Theorem 3.1 (noting, for the second claim, that also  $|\mathcal{F}| \leq 2^{n^2}$  and hence  $\log \log |\mathcal{F}| \leq 2 \log n$ ). ■

We next consider graphs with  $r(n)$ -separator. Let  $\mathcal{B}_k$  denote the set of bipartite graphs whose smaller part is of size  $k$ . Clearly,  $\mathcal{B}_k$  has a  $k$ -separator. Denote by  $\mathcal{B}_{k,n}$  the subcollection of graphs of  $\mathcal{B}_k$  on the set of nodes  $V_n$ . We next bound the total label length required by distance labeling schemes for  $\mathcal{B}_{k,n}$ .

For every bipartite graph in  $\mathcal{B}_{k,n}$ , let  $X$  and  $Y$  denote the two parts of nodes, with  $|X| = k$ , and  $|Y| = n - k$ . Consider the subset of graphs  $\mathcal{F} \subset \mathcal{B}_{k,n}$  whose diameter is bounded by 3. Note that  $\mathcal{F}$  is a  $(V_n, 3)$ -family, because for every two distinct  $G, H \in \mathcal{F}$ , there exists a pair  $(x, y) \in X \times Y$  such that  $d_G(x, y) = 1$  and  $d_H(x, y) \neq 1$  (or the reverse). Since  $H$  is of diameter 3, the fact that there is no edge between  $x$  and  $y$  necessitates  $d_H(x, y) = 3$ .

**LEMMA 3.2.** *For sufficiently large  $n$  and for  $2 \log n \leq k \leq n/2$ ,  $|\mathcal{F}| \geq 2^{k(n-k)-1}$ .*

**THEOREM 3.3.** *Let  $s < 3$ . Then for every sufficiently large  $n$  and for each  $k$  such that  $2 \log n \leq k \leq n/2$ ,*

$$\bar{\ell}^s(\mathcal{B}_{k,n}) \geq k(n-k) - 2n \log n.$$

Since  $\mathcal{B}_k$  has a  $k$ -separator, we have that in general, graphs with an  $n^\epsilon$ -separator for constant  $\epsilon \leq 1$  have

*distance labeling scheme* for the graph family  $\mathcal{G}$  if it is a distance labeling for every graph  $G \in \mathcal{G}$ .

It is important to note that the function  $f$ , responsible of the distance computation, is independent of  $G$  or of its cardinality  $|V(G)|$ . Thus  $f$  can be seen as a method used to compute the distances in a distributed fashion, given any two labels and knowing that the graph belongs to some specific family. In particular, it must be possible to define  $f$  by a constant size algorithm. In contrast, the labels contain some information that can be pre-computed by considering the whole graph structure.

Clearly, a distance labeling scheme always exists for any graph family if one allows arbitrarily large labels. In this paper we are interested in the existence of distance labeling schemes which use short labels. Let  $|L(u, G)|$  denote the length of the binary label  $L(u, G)$  associated with  $u$ , and denote  $L_{\max}(G) = \max_{u \in V(G)} |L(u, G)|$ . Given a finite graph family  $\mathcal{G}$  and a distance labeling scheme  $\langle L, f \rangle$ , denote

$$\begin{aligned} \ell_{\langle L, f \rangle}(\mathcal{G}) &= \max\{L_{\max}(G) \mid G \in \mathcal{G}\}, \\ \ell(\mathcal{G}) &= \min\{\ell_{\langle L, f \rangle}(\mathcal{G}) \mid \langle L, f \rangle \text{ is a} \\ &\quad \text{distance labeling scheme for } \mathcal{G}\}. \end{aligned}$$

For an arbitrary graph family  $\mathcal{G}$ , let  $\mathcal{G}_n$  denote the subfamily containing the  $n$ -node graphs of  $\mathcal{G}$ , and define  $\ell_{\langle L, f \rangle}(\mathcal{G}, n) = \ell_{\langle L, f \rangle}(\mathcal{G}_n)$  and  $\ell(\mathcal{G}, n) = \ell(\mathcal{G}_n)$ .

Instead of considering the maximal label length one can prefer the *total* label length. For every node subset  $W \subseteq V(G)$ , let  $L_{\text{sum}}(W, G) = \sum_{u \in W} |L(u, G)|$ , and  $L_{\text{sum}}(G) = L_{\text{sum}}(V(G), G)$ . Given a distance labeling scheme  $\langle L, f \rangle$ , denote

$$\begin{aligned} \bar{\ell}_{\langle L, f \rangle}(\mathcal{G}) &= \max\{L_{\text{sum}}(G) \mid G \in \mathcal{G}\}, \\ \bar{\ell}(\mathcal{G}) &= \min\{\bar{\ell}_{\langle L, f \rangle}(\mathcal{G}) \mid \langle L, f \rangle \text{ is a} \\ &\quad \text{distance labeling scheme for } \mathcal{G}\}. \end{aligned}$$

and let  $\bar{\ell}_{\langle L, f \rangle}(\mathcal{G}, n) = \bar{\ell}_{\langle L, f \rangle}(\mathcal{G}_n)$  and  $\bar{\ell}(\mathcal{G}, n) = \bar{\ell}(\mathcal{G}_n)$ . We are also interested in the efficiency of the distance computation. In a *linear* distance labeling, the worst-case time complexity is proportional to the size of the inputs, i.e., to the length of the longest label.

Distance labelings can also be defined up to multiplicative stretch factor  $s$ . That is, given a distance decoder  $f$ , a node-labeling  $L$  and a real  $s \geq 1$ , we say that  $\langle L, f \rangle$  is an *s-stretched distance labeling* for  $G$  if for any pair of nodes  $u, v$  of  $G$ ,

$$d_G(u, v) \leq f(L(u, G), L(v, G)) \leq s \cdot d_G(u, v).$$

All the above parameters are extended to this case by adding a superscript  $s$ .

The above definitions are for the general case of weighted graphs. Below, we will work mainly with classes of unweighted graphs (unless said otherwise).

### 1.3 Related work.

Many online problems on static graph collections can be solved efficiently using preprocessing and auxiliary space. However, here we insist on more localized processing, namely, answering online queries with local information (or labels) associated to the nodes involved in the query alone. *Adjacency* labeling schemes are studied in [7]. Specifically, it is shown how to construct  $O(\log n)$ -bit adjacency labeling schemes for a number of graph families, including trees, bounded arboricity graphs (including, in particular, graphs of bounded degree and graphs of bounded genus, e.g., planar graphs), various intersection-based graphs such as interval graphs, and  $c$ -decomposable graphs. It is also easy to encode the *ancestry* (or *descendance*) relation in a tree using interval-based schemes (cf. [13]).

Concerning *distance* query on general  $n$ -node graphs, Graham and Pollak proposed to label each node by a word of  $q_n$  symbols taken in  $\{0, 1, *\}$  such that the distance between two nodes corresponds to the Hamming distance of the two words (the distance between  $*$  and any symbol is null) [6]. Referenced as the *Squashed Cube Conjecture*, Winkler has proved that  $q_n \leq n - 1$  for every  $n$ , implying a scheme with labels of  $n \log_2(3) \approx 1.58n$  bits, although with a prohibitive  $\Theta(n)$  query time to decode the distance [14].

More recently, a distance labeling scheme for weighted trees with weights from the range  $[0, M - 1]$  using  $O(\log^2 n + \log n \log M)$  bit labels has been given in [9], and  $O(\log^2 n)$  distance labeling schemes for interval graphs and permutation graphs were presented in [8], all with  $O(\log n)$  query time. Queries concerning the least-common ancestor of two nodes, and related functions, can be answered with labels of length  $O(\log^2 n)$  bits with  $O(\log n)$  query time [10].

### 1.4 Our contribution.

We first present some upper bounds. For the class  $\mathcal{G}$  of all graphs, Winkler showed in [14] that  $\ell(\mathcal{G}, n) \leq 1.58n$ , however with a  $\Theta(n)$  time to decode the distance. We show that  $n$ -node graphs can be labeled with labels of size  $9n$  bits so that in time  $O(\log \log n)$  the distance between two nodes can be computed given their labels only. This result is complemented by the fact that the class  $\mathcal{G}$  of all  $n$ -node graphs requires labels of size  $\Omega(n)$ . Hence  $\ell(\mathcal{G}, n) = \Theta(n)$ .

We also show that classes of graphs with (recursive)  $r(n)$ -separators support distance labeling scheme with labels of size  $O(r(n) \log n + \log^2 n)$  such that the distance can be computed in time  $O(\log n)$ . This general upper bound implies several results. For instance, it implies that for the family  $\mathcal{G}_P$  of planar graphs  $\ell(\mathcal{G}_P, n) = O(\sqrt{n} \log n)$ , and for the family  $\mathcal{G}_{BTW}$  of graphs with

distance labeling with labels of size  $\Omega(n^\epsilon)$ . The extremal case  $k = n/2$  yields an alternative proof for the  $\Omega(n)$  lower bound for general graphs, which in fact holds for larger stretch values, albeit with a slightly weaker constant in the leading term.

**COROLLARY 3.1.** *Let  $\mathcal{G}$  be the family of general graphs, and let  $s < 3$ . Then for every sufficiently large  $n$ ,  $\bar{\ell}^s(\mathcal{G}, n) \geq n^2/4 - 2n \log n$ .*

Our next question is whether there exists a distance labeling scheme with short labels, say of length  $O(n^\epsilon)$  for constant  $\epsilon < 1$ , for the class of  $n$ -node graphs with  $O(n)$  edges. The following theorem answers this question negatively for every  $\epsilon < 1/2$ . Let  $\mathcal{D}_3$  be the class of graphs of maximum degree three.

**THEOREM 3.4.** *For every sufficiently large  $n$ ,  $\bar{\ell}(\mathcal{D}_3, n) = \Omega(n^{3/2})$ .*

**Proof.** Let  $X = \{x_1, \dots, x_k\}$  and  $Y = \{y_1, \dots, y_m\}$  where  $k + m = n$ . For an integer  $t \geq 1$ , let  $T(t)$  be a complete binary tree of height  $\lceil \log t \rceil$ . We assume that the first  $m$  leaves of the tree  $T(m)$  are numbered 1 through  $m$ . We construct a family  $\mathcal{H}$  of graphs defined as follows. With each  $x_i \in X$  we associate a copy  $T_{x_i}$  of  $T(m)$ . Similarly, with each  $y_j \in Y$  we associate a copy  $T_{y_j}$  of  $T(k)$ . The union of these  $n$  trees forms the set of nodes and a part of the edge-set of all the graphs of  $\mathcal{H}$ . In addition, for every  $(x_i, y_j) \in X \times Y$ , a graph  $H \in \mathcal{H}$  may or may not contain a *cross edge*  $e_{i,j}$  connecting the  $j$ th leaf of  $T_{x_i}$  with the  $i$ th leaf of  $T_{y_j}$ . Thus the class  $\mathcal{H}$  consists of the  $2^{km}$  graphs generated by considering all possible such choices. Alternatively,  $\mathcal{H}$  can be viewed as the class of all bipartite graphs with parts  $X$  and  $Y$ , in which every node  $x_i \in X$  (respectively,  $y_j \in Y$ ) is replaced by a complete binary tree of height  $\lceil \log m \rceil$  (resp.,  $\lceil \log k \rceil$ ). (See Fig. 1.)

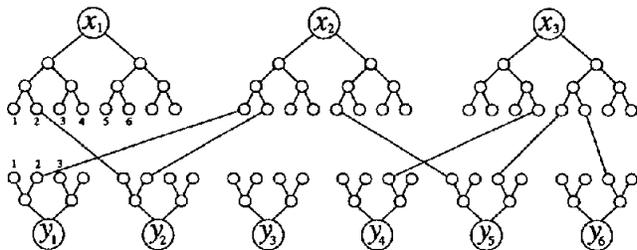


Figure 1: A possible graph  $H$  for  $k = 3$  and  $m = 6$ .

The number of leaves in a complete binary tree of height  $h$  is  $2^h$ , and the number of nodes is  $2^{h+1} - 1$ . So, the number of nodes in each graph of  $\mathcal{H}$  is

$$\begin{aligned} N &= k \cdot 2^{\lceil \log m \rceil + 1} - k + m \cdot 2^{\lceil \log k \rceil + 1} - m \\ &< 4km + 4mk - n < 8km . \end{aligned}$$

Moreover, the maximum degree is three (as in particular, there is at most one cross edge touching any leaf). Let  $V_N = \{1, \dots, N\}$  and  $A = X \cup Y$ . For every  $a \in A$ , the root of  $T_a$  is labeled  $a$ . By the above definition, every two graphs  $G, H \in \mathcal{H}$  differ on some cross edge  $e_{i,j}$ , and subsequently, exhibit a gap, as  $d_G(x_i, y_j) \neq d_H(x_i, y_j)$ . Hence  $\mathcal{H}$  is an  $A$ -family. The number of graphs in  $\mathcal{H}$  equals  $2^{km}$ , the number of bipartite graphs defined on  $X \cup Y$ , and can be bounded from above by  $|\mathcal{H}| < 2^{n^2}$ . Hence by part 2 of Theorem 3.1, any distance labeling scheme  $(L, f)$  on  $\mathcal{H}$  requires  $\bar{\ell}(A, \mathcal{H}) \geq km - 2n \log n$ .

Consider now a complete binary tree  $T$  of height  $h_T$  with a node  $r$  of degree one attached to its root. We choose  $h_T$  such that its set of leaves,  $B$ , has cardinality at least  $km/n$  and less than  $2km/n$ .  $T$  is a  $(B, r)$ -graph, and  $\mathcal{H} \circ T \subset \mathcal{D}_3$ . Thus, by Lemma 3.1, the family  $\mathcal{H} \circ T$  satisfies  $\bar{\ell}(\mathcal{H} \circ T) \geq |B| \cdot \bar{\ell}(A, \mathcal{H}) - |A| \cdot |B| \cdot \lceil \log(h_T + 2) \rceil$ .

Let  $W$  denote the set of nodes for every graph in  $\mathcal{H} \circ T$ . Then  $|W| \leq |V_N| + |A| \cdot |V(T)| < 12km$ , because each copy of tree  $T$  has at most  $2km/n$  leaves, thus at most  $4km/n$  nodes. Moreover,  $|B| < 2km/n$  and  $\lceil \log(h_T + 2) \rceil < 2 \log \log(km/n)$ . It follows that

$$\begin{aligned} \bar{\ell}(\mathcal{H} \circ T) &\geq |B| (km - 2n \log n - 2n \log \log(km/n)) \\ &\geq \frac{(km)^2}{n} - 2km (\log n + \log \log(km/n)) . \end{aligned}$$

Choosing  $k = n/2$ , we obtain  $\bar{\ell}(\mathcal{H} \circ T) \geq n^3/16 - n^2 \log n$  and  $|W| < 3n^2$ , i.e.,  $\bar{\ell}(\mathcal{H} \circ T) = \Omega(|W|^{3/2})$ . ■

### 3.3 A Lower Bound for Planar Graphs.

In this subsection we provide a lower bound for planar graphs. Note that a graph with a  $O(\sqrt{n})$ -separator is not necessarily planar. In particular, almost all the subgraphs of the complete bipartite graph  $K_{\sqrt{n}, n-\sqrt{n}}$  are not planar (because they contain  $K_{3,3}$ ), and yet they have a  $\sqrt{n}$ -separator. So the lower bound of Theorem 3.3 cannot be applied.

**THEOREM 3.5.** *There exists a graph family  $\mathcal{P}$  consisting of bounded degree planar graphs, such that for sufficiently large  $n$ ,  $\bar{\ell}(\mathcal{P}, n) = \Omega(n^{4/3})$ .*

**Proof.** We first construct a class  $\mathcal{G}$  of planar  $n$ -node graphs of bounded degree, which is an  $S$ -family for a node set  $S$  of size  $|S| = O(n^{1/3})$ , and such that  $\log |\mathcal{G}| = \Omega(n^{2/3})$ . Since the size of any family of  $n$ -node bounded-degree graphs is at most  $2^{O(n \log n)}$ , it follows by part 2 of Theorem 3.1 that every distance labeling scheme on  $\mathcal{G}$  requires  $\bar{\ell}(S, \mathcal{G}) \geq \Omega(n^{2/3})$ . Then it remains to consider the family  $\mathcal{P} = \mathcal{G} \circ T$ , where  $T$  is a complete binary tree with  $\Theta(n^{2/3})$  leaves.  $\mathcal{P}$  is

composed of planar  $O(n)$ -node bounded degree graphs, and by Lemma 3.1 every distance labeling on  $\mathcal{P}$  requires  $\bar{\ell}(\mathcal{G} \circ T) = \Omega(n^{4/3})$ .

**Description of an  $S$ -family  $\mathcal{G}$ .** Consider the upper-left half of a grid of  $k$  columns and  $k$  rows (see Fig. 2). The node with coordinates  $(i, j)$ , i.e., residing on the  $i$ th column and  $j$ th row of the grid, is named  $z_{i,j}$ . The set of nodes we consider in the grid is  $Z = \{z_{i,j} \mid 2 \leq i + j \leq k + 1\}$  (drawn in gray in Fig. 2). At every node  $z_{i,1}$ , for  $1 \leq i \leq k$ , we attach a node  $u_i$  of degree one, and at every node  $z_{k+1-j,j}$ , for  $1 \leq j \leq k$ , we attach a node  $v_j$  of degree one. To lighten notations  $u_i$  is also named  $z_{i,0}$  and  $v_j$  named  $z_{k+2-j,j}$ . For every  $z_{i,j} \in Z$ , the edge  $(z_{i,j}, z_{i,j-1})$  is subdivided into two edges  $(z_{i,j}, x_{i,j})$  and  $(x_{i,j}, z_{i,j-1})$ , adding the node  $x_{i,j}$ . Moreover the edge  $(z_{i,j}, z_{i+1,j})$  is subdivided into the edges  $(z_{i,j}, y_{i,j})$  and  $(y_{i,j}, z_{i+1,j})$ , adding the node  $y_{i,j}$ . Finally we add the edge  $e_{i,j} = (x_{i,j}, y_{i,j})$  for all  $i, j$ . For simplification, we consider the graph to be weighted, and assign an integer weight  $w(e) \geq 1$  to each edge  $e$ . Specifically, let  $w(e) = 1$  for every edge  $e$ , except for the edges  $(x_{i,j}, z_{i,j-1})$  which are assigned the weight  $2i - 1$ , and the edges  $(y_{i,j}, z_{i+1,j})$  which are assigned the weight  $2j - 1$ , for all  $i, j$  such that  $2 \leq i + j \leq k + 1$ . The resulting labeled graph is denoted by  $G_k$ . It is planar and of degree bounded by 4. It is depicted on Fig. 2 with  $k = 6$ .

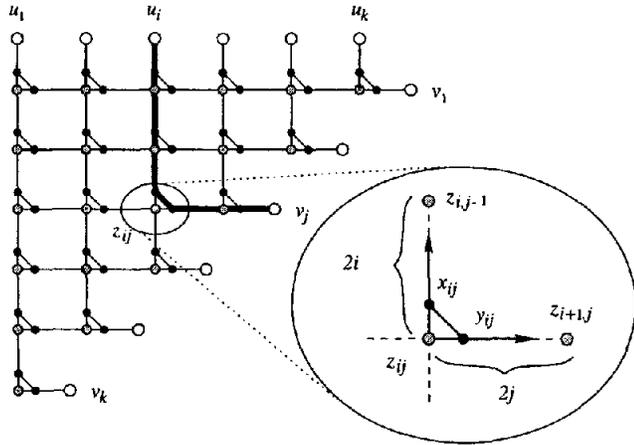


Figure 2: The graph  $G_k$  defining  $\mathcal{G}$ .

It should be clear that the graph  $G_k$  can be transformed back into an unweighted graph, by replacing each edge  $e$  of weight  $w(e)$  with a simple path of  $w(e)$  edges. Since an edge with weight  $w$  contributes  $w - 1$  new nodes, the total number of nodes in the unweighted version of  $G_k$  is

$$n = \sum_{2 \leq i+j \leq k+1} (2i + 2j + O(1)) + O(k^2)$$

$$= \frac{2}{3} \cdot k^3 + O(k^2).$$

For convenience, we henceforth discuss the graph in its weighted form.

Let  $S = \{u_1, \dots, u_k, v_1, \dots, v_k\}$ . The family  $\mathcal{G}$  is composed of all graphs  $G_k$  in which we decide to remove or not each edge  $e_{i,j}$ . The number of edges  $e_{i,j}$  in  $G_k$  is  $|Z| = k(k+1)/2$ , thus  $|\mathcal{G}| = 2^{k(k+1)/2}$ . We need to show that  $\mathcal{G}$  is an  $S$ -family. Towards proving this, we establish the following two lemmas.

**LEMMA 3.3.** *Any shortest path in  $G_k$  from  $u_i$  to any of the nodes  $\{x_{i,j}, y_{i,j}, z_{i,j}\}$ , for every  $j$ , must go through the nodes of the  $i$ th column only.*

The next lemma states that the shortest path in  $G_k$  from  $u_i$  to  $v_j$  is precisely the one highlighted in Fig. 2.

**LEMMA 3.4.** *For any  $i, j$ ,  $2 \leq i + j \leq k + 1$ , every shortest path in  $G_k$  from  $u_i$  to  $v_j$  goes through the sequence of nodes  $x_{i,1}, z_{i,1}, x_{i,2}, z_{i,2}, \dots, z_{i,j-1}, x_{i,j}, y_{i,j}, z_{i+1,j}, y_{i+1,j}, \dots, z_{k+1-i,j}, y_{k+1-i,j}$ .*

It follows from Lemma 3.4 that any shortest path from  $u_i$  to  $v_j$  must use the edge  $e_{i,j}$ , so removing this edge from the graph increases the distance by at least 1. Moreover, this shortest path does not go through any other edge  $e_{i',j'}$ , showing that  $d_G(u_i, v_j)$  depends only on whether  $e_{i,j}$  exists or not. So, given two graphs  $G, H \in \mathcal{G}$  that differs by the edge  $e_{i,j}$  we have  $d_G(u_i, v_j) \neq d_H(u_i, v_j)$ .

**Application of the Lower-Bound Theorem.** Because  $\mathcal{G}$  is an  $S$ -family, we have by part 3 of Theorem 3.1 that every distance labeling scheme on  $\mathcal{G}$  requires  $\bar{\ell}(S, \mathcal{G}) \geq k(k+1)/2 - O(|S| \log k)$ . We have  $|S| = 2k$ ,  $n = (2/3) \cdot k^3 + O(k^2)$ , thus  $k > n^{1/3} - O(n^{2/9})$ , and finally  $L_{\text{sum}}(S, \mathcal{G}) \geq n^{2/3}/2 - O(n^{4/9})$ , completing the proof.  $\blacksquare$

We also trivially have a  $\sqrt{n}$ -lower bound for non-uniform weighted planar graphs.

**COROLLARY 3.2.** *There exists a graph family  $\mathcal{WP}$  consisting of bounded degree weighted planar graphs, such that for every distance labeling scheme  $\langle L, f \rangle$  for  $\mathcal{WP}$ , and for sufficiently large  $n$ , there exists an  $n$ -node graph  $G$  of  $\mathcal{WP}$  on which  $\langle L, f \rangle$  requires  $L_{\text{sum}}(G) = \Omega(n^{3/2})$ . Moreover, the weights are non negative integers that do not exceed  $O(\sqrt{n})$ .*

### 3.4 A Lower Bound on Trees.

When applying the general approach for trees, considering the set  $\mathcal{F}$  of all labeled trees on the set  $V_n = \{1, \dots, n\}$  as a  $(V_n, 1)$ -family, one gets  $|\mathcal{F}| = n^{n-2}$  (known as Cayley's formula). Unfortunately, this

implies only the trivial  $\log n$  lower bound on the average or maximum label length.

In this section we prove a stronger lower bound, namely, that for the family  $\mathcal{T}$  of weighted trees with weights from the range  $[0, M - 1]$ , any distance labeling scheme requires  $\ell(\mathcal{T}, n) = \Omega((\log M + \log n) \log n)$ . This bound is tight given the  $O((\log M + \log n) \log n)$  distance labeling scheme given for this class in [9]. Note that for unweighted trees we obtain a lower bound of  $\Omega(\log^2 n)$ .

For the lower bound proof we focus on a special class of binary weighted trees called  $(h, M)$ -trees, defined as follows. For  $h = 1$ , a  $(1, M)$ -tree  $T$  is composed of a root with a single child and two grandchildren. An integral weight  $x \in [0, M - 1]$  is associated with each of the two edges connecting the child to the two grandchildren, and the weight  $M - x$  is associated with the edge connecting the root to the child.

For  $h \geq 2$ , a  $(h, M)$ -tree is constructed by taking a  $(1, M)$ -tree and attaching to each of its two leaves an  $(h - 1, M)$ -tree. Hence an  $(h, M)$ -tree contains  $2^h$  leaves, denoted  $a_1, \dots, a_{2^h}$ . Let  $\mathcal{C}(h, M)$  denote the class of all  $(h, M)$ -trees. Note that all of those trees have the same structure, and they differ only in their weight assignment. Fig. 3 depicts a  $(3, M)$ -tree.

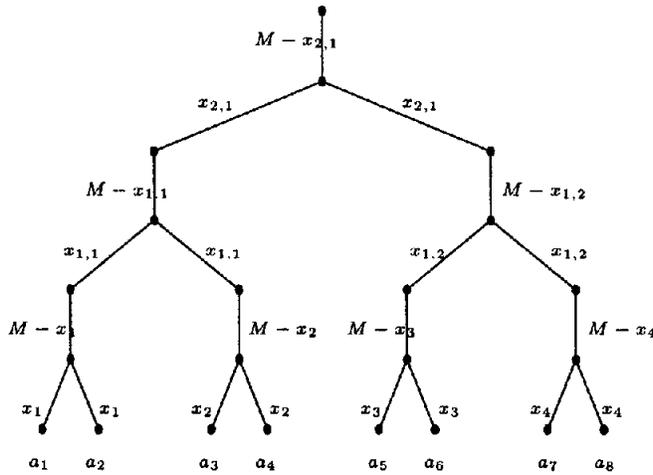


Figure 3: A  $(3, M)$ -tree.

Note that a  $(h, M)$ -tree  $T$  is completely defined by the triple  $T = (T_0, T_1, x)$ , where  $x$  is the weight associated with the two edges of the top  $(1, M)$ -tree, and  $T_0$  and  $T_1$  are the two  $(h - 1, M)$ -trees attached to the leaves of the top tree. The subclass of  $\mathcal{C}(h, M)$  consisting of  $(h, M)$ -trees with topmost weight  $x$  is denoted  $\mathcal{C}(h, M, x)$ . Hence  $\mathcal{C}(h, M) = \bigcup_{x=0}^{M-1} \mathcal{C}(h, M, x)$ .

By the definition of these binary trees we have

LEMMA 3.5. For every two leaves  $a, a'$  of a tree  $T \in \mathcal{C}(h, M, x)$ , (1) If  $a \in T_0$  and  $a' \in T_1$  then

$d_T(a, a') = 2(h - 1)M + 2x$ , and (2) If  $a, a' \in T_i$  (for  $i \in \{0, 1\}$ ) then  $d_T(a, a') = d_{T_i}(a, a')$ . ■

LEMMA 3.6. Consider two  $(h, M)$ -trees  $T = (T_0, T_1, x)$  and  $T' = (T'_0, T'_1, x')$ . For any leaves  $a_0 \in T_0, a_1 \in T_1, a'_0 \in T'_0$  and  $a'_1 \in T'_1$ ,  $d_T(a_0, a_1) = d_{T'}(a'_0, a'_1) \iff x = x'$ . ■

For a distance labeling scheme  $(L, f)$  on  $\mathcal{C}(h, M)$ , let  $W(L, h, M)$  denote the set of all labels assigned by  $L$  to nodes in trees of  $\mathcal{C}(h, M)$ , and let  $g(h, M)$  denote the minimum cardinality  $|W(L, h, M)|$  over all distance labeling schemes on  $\mathcal{C}(h, M)$ .

Hereafter, we fix  $(L, f)$  to be some distance labeling scheme attaining  $g(h, M)$ , i.e., such that  $|W(L, h, M)| = g(h, M)$ . Let  $W(x)$  denote the set of possible pairs of labels  $(v_0, v_1)$  assigned by  $L$  to some leaves  $a_j \in T_0$  and  $a_i \in T_1$  respectively, for some tree  $T = (T_0, T_1, x) \in \mathcal{C}(h, M, x)$ . Let  $\mathcal{W} = \bigcup_{x=0}^{M-1} W(x)$ . As  $\mathcal{W} \subseteq W(L, h, M) \times W(L, h, M)$  we have

LEMMA 3.7.  $|\mathcal{W}| \leq g(h, M)^2$ .

LEMMA 3.8. For every  $0 \leq x \neq x' < M$ , the sets  $W(x)$  and  $W(x')$  are disjoint.

Proof. Consider two different weights  $0 \leq x \neq x' < M$ , and assume by way of contradiction that there exists a pair  $(\lambda_1, \lambda_2) \in W(x) \cap W(x')$ . Then there exist two  $(h - 1, M)$ -trees  $T_0, T_1$  such that  $T = (T_0, T_1, x)$  uses the label  $\lambda_1$  for some leaf  $a_{j_1} \in T_0$  and the label  $\lambda_2$  for some leaf  $a_{j_2} \in T_1$ , and there exist two  $(h - 1, M)$ -trees  $T'_0, T'_1$  such that  $T' = (T'_0, T'_1, x')$  uses the label  $\lambda_1$  for some leaf  $a_{j_3} \in T'_0$  and the label  $\lambda_2$  for some leaf  $a_{j_4} \in T'_1$ . Therefore, by the definition of  $f$ ,

$$\begin{aligned} 2(h - 1)M + 2x &= d(a_{j_1}, a_{j_2}) = f(\lambda_1, \lambda_2) \\ &= d(a_{j_3}, a_{j_4}) = 2(h - 1)M + 2x', \end{aligned}$$

implying  $x = x'$ , contradiction. ■

The crux of the analysis lies in the following lemma.

LEMMA 3.9. For  $0 \leq x < M$ ,  $|W(x)| \geq g(h - 1, M^2)$ .

Proof. In any  $(h - 1, M^2)$ -tree, a weight  $w \in [0, M^2 - 1]$  can be represented by the pair of weights  $w_0 = w \bmod M, w_1 = \lfloor w/M \rfloor$ , such that  $w_0, w_1 \in [0, M - 1]$  and  $w = w_0 + Mw_1$ .

Consequently, one can associate with any  $(h - 1, M^2)$ -tree  $T'$  a pair of  $(h - 1, M)$ -trees  $T_0$  and  $T_1$  as follows. For any edge  $e$  of  $T'$  with weight  $w = w_0 + M \cdot w_1$ , let the corresponding weight of  $e$  in  $T_0$  (respectively,  $T_1$ ) be  $w_0$  (resp.,  $w_1$ ). These two trees define also a  $(h, M)$ -tree  $T = (T_0, T_1, x)$  in  $\mathcal{C}(h, M, x)$ .

Every leaf  $a_j$  of  $T'$  is now associated with two homologous leaves of  $T$ , namely, the leaf  $a_j^0 = a_j$  (occurring in the left part of  $T$ , i.e.,  $T_0$ ), and the leaf  $a_j^1 = a_{j+2^{h-1}}$  (occurring in  $T_1$ ). For every two leaves  $a_j, a_t$  of  $T'$  we now have

$$(3.3) \quad \begin{aligned} d_{T'}(a_j, a_t) &= d_{T_0}(a_j^0, a_t^0) + M \cdot d_{T_1}(a_j^1, a_t^1) \\ &= d_T(a_j^0, a_t^0) + M \cdot d_T(a_j^1, a_t^1). \end{aligned}$$

We use this observation to derive a labeling scheme for all  $(h-1, M^2)$ -trees using at most  $|W(x)|$  labels. Given an  $(h-1, M^2)$ -tree  $T'$ , consider the pair of  $(h-1, M)$ -trees  $T_0, T_1$  defined above, and use the labeling  $L$  to label the tree  $T = (T_0, T_1, x)$ . Now use the resulting labeling to define a labeling function  $L'$  for the nodes of  $T'$  as follows. A leaf  $a_j \in T'$  receives as its label the pair  $L'(a_j, T') = \langle L(a_j^0, T), L(a_j^1, T) \rangle$ . Note that this pair belongs to  $W(x)$ .

The distance decoder  $f'$  for  $(h-1, M^2)$ -trees is now obtained by setting  $f'(L'(a_j, T'), L'(a_t, T')) = f'(\langle L(a_j^0, T), L(a_j^1, T) \rangle, \langle L(a_t^0, T), L(a_t^1, T) \rangle) = f(L(a_j^0, T), L(a_t^0, T)) + M \cdot f(L(a_j^1, T), L(a_t^1, T))$ . As  $L$  is a distance labeling scheme for  $(h, M)$ -trees we have  $f(L(a_j^0, T), L(a_t^0, T)) = d_T(a_j^0, a_t^0)$  and  $f(L(a_j^1, T), L(a_t^1, T)) = d_T(a_j^1, a_t^1)$ , so by Eq. (3.3),  $f'(L'(a_j, T'), L'(a_t, T')) = d_T(a_j^0, a_t^0) + M \cdot d_T(a_j^1, a_t^1) = d_{T'}(a_j, a_t)$ .

So we have obtained a labeling scheme  $\langle L', f' \rangle$  labeling any  $(h-1, M^2)$ -tree with labels taken from  $W(x)$ . It follows that  $|W(x)| \geq g(h-1, M^2)$ . ■

LEMMA 3.10.  $g(h, M) \geq M^{h/2}$ .

THEOREM 3.6. For the family  $\mathcal{T}$  of binary trees with weights from the range  $[0, M-1]$ ,

$$\ell(\mathcal{T}, n) \geq \frac{1}{2}(\log n - 2) \log M.$$

**Proof.** By Lemma 3.10, for the class  $\mathcal{C}(h, M)$  we have  $\ell(\mathcal{C}(h, M)) \geq \frac{h}{2} \cdot \log M$ . The number of nodes of the an unweighted  $(h, M)$ -tree is  $n = 3 \cdot 2^h - 2$ . This yields the theorem, as

$$\ell(\mathcal{T}, n) \geq \frac{1}{2} \cdot \log \left( \frac{n+2}{3} \right) \cdot \log M \geq \frac{1}{2}(\log n - 2) \log M \quad \blacksquare$$

COROLLARY 3.3. For the family  $\mathcal{T}$  of unweighted binary trees,  $\ell(\mathcal{T}, n) \geq \frac{1}{8} \log^2 n - O(\log n)$ .

**Proof.** A  $(h, M)$ -tree can be transformed into an unweighted tree by replacing each edge  $e$  of weight  $w$  with a path of  $w$  (unweighted) edges. Let  $t(h, M)$  be the maximal number of nodes of the unweighted tree corresponding to the construction of a  $(h, M)$ -tree. Then  $t(h, M) \leq 3 \cdot 2^h \cdot M$ . Taking  $h = \log \sqrt{n}/3$  and  $M = \sqrt{n}/3$ , and applying Lemma 3.10, we obtain

$$g(h, M) \geq 2^{\frac{h}{2} \log M} \geq 2^{\frac{1}{8} \log^2 (n/3)}$$

for sufficiently large  $n$ . Moreover, we obtain an unweighted tree with at most  $t(h, M) \leq n$  nodes. Note that the depth is at most  $2hM < \sqrt{n} \log n$ . So for unweighted binary trees with  $n$  nodes and depth  $O(\sqrt{n} \log n)$ , at least  $\frac{1}{8} \log^2 n - O(\log n)$  bits may be necessary. ■

## References

- [1] N. ALON, P. D. SEYMOUR, AND R. THOMAS, *A separator theorem for graphs with an excluded minor and its applications*, in 22<sup>nd</sup> Symp. on Theory of Computing, ACM, May 1990, pp. 293–299.
- [2] M. A. BREUER, *Coding the vertexes of a graph*, IEEE Trans. Inf. Theory, IT-12 (1966), pp. 148–153.
- [3] M. A. BREUER AND J. FOLKMAN, *An unexpected result on coding the vertices of a graph*, J. of Mathematical Analysis and Applications, 20 (1967), pp. 583–600.
- [4] C. GAVOILLE, D. PELEG, S. PÉRENNES, AND R. RAZ, *Distance labeling in graphs*, Research Report RR-1182-97, LaBRI, Univ. of Bordeaux, Talence, France, 1999.
- [5] J. R. GILBERT, J. P. HUTCHINSON, AND R. E. TARJAN, *A separation theorem for graphs of bounded genus*, J. of Algorithms, 5 (1984), pp. 391–407.
- [6] R. L. GRAHAM AND H. O. POLLAK, *On embedding graphs in squashed cubes*, Lecture Notes in Mathematics, 303 (1972), pp. 99–110.
- [7] S. KANNAN, M. NAOR, AND S. RUDICH, *Implicit representation of graphs*, in 20<sup>th</sup> Symp. on Theory of Computing, ACM, May 1988, pp. 334–343.
- [8] M. KATZ, N. KATZ, AND D. PELEG, *Distance labeling schemes for well-separated graph classes*, in 17<sup>th</sup> Symp. on Theoretical Aspects of Computer Science, vol. LNCS 1770, Springer, Feb. 2000, pp. 516–528.
- [9] D. PELEG, *Proximity-preserving labeling schemes and their applications*, in 25<sup>th</sup> Workshop on Graph-Theoretic Concepts in Computer Science, vol. LNCS 1665, Springer, June 1999, pp. 30–41.
- [10] ———, *Informative labeling schemes for graphs*, in 25<sup>th</sup> Symp. on Mathematical Foundations of Computer Sci., vol. LNCS 1893, Springer, Aug. 2000, pp. 579–588.
- [11] D. PELEG AND E. UPFAL, *A trade-off between space and efficiency for routing tables*, J. ACM, 36 (1989), pp. 510–530.
- [12] S. PLOTKIN, S. RAO, AND W. D. SMITH, *Shallow excluded minors and improved graph decomposition*, in 5<sup>th</sup> Symp. on Discrete Algorithms, ACM-SIAM, 1994, pp. 462–470.
- [13] N. SANTORO AND R. KHATIB, *Labelling and implicit routing in networks*, The Computer Journal, 28 (1985), pp. 5–8.
- [14] P. WINKLER, *Proof of the squashed cube conjecture*, Combinatorica, 3 (1983), pp. 135–139.