

---

This is an electronic reprint of the original article.  
This reprint may differ from the original in pagination and typographic detail.

Balliu, Alkida; Brandt, Sebastian; Chang, Yi Jun; Olivetti, Dennis; Studený, Jan; Suomela, Jukka; Tereshchenko, Aleksandr

## Locally checkable problems in rooted trees

*Published in:*  
Distributed Computing

*DOI:*  
[10.1007/s00446-022-00435-9](https://doi.org/10.1007/s00446-022-00435-9)

Published: 01/09/2023

*Document Version*  
Publisher's PDF, also known as Version of record

*Published under the following license:*  
CC BY

*Please cite the original version:*  
Balliu, A., Brandt, S., Chang, Y. J., Olivetti, D., Studený, J., Suomela, J., & Tereshchenko, A. (2023). Locally checkable problems in rooted trees. *Distributed Computing*, 36(3), 277–311. <https://doi.org/10.1007/s00446-022-00435-9>

---

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.



# Locally checkable problems in rooted trees

Alkida Balliu<sup>1</sup> · Sebastian Brandt<sup>2</sup> · Yi-Jun Chang<sup>3</sup> · Dennis Olivetti<sup>1</sup> · Jan Studený<sup>4</sup> · Jukka Suomela<sup>4</sup> · Aleksandr Tereshchenko<sup>4</sup>

Received: 15 November 2021 / Accepted: 11 July 2022 / Published online: 22 August 2022  
© The Author(s) 2022

## Abstract

Consider any locally checkable labeling problem  $\Pi$  in *rooted regular trees*: there is a finite set of labels  $\Sigma$ , and for each label  $x \in \Sigma$  we specify what are permitted label combinations of the children for an internal node of label  $x$  (the leaf nodes are unconstrained). This formalism is expressive enough to capture many classic problems studied in distributed computing, including vertex coloring, edge coloring, and maximal independent set. We show that the distributed computational complexity of any such problem  $\Pi$  falls in one of the following classes: it is  $O(1)$ ,  $\Theta(\log^* n)$ ,  $\Theta(\log n)$ , or  $n^{\Theta(1)}$  rounds in trees with  $n$  nodes (and all of these classes are nonempty). We show that the complexity of any given problem is the same in all four standard models of distributed graph algorithms: deterministic LOCAL, randomized LOCAL, deterministic CONGEST, and randomized CONGEST model. In particular, we show that randomness does not help in this setting, and the complexity class  $\Theta(\log \log n)$  does not exist (while it does exist in the broader setting of general trees). We also show how to systematically determine the complexity class of any such problem  $\Pi$ , i.e., whether  $\Pi$  takes  $O(1)$ ,  $\Theta(\log^* n)$ ,  $\Theta(\log n)$ , or  $n^{\Theta(1)}$  rounds. While the algorithm may take exponential time in the size of the description of  $\Pi$ , it is nevertheless practical: we provide a freely available implementation of the classifier algorithm, and it is fast enough to classify many problems of interest.

## 1 Introduction

We aim at *systematizing* and *automating* the study of computational complexity in the field of distributed graph algo-

rithms. Many key problems of interest in the field are *locally checkable*. While it is known that questions related to the distributed computational complexity of locally checkable problems are *undecidable* in general graphs [13,24], there is no known obstacle that would prevent one from completely automating the study of locally checkable problems in *trees*. Achieving this is one of the major open problems in the field: currently only parts of the complexity landscape are known to be decidable [16], and the general decidability results are primarily of theoretical interest; *practical* automatic techniques are only known for specific families of problems [2,13,17].

In this work we show that the study of locally checkable graph problems can be completely automated in *regular rooted trees*. We not only give a full classification of the distributed complexity of any such problem (in all the usual models of distributed computing: deterministic and randomized LOCAL and CONGEST), but we also present an algorithm that can automatically determine the complexity class of any given problem (with one caveat: our algorithm determines if the complexity is  $n^{\Theta(1)}$ , but not the precise exponent in this case). Even though the algorithm takes in the worst case exponential time in the size of the problem description, it is nevertheless practical: we have implemented it for the case of binary trees, and it is in practice very fast, classifying e.g.

---

✉ Jukka Suomela  
jukka.suomela@aalto.fi

Alkida Balliu  
alkida.balliu@gssi.it

Sebastian Brandt  
brandt@cispa.de

Yi-Jun Chang  
cyjjun@nus.edu.sg

Dennis Olivetti  
dennis.olivetti@gssi.it

Jan Studený  
jan.studený@aalto.fi

Aleksandr Tereshchenko  
aleksandr.tereshchenko@aalto.fi

- <sup>1</sup> Gran Sasso Science Institute, L'Aquila, Italy
- <sup>2</sup> CISPA Helmholtz Center for Information Security, Saarbrücken, Germany
- <sup>3</sup> National University of Singapore, Queenstown, Singapore
- <sup>4</sup> Aalto University, Espoo, Finland

the sample problems that we present here in a matter of milliseconds [27].

### 1.1 Setting

In this work we study locally checkable problems defined in regular, unlabeled, not necessarily balanced, rooted trees of bounded degree. For our purposes, such a problem  $\Pi$  is specified as a triple  $(\delta, \Sigma, C)$ , where  $\delta \in \mathbb{N}$  is the number of children for the internal nodes,  $\Sigma$  is a finite set of *labels*, and  $C$  is the set of permitted *configurations*. Each configuration looks like  $x : y_1 y_2 \cdots y_\delta$ , indicating that if the label of an internal node is  $x$ , then one of the possible labelings for its  $\delta$  children is  $y_1, y_2, \dots, y_\delta$ , in some order (that is, the order of the children does not matter). The leaf nodes are unconstrained.

The reason why we choose this specific setting is the following. As soon as we consider *inputs*, it is known that decidability questions become much harder [1,14], and since even the case with no inputs is still not understood, we try to understand this setting first. Moreover, it is possible to use non-regular trees to encode trees with inputs, and for this reason we constrain only nodes with exactly  $\delta$  children.

### 1.2 Example: 3-coloring

Consider the problem of 3-coloring binary trees, i.e., trees in which internal nodes have  $\delta = 2$  children. The possible labels of the nodes are  $\Sigma = \{1, 2, 3\}$ . The color of a node has to be different from the colors of any of its children; hence we can write down the set of configurations e.g. as follows:

$$C = \{1 : 22, 1 : 23, 1 : 33, \\ 2 : 11, 2 : 13, 2 : 33, \\ 3 : 11, 3 : 12, 3 : 22\}. \quad (1)$$

We emphasize that the ordering of the children is irrelevant here; hence  $1 : 23$  and  $1 : 32$  are the same configuration. It is easy to verify that this is a straightforward correct encoding of the 3-coloring problem in binary trees.

It is well-known that this problem can be solved in the LOCAL model of distributed computing in  $O(\log^* n)$  rounds in rooted trees [9, Section 3.4], using the technique by Cole and Vishkin [19], and this is also known to be tight, both for deterministic and randomized algorithms [21,23].

One can also in a similar way define the problem of 2-coloring binary trees; it is easy to check that this is a global problem, with complexity  $\Theta(n)$  rounds:

$$C = \{1 : 22, 2 : 11\}. \quad (2)$$

### 1.3 Example: maximal independent set

Let us now look at a bit more interesting problem: maximal independent sets (MIS). Let us again stick to binary trees, i.e.,  $\delta = 2$  children. The first natural idea for encoding MIS as a locally checkable problem would be to try to use only two labels, 0 and 1, with 1 indicating that a node is in the independent set, but this is not sufficient to express both the notion of independence and the notion of maximality. However, three labels will be sufficient to correctly capture the problem. We set  $\Sigma = \{1, a, b\}$ , with 1 indicating that a node is in the independent set, and choose the following configurations:

$$C = \{1 : aa, 1 : ab, 1 : bb, a : bb, b : b1, b : 11\}. \quad (3)$$

Now it takes a bit more effort to convince oneself that this indeed correctly captures the idea of maximal independent sets. The key observations are these: a node with label 1 cannot be adjacent to another node with label 1, a node with label  $a$  has to have 1 above it, and a node with label  $b$  has to have 1 below it, so nodes with label 1 clearly form a maximal independent set. Conversely, given any maximal independent set  $X$  we can find a corresponding label assignment if we first assign labels 1 to nodes in  $X$ , then assign labels  $b$  to the parents of the nodes in  $X$ , and finally label the remaining nodes with label  $a$ . The only minor technicality is that this labeling corresponds to an MIS only for internal nodes of the tree, but as is often the case, once the internal parts are solved correctly, one can locally fix the labels near the root and the leaves.

Maximal independent set is a well-known symmetry-breaking problem, and e.g. in the case of a directed path ( $\delta = 1$ ) it is known to be as hard as e.g. 3-coloring. Hence one might expect that MIS on rooted regular binary trees also has got the complexity of  $\Theta(\log^* n)$  rounds in the LOCAL model. *This is not the case—maximal independent set in rooted binary trees can be solved in constant time!* Indeed, this is a good example of a non-trivial constant-time-solvable problem. It can be solved in exactly 4 rounds, using the following idea (again, omitting some minor details related to what happens e.g. near the root).

First, we need to pick some consistent way of referring to your “left” child and the “right” child (for this reason, we can assume that a port numbering is available, that is, a node can send a message to a specific child, by indexing it with a number from 1 to  $\delta$ , or we can assume that nodes have unique identifiers, and then we can order the children by their unique identifiers). Label all nodes first with an empty string. Then we repeat the following step for 4 times: add 0 to your string and send it to your left child, and add 1 to your string and send it to your right child. Your new label is the label that you received from your parent. This way all nodes get



labeled with a 4-bit string (see Fig. 1a). A key property is this: if my string is  $xyzw$ , the string of my parent is  $0xyz$  or  $1xyz$ . Finally, interpret the binary string as a number between 0 and 15, and output the corresponding element of the following string (using 0-based indexing; see Fig. 1b):

$$b1abbb1bb11bbb1b. \quad (4)$$

One can verify the correctness of the algorithm by checking all  $2^3$  possible cases: for example, if a node is labeled with  $x010$ , it will output either symbol 2 of (4), which is  $a$ , or symbol 10, which is 1. Its two children will have labels 0100 and 0101, so they will output symbols 4 and 5 of (4), which are  $b$  and  $b$ . This results in a configuration  $a : bb$  or  $1 : bb$ , both of which are valid in (3).

The key point of the example is this: even though the algorithm is somewhat involved, we can use the computer program accompanying in this work to *automatically* discover this algorithm and to determine that this problem is indeed constant-time solvable! Also, this problem demonstrates that there are  $O(1)$ -round-solvable locally checkable problems in rooted regular trees that require strictly more than zero rounds, while e.g. in the previously-studied family of binary labeling problems [2] all  $O(1)$ -round-solvable problems are known to be zero-round solvable.

#### 1.4 Example: branch 2-coloring

As the final example, let us consider the following problem, with  $\delta = 2$  and  $\Sigma = \{1, 2\}$ :

$$C = \{1 : 12, 2 : 11\}. \quad (5)$$

This problem is, in essence, 2-coloring with a choice: starting with a node of label 1 and going downwards, there is always a monochromatic path labeled with 1, 1, 1, 1,  $\dots$ , and a properly colored path labeled with 1, 2, 1, 2,  $\dots$ . It turns out that the choice makes enough of a difference: the complexity of this problem is  $\Theta(\log n)$  rounds. We encourage the reader to come up with an algorithm and a matching lower bound—with our techniques we get a tight result immediately.

#### 1.5 Contributions

As we have seen, the family of locally checkable problems in regular rooted trees is rich and expressive. Using auxiliary labels similar to what we saw in the MIS example in Sect. 1.3, we can encode, in essence, any locally checkable labeling problem (LCL problem) [24] in the classic sense, as long as the problem is such that the interesting part is related to what happens in the internal parts of regular trees. We have already seen that there are problems with at least four distinct complexity classes:  $O(1)$ ,  $\Theta(\log^* n)$ ,  $\Theta(\log n)$ , and  $\Theta(n)$ . In

Sect. 8 we also show how to generate problems of complexity  $\Theta(n^{1/k})$  for any  $k = 1, 2, 3, \dots$

We prove in this work that *this list is exhaustive*: any problem that can be represented in our formalism has complexity  $O(1)$ ,  $\Theta(\log^* n)$ ,  $\Theta(\log n)$ , or  $\Theta(n^{1/k})$  in rooted regular trees with  $n$  nodes. This is a *robust* result that does not depend on the specific choice of the model of computing: the complexity of a given problem is the same, regardless of whether we are looking at the LOCAL model or the CONGEST model, and regardless of whether we are using deterministic or randomized algorithms.

One of the surprising consequences is that *randomness does not help* in rooted regular trees. In unrooted regular trees there are problems (the canonical example being the *sinkless orientation* problem) that can be solved with the help of randomness in  $\Theta(\log \log n)$  rounds, while the deterministic complexity is  $\Theta(\log n)$  [13]. This class of problems disappears in rooted trees.

Our main contribution is that the *complexity of any given problem in this formalism is decidable*: there is an algorithm that, given the description of a problem  $\Pi$  as a list of permitted configurations, outputs the computational complexity of problem  $\Pi$ , putting it in one of the four possible classes, i.e., determines whether the complexity is  $O(1)$ ,  $\Theta(\log^* n)$ ,  $\Theta(\log n)$ , or  $\Theta(n^{1/k})$  for some  $k$ ; in the fourth case our algorithm does not determine the exponent  $k$ , but then one could (at least in principle) use the more general decision procedure by Chang [14] to determine the value of  $k$ .

While our algorithm takes in the worst case exponential time in the size of the description of  $\Pi$ , the approach is nevertheless *practical*. We have *implemented* the algorithm for the case of  $\delta = 2$ , and made it freely available online [27]. Even though it is not at all optimized for performance, it classifies for example all of our sample problems above in a matter of *milliseconds*.

We summarize our key results and compare them with prior work in Table 1.

## 2 Related work

### 2.1 Landscape of LCL problems in the LOCAL model

#### 2.1.1 Paths and cycles

We know that, on graph families such as paths and cycles, there are LCLs with complexities (both deterministic and randomized) of  $O(1)$  (e.g., trivial problems),  $\Theta(\log^* n)$  [19,21,23] (e.g., 3-coloring), and  $\Theta(n)$  (e.g., global problems such as properly orienting a path/cycle). Moreover, there are complexity *gaps*, that is, in these families of graphs, there are no LCLs with round complexity between  $\omega(1)$  and  $o(\log^* n)$  [24], and between  $\omega(\log^* n)$  and  $o(n)$  [15]. These works

**Table 1** An overview of the landscape and decidability of the round complexity of LCL problems in the LOCAL model. The case studied in the present work (unlabeled, rooted, regular trees) is highlighted with shading, and the darker shade indicates the key new results. The decid-

ability is given assuming  $P \neq PSPACE \neq EXPTIME$ . We have listed a few key references for each column, focusing on decidability aspects; the overall picture of the complexity landscape is the result of a long sequence of papers, including [4,5,7,12,15,19,21,23]

Setting	Paths and cycles				Trees				General		
no input	✓	✓	✓		✓	✓	✓	✓			
regular	✓	✓			✓	✓	✓	✓			
directed or rooted	✓	✓					✓				
binary output	✓					✓					
homogeneous					✓						
<b>Complexity classes</b>	$O(1)$	D+R	D+R	D+R	D+R	D+R	D+R	D+R	D+R	D+R	
...		—	—	—	—	—	—	—	—	—	
	$\Theta(\log \log^* n)$	—	—	—	—	—	—	?	?	D+R	
...		—	—	—	—	—	—	?	?	D+R	
	$\Theta(\log^* n)$	—	D+R	D+R	D+R	D+R	—	D+R	D+R	D+R	
...		—	—	—	—	—	—	—	—	—	
	$\Theta(\log \log n)$	—	—	—	—	R	R	—	R	R	
	$\Theta(\log^\alpha \log n)$	—	—	—	—	—	—	—	—	?	
...		—	—	—	—	—	—	—	—	—	
	$\Theta(\log n)$	—	—	—	—	D+R	D+R	D+R	D+R	D+R	
...		—	—	—	—	—	—	—	—	D+R	
	$\Theta(n^{1/k})$	—	—	—	—	—	—	(n)	?	(n)	
...		—	—	—	—	—	—	—	—	D+R	
	$\Theta(n)$	D+R	D+R	D+R	D+R	—	D+R	D+R	D+R	D+R	
<b>Decidability</b>	P	✓	✓	✓	—	?	(D)	?	?	—	
	PSPACE	✓	✓	✓	?	?	(D)	?	?	—	
	EXPTIME	✓	✓	✓	?	?	(D)	(k)	?	?	
	decidable	✓	✓	✓	✓	?	(D)	✓	(H)	(H)	
<b>References</b>		[2, 13, 24]	[13, 24]	[17]	[1]	[8]	[2]	this work	[14, 16]	[14, 16]	[13, 24]
<b>Legend</b>	✓ = yes ? = unknown — = not possible $\alpha > 1$ $k = 2, 3, \dots$ D = class exists for deterministic algorithms R = class exists for randomized algorithms (n) = the current construction assumes the knowledge of n; unknown without this information (k) = does not determine the value of k for the class $\Theta(n^{1/k})$ (D) = known only for deterministic complexities, unknown for randomized (H) = known only for classes between $\Omega(\log n)$ and $O(n)$										

show that the only possible complexities for LCL problems on paths and cycles are  $O(1)$ ,  $\Theta(\log^* n)$ , and  $\Theta(n)$ , and randomness does not help in solving problems faster.

### 2.1.2 Trees

For the case of the graph family of trees almost everything is understood nowadays. As in the case of paths and cycles, we have LCLs with time complexities (both deterministic and randomized)  $O(1)$ ,  $\Theta(\log^* n)$ , and  $\Theta(n)$ . On trees, we know

that there is more: there are LCL problems with both deterministic and randomized complexity of  $\Theta(\log n)$  (e.g., problems of the form “copy the input of the nearest leaf”), and  $\Theta(n^{1/k})$  for any  $k \geq 2$  [16]. Moreover, there are cases where randomness helps, in fact there are problems that have  $\Theta(\log n)$  deterministic and  $\Theta(\log \log n)$  randomized complexity [12]. As far as gaps are concerned, let us first consider the spectrum of time complexities of  $\omega(\log^* n)$ , and then the one of  $o(\log^* n)$ . Chang et al. [15] showed that the deterministic complexity of any LCL problem on bounded-degree trees is either  $O(\log^* n)$  or  $\Omega(\log n)$ , while its randomized complexity is either  $O(\log^* n)$  or  $\Omega(\log \log n)$ . Moreover, Chang and Pettie [16] showed that any algorithm that takes  $n^{o(1)}$  rounds can be sped up to run in  $O(\log n)$  rounds. Balliu et al. [4] showed that there is a gap between  $\omega(\sqrt{n})$  and  $o(n)$  for deterministic algorithms, and Chang [14] extended these results and showed that there is a gap between  $\omega(n^{1/k})$  and  $o(n^{1/(k-1)})$ , for any  $k \geq 2$ , for both deterministic and randomized algorithms. The spectrum of time complexities of  $o(\log^* n)$  is still not entirely understood. Chang and Pettie [16] showed that ideas similar to Naor and Stockmeyer [24] can be used to prove that there are no LCLs on bounded-degree trees with time complexity between  $\omega(1)$  and  $o(\log \log^* n)$ . Also, in the same paper, the authors conjectured that it should be possible to extend this gap up to  $o(\log^* n)$ . While this still remains an open question, Balliu et al. [8] showed that such a gap exists for a special subclass of LCLs, called *homogeneous* LCLs. Observe that all the mentioned results hold for the setting of *unrooted* trees, and in this setting there are still many open questions related to *decidability*. In this work, we prove decidability results for a restriction of this setting, that is, for *rooted* trees.

### 2.1.3 General graphs

In general bounded-degree graphs there are LCLs with the same time complexity as in trees, so the question is if there are also the same gaps, or if in the case of general graphs we have a denser spectrum of complexities. First of all, the gaps of the lower spectrum on trees hold also on general graphs: we still have the  $\omega(1) - o(\log \log^* n)$  gap for both deterministic and randomized algorithms, the  $\omega(\log^* n) - o(\log n)$  gap for deterministic algorithms, and the  $\omega(\log^* n) - o(\log \log n)$  gap for randomized algorithms. Also, Chang and Pettie [16] showed that any  $o(\log n)$ -round randomized algorithm can be sped up to run in  $O(T_{LLL})$  rounds, where  $T_{LLL}$  is the time required for solving with randomized algorithms the distributed constructive Lovász Local Lemma problem (LLL) [18] under a polynomial criterion. By combining this result with the results on the complexity of LLL by Fischer and Ghaffari [20] and the network decomposition one by Rozhoňand Ghaffari [26], we get a gap for randomized algorithms between  $\omega(\text{poly}(\log \log n))$  and  $o(\log n)$ . Balliu et al.

[7] showed that, differently from the case of trees, the regions between  $\omega(\log \log^* n)$  and  $o(\log^* n)$  and between  $\omega(\log n)$  and  $o(n)$  are dense. In fact, for any complexity  $T$  in these regions, it is possible to define an LCL with a time complexity that is arbitrary close to  $T$ . Also, in the case of trees, randomness either helps exponentially or not at all, while in the case of general graphs this is not the case anymore. In fact, Balliu et al. [5] showed that there are LCL problems on general graphs where randomness helps only polynomially by defining LCLs with deterministic complexity  $\Theta(\log^k n)$  and randomized complexity  $\Theta(\log^{k-1} n \log \log n)$ , for any integer  $k \geq 1$ .

### 2.1.4 Special settings

Over the years, researchers have investigated the complexity of interesting subclasses of LCLs. We already mentioned homogeneous LCLs on trees [8], that, on a high level, are LCLs for which the hard instances are  $\Delta$ -regular trees. For this subclass of LCL problems, the spectrum of deterministic complexities consists of  $O(1)$ ,  $\Theta(\log^* n)$ , and  $\Theta(\log n)$ . Also, as in the case of trees, there are cases where randomness helps: there are homogeneous LCLs with  $\Theta(\log n)$  deterministic and  $\Theta(\log \log n)$  randomized complexity. These are the only possible complexities for homogeneous LCLs. Brandt et al. [13] studied LCLs on  $d$ -dimensional torus grids, and showed that there are LCLs with complexity (both deterministic and randomized)  $O(1)$ ,  $\Theta(\log^* n)$ , and  $\Theta(n^{1/d})$ . The authors showed that these are the only possible complexities, implying that randomness does not help. Balliu et al. [2] studied *binary labeling problems*, that are LCLs that can be expressed with no more than two labels in the *edge labeling* formalism [3,25] (such LCLs include, for example, sinkless orientation). The authors showed that, in trees, there are no such LCLs with deterministic round complexity between  $\omega(1)$  and  $o(\log n)$ , and between  $\omega(\log n)$  and  $o(n)$ , proving that the spectrum of deterministic complexities of binary labeling problems in bounded-degree trees consists of  $O(1)$ ,  $\Theta(\log n)$  and  $\Theta(n)$ . The authors also studied the randomized complexity of binary labeling problems that have deterministic complexity  $\Theta(\log n)$ , showing that for some of them randomness does not help, while for some others it does help (note that from previous work we know that, in this case, randomness either helps exponentially or not at all). Determining the tight randomized complexity of all binary labeling problems is still an open question.

## 2.2 Decidability of LCL problems

As we have seen, there are often gaps in the spectrum of distributed complexities of LCLs. Hence, a natural question that arises is the following: given a specific LCL, can we decide on which side of the gap it falls? In other words, are

these classifications of LCL problems decidable? We can push this question further and ask whether it is possible to automate the design of distributed algorithms for optimally solving LCLs. There is a long line of research that has investigated these kind of questions.

For graph families that consist of *unlabeled* paths and cycles (that is, nodes do not have any label in input), the complexity of a given LCL is decidable [13,17,24]. The next natural question is whether we have decidability in the case of trees (rooted or not). Because the structure of a tree can be used to encode input labels, researchers had to first understand the role of input labels in decidability. For this purpose, Balliu et al. [1] studied the decidability of *labeled* paths and cycles, showing that the complexity of LCLs in this setting is decidable, but it is PSPACE-hard to decide it, and this PSPACE-hardness result extends also for the case of bounded-degree unlabeled trees (since the structure of the tree may encode input labels). The authors also show how to automate the design of asymptotically optimal distributed algorithms for solving LCLs in this context. Later, Chang [14] improved these results showing that, in this setting, it is EXPTIME-hard to decide the complexity of LCLs. While the decidability on bounded degree trees is still an open question, there are some positive partial results in this direction. In fact, Chang and Pettie [16] along with the  $\omega(\log n) - n^{o(1)}$  gap, showed also that we can decide on which side of the gap the complexity of an LCL lies. Moreover, Balliu et al. [2] showed that, the deterministic complexity of binary labeling problems on trees is decidable and we can automatically find optimal algorithms that solve such LCLs. The works of Brandt [10] and Olivetti [25] played a fundamental role in further understanding to which extent we can automate the design of algorithms that optimally solve LCLs.

Unfortunately, in general, the complexity of an LCL is not decidable. In fact, Naor and Stockmeyer showed that, even on unlabeled non-toroidal grid graphs, it is undecidable whether the complexity of a given LCL is  $O(1)$  [24]. For unlabeled toroidal grids, Brandt et al. [13] showed that, given an LCL, it is decidable whether its complexity is  $O(1)$ , but it is undecidable whether its complexity is  $\Theta(\log^* n)$  or  $\Theta(n)$ . On the positive side, the authors showed that, given an LCL with round complexity  $O(\log^* n)$ , one can automatically find an  $O(\log^* n)$  rounds algorithm that solves it.

### 3 Road map

We will start by providing some useful definitions in Sect. 4. Then, in Sect. 5 we will consider the spectrum of complexities in the  $\Omega(\log n)$  region (that is,  $\Theta(\log n)$  and above). We will define an object called *certificate for  $O(\log n)$  solvability*, for which we will prove, in Theorem 5.3, that we can decide the existence in polynomial time. We will prove in Theorem

5.1 that, if such a certificate for a problem exists, then the problem can be solved in  $O(\log n)$  time with a deterministic algorithm, even in the CONGEST model, while if such a certificate does not exist then we will prove in Theorem 5.2 that the problem requires  $n^{\Omega(1)}$  rounds, even in the LOCAL model and even for randomized algorithms. By combining these results, we will essentially obtain a *decidable gap* between  $\omega(\log n)$  and  $n^{o(1)}$  that is robust on the choice of the model.

We will then consider, in Sect. 6, the spectrum of complexities in the  $O(\log n)$  region. We will define the notion of *certificate for  $O(\log^* n)$  solvability*, and we will prove, in Theorem 6.10, that we can decide in exponential time if such a certificate exists. We will also prove, in Theorem 6.3, that the existence of such a certificate implies a deterministic  $O(\log^* n)$  algorithm for the CONGEST model, while we will prove in Lemma 6.7 that the non-existence of such a certificate implies an  $\Omega(\log n)$  randomized lower bound for the LOCAL model. Hence, also in this case we obtain a decidable gap that is robust on the choice of the model.

In Sect. 7, we consider the spectrum of complexities in the  $O(\log^* n)$  region. We will define the notion of *certificate for  $O(1)$  solvability*, that will be nothing else than a certificate for  $O(\log^* n)$  solvability that has some special property. We will show, in Theorem 7.8, that also in this case, we can decide its existence in exponential time, and we will show in Theorem 7.2 that its existence implies a constant-time deterministic algorithm for the CONGEST model, while we will show in Theorem 7.7 that the non-existence implies an  $\Omega(\log^* n)$  lower bound for the LOCAL model. Hence, we will obtain that there are only four possible complexities,  $O(1)$ ,  $\Theta(\log^* n)$ ,  $\Theta(\log n)$ , and  $n^{\Omega(1)}$ , and that for all problems we can decide which of these four complexities is the right one.

For the fine-grained structure inside the  $n^{\Omega(1)}$  class we refer to the prior work [6,14,16]; while these papers study the case of *unrooted* trees, we note that the orientation can be encoded as a locally checkable input, and the results are also applicable here. It follows that there are only classes  $O(1)$ ,  $\Theta(\log^* n)$ ,  $\Theta(\log n)$ , and  $\Theta(n^{1/k})$  for  $k = 1, 2, \dots$ , and the exact class (including the value of  $k$ ) is decidable. Although the existence of the gap  $\omega(n^{1/(k+1)}) - o(n^{1/k})$  [14] applies to regular rooted trees, the problems with complexity  $\Theta(n^{1/k})$  that have been shown to exist in [16] are not defined on regular rooted trees (e.g., nodes of different degrees may have different constraints). In Sect. 8, we define problems with complexity  $\Theta(n^{1/k})$  in regular rooted trees, showing that the complexity class  $\Theta(n^{1/k})$  is non-empty for regular rooted trees.

### 4 Definitions

In this section we define some notions that will be used in the following sections.

## 4.1 Input graphs

We assume that all input graphs will be *unlabeled rooted trees* where each node has either exactly  $\delta$  or 0 children for some positive integer  $\delta$ . That is, input graphs are *full  $\delta$ -ary trees*. For convenience, when not specified otherwise, a tree  $T$  is assumed to be a full  $\delta$ -ary tree.

## 4.2 Models of computing

The models that we consider in this work are the classical LOCAL and CONGEST model of distributed computing. Let  $G$  be any graph with  $n$  nodes and maximum degree  $\Delta$ . In the LOCAL model, each node of  $G$  is equipped with an identifier in  $\{1, 2, \dots, \text{poly}(n)\}$ , and the initial knowledge of a node consists of its own identifier, its degree (i.e., the number of incident edges), the total number  $n$  of nodes, and  $\Delta$  (in the case of rooted trees, each node knows also which of its incident edges connects it to its parent). Nodes try to learn more about the input instance by communicating with the neighbors. The computation proceeds in synchronous rounds, and at each round nodes send messages to neighbors, receive messages from them, and perform local computation. Messages can be arbitrarily large and the local computational can be of arbitrary complexity. Each node must terminate its computation at some point and decide on its local output. The running time of a distributed algorithm running at each node in the LOCAL model is determined by the number of rounds needed such that all nodes have produced their local output. In the randomized version of the LOCAL model, each node has access to its own stream of random bits. The randomized algorithms considered in this paper are Monte Carlo ones, that is, a randomized algorithm of complexity  $T$  that solves a problem  $P$  must terminate at all nodes upon  $T$  rounds and this should result in a global solution for  $P$  that is correct with probability at least  $1 - 1/n$ .

There is only one difference between the CONGEST and the LOCAL model, and it lies in the size of the messages. While in the LOCAL model messages can be arbitrarily large, in the CONGEST model the size of the messages is bounded by  $O(\log n)$  bits.

## 4.3 LCL problems

We define LCL problems as follows.

**Definition 4.1** (*LCL problem*) An LCL problem is a triple  $\Pi = (\delta, \Sigma, C)$  where:

- $\delta$  is the number of allowed children;
- $\Sigma$  is a finite set of (output) labels;
- $C$  is a set of tuples of size  $\delta + 1$  from  $\Sigma^{\delta+1}$  called *allowed configurations*.

A configuration  $(a, b_1, \dots, b_\delta)$  will also be written as  $(a : b_1, \dots, b_\delta)$ , in order to highlight that the label  $a$  is for the parent and  $b_1, \dots, b_\delta$  are labels of the leaves. Sometimes we will omit the commas, and just write  $(a : b_1 \dots b_\delta)$ . Sometimes even the parenthesis will be omitted, obtaining  $a : b_1 \dots b_\delta$ , that is the notation used in e.g. Sect. 1.3. As a shorthand notation, for an LCL problem  $\Pi$ , we will also denote the labels and configurations of  $\Pi$  by  $\Sigma(\Pi)$  and  $C(\Pi)$ .

**Definition 4.2** (*solution*) A solution to an LCL problem  $\Pi$  for a tree  $T$  is a labeling function  $\lambda$  for which:

- every node  $v \in T$  is labeled with a label  $\lambda(v)$  from  $\Sigma(\Pi)$ ;
- every node  $v \in T$  with  $\delta$  children  $v_1, \dots, v_\delta$  satisfies that there exists a permutation  $\rho: \{1, \dots, \delta\} \rightarrow \{1, \dots, \delta\}$  such that  $(\lambda(v) : \lambda(v_{\rho(1)}), \dots, \lambda(v_{\rho(\delta)}))$  is in  $C(\Pi)$ .

In other words, a solution is a labeling for the nodes that must satisfy some local constraints. Note that only nodes with  $\delta$  children are constrained, but that such LCL problems could be well-defined even on non-full  $\delta$ -ary trees (nodes with a number of children different from  $\delta$  are unconstrained). Full  $\delta$ -ary trees are the hardest instances for the problems as every node is constrained.

**Definition 4.3** (*restriction*) Given an LCL problem  $\Pi = (\delta, \Sigma, C)$ , a restriction of  $\Pi$  to labels  $\Sigma' \subseteq \Sigma$  is a new LCL problem  $\Pi' = (\delta, \Sigma', C')$ , where  $C'$  consists of all configurations in  $C$  that only use labels in  $\Sigma'$ .

**Definition 4.4** (*continuation below*) Let  $\Pi$  be an LCL problem. Label  $\sigma \in \Sigma(\Pi)$  has a *continuation below* if there exists a configuration  $(\sigma : \sigma_1, \dots, \sigma_\delta) \in C(\Pi)$ .

**Definition 4.5** (*continuation below with specific labels*) Let  $\Pi = (\delta, \Sigma, C)$  be an LCL problem. Label  $\sigma \in \Sigma(\Pi)$  has a *continuation below with labels in  $\Sigma' \subseteq \Sigma$*  if there exists a configuration  $(\sigma : \sigma_1, \dots, \sigma_\delta) \in C(\Pi)$  such that  $\{\sigma, \sigma_1, \dots, \sigma_\delta\} \subseteq \Sigma'$ .

**Definition 4.6** (*path-form of an LCL problem*) Let  $\Pi = (\delta, \Sigma, C)$  be an LCL problem. The *path-form* of  $\Pi$  is the LCL problem  $\Pi^{\text{path}} = (1, \Sigma, C')$ , where  $(a : b) \in C'$  if and only if there exists a configuration  $(a : b_1, b_2, \dots, b_\delta) \in C$  with  $b = b_i$  for some  $i$ .

See Fig. 2b for an illustration of the path-form.

## 4.4 Automata and flexibility

**Definition 4.7** (*automaton associated with path-form of an LCL problem; [17]*) Let  $\Pi$  be an LCL problem. The automaton  $\mathcal{M}(\Pi)$  associated with the path-form of  $\Pi$  is a nondeterministic unary semiautomaton defined as follows:

- The set of states is  $\Sigma(\Pi)$ .
- There is a transition from state  $a$  to state  $b$  if there is a configuration  $(a : b)$  in the path-form  $\Pi^{\text{path}}$  of  $\Pi$ .

See Fig. 2c for an illustration of the automaton.

**Definition 4.8** (*flexible state of an automaton*; [17]) A state  $a$  from  $\mathcal{M}(\Pi)$  is *flexible* if there is a natural number  $K$  such that for all  $k \geq K$  there is a walk  $a \rightsquigarrow a$  of length exactly  $k$  in  $\mathcal{M}(\Pi)$ . The smallest number  $K$  that satisfies this property is the *flexibility* of state  $a$ , in notation  $\text{flexibility}(a)$ .

As the set of states of the automaton is the set of labels, we can expand the notion of flexibility of a state to the notion of flexibility of a label.

**Definition 4.9** (*path-flexibility*) Let  $\Pi$  be an LCL problem and  $\Pi^{\text{path}}$  its path-form. A label  $\sigma \in \Sigma(\Pi)$  is *path-flexible* if  $\sigma$  is a flexible state in automaton  $\mathcal{M}(\Pi)$ , and *path-inflexible* otherwise.

Moreover, an LCL problem  $\Pi$  is *path-flexible* if all labels are path-flexible labels and its automaton  $\mathcal{M}(\Pi)$  has one strongly connected component.

### 4.5 Graph-theoretic definitions

**Definition 4.10** (*root-to-leaf path*) A *root-to-leaf* path in a tree is a path that starts at the root and ends at one of its leaves.

**Definition 4.11** (*hairy path*) A full  $\delta$ -ary tree  $T$  is called a *hairy path* if it can be obtained by attaching leaves to a directed path such that all nodes of the path have exactly  $\delta$  children.

**Definition 4.12** (*minimal absorbing subgraph*) Let  $G$  be a directed graph. A subgraph  $G' \subseteq G$ , is called a *minimal absorbing subgraph* if  $G'$  is a strongly connected component of  $G$  and  $G'$  does not have any outgoing edges.

We note that a minimal absorbing subgraph exists for any directed graph.

**Definition 4.13** (*ruling set*) Let  $G$  be a graph. A  $(k, l)$ -*ruling set* is a subset  $S$  of nodes of  $G$  such that the distance between any two nodes in  $S$  is at least  $k$ , and the distance between any node in  $G$  and the closest node in  $S$  is at most  $l$ .

## 5 Super-logarithmic region

In this section we prove that there is no LCL problem  $\Pi$  with distributed time complexity between  $\omega(\log n)$  and  $n^{o(1)}$ . Also, we prove that, given a problem  $\Pi$ , we can *decide* if its complexity is  $O(\log n)$  or  $n^{\Omega(1)}$ . In view of [14,16], randomness does not help for LCL problems with round complexity  $\Omega(\log n)$ , so we focus on the deterministic setting.

### 5.1 High-level idea

The key idea is that we *iteratively prune* the description of problem  $\Pi$  by removing subsets of labels that we call *path-inflexible*—these are sets of labels that require long-distance coordination (cf. 2-coloring). After each such step, we may arrive at a subproblem that contains a new path-inflexible set, but eventually the pruning process will terminate, as there is only a finite number of labels.

Assume the pruning process terminates after  $k$  steps. Let  $X_1, X_2, \dots, X_k$  be the sets of labels we removed during the process, and let  $X'$  be the set of labels that is left after no path-inflexible labels remain. We have two cases:

- (1) Set  $X'$  is empty. In this case we can show that the round complexity of the problem  $\Pi$  is at least  $\Omega(n^{1/k})$ . To prove this, we make use of a  $k$ -level construction that generalizes the one used for  $2\frac{1}{2}$ -coloring in [16]. We argue that, roughly speaking, for  $o(n^{1/k})$ -round algorithms, no label from set  $X_i$  can be used for labeling the level- $j$  nodes, for each  $j \geq i$ , as this requires coordination over distance  $\Theta(n^{1/k})$ .
- (2) Set  $X'$  is non-empty. In this case, after removing the sets  $X_i$ , we are left with a non-empty *path-flexible* subproblem  $\Pi' \subseteq \Pi$ , and we can make use of the flexibility to solve  $\Pi'$  in  $O(\log n)$  rounds. Hence the original problem  $\Pi$  is also solvable in  $O(\log n)$  rounds.

We say that problem  $\Pi$  has a *certificate for  $O(\log n)$  solvability* if and only if the set  $X'$  is non-empty.

Formally, we prove the following theorems.

**Theorem 5.1** *Let  $\Pi$  be a problem having a certificate for  $O(\log n)$  solvability. Then  $\Pi$  is solvable in  $O(\log n)$  rounds in the CONGEST model.*

**Theorem 5.2** *Let  $\Pi$  be an LCL problem having no certificate for  $O(\log n)$  solvability. Then both the randomized and the deterministic complexity of  $\Pi$  in the LOCAL model are  $\Omega(n^{1/k})$  for some  $k \geq 1$ .*

**Theorem 5.3** *Whether an LCL problem  $\Pi$  has round complexity  $O(\log n)$  or  $n^{\Omega(1)}$  can be decided in polynomial time.*

### 5.2 Certificate

We present an algorithm that decides whether the complexity of a given problem  $\Pi$  is  $O(\log n)$  or  $n^{\Omega(1)}$  rounds.

We start by defining a procedure that, given a problem  $\Pi$ , creates its restriction  $\Pi'$  to path-flexible states of  $\Pi$ ; see Algorithm 1. However, note that states that were path-flexible in  $\Pi$  may become path-inflexible in  $\Pi'$ ; hence problem  $\Pi'$  may still contain path-inflexible states.

**Algorithm 1:** removePathInflexibleConfigurations( $\Pi$ )

**Input:** LCL problem  $\Pi$   
**Output:** LCL problem  $\Pi'$ , a restriction of  $\Pi$  to its path-flexible states

Construct  $\Pi^{\text{path}}$ , the path-form of  $\Pi$ .  $\triangleright$  See Definition 4.6.  
 Construct the automaton  $\mathcal{M}(\Pi)$ .  $\triangleright$  See Definition 4.7.  
 $\Sigma' \leftarrow$  the set of path-flexible states of  $\Pi$ .  $\triangleright$  See Definition 4.9.  
 $\Pi' \leftarrow$  the restriction of  $\Pi$  to labels  $\Sigma'$ .  $\triangleright$  See Definition 4.3.  
**return**  $\Pi'$

Next we describe a new procedure findLogCertificate that uses Algorithm 1 to analyze the complexity of a given problem. This procedure either returns  $\epsilon$  to indicate that the problem requires  $n^{\Omega(1)}$  rounds, or it returns a new problem  $\Pi_{\text{pf}}$  that is a restriction of  $\Pi$ , but that will be nevertheless solvable in  $O(\log n)$  rounds (and therefore  $\Pi$  is also solvable in this time).

Informally, procedure findLogCertificate applies iteratively Algorithm 1 until one of the following happens:

- We obtain an empty problem. In this case we return  $\epsilon$ . We will show that this can only happen if  $\Pi$  requires  $n^{\Omega(1)}$  rounds.
- We reach a non-empty fixed point  $\Pi_i$ . In this case we further restrict  $\Pi_i$  to the labels that induce a minimal absorbing subgraph in the automaton associated with its path-form. Let  $\Pi_{\text{pf}}$  be the problem constructed this way. We return  $\Pi_{\text{pf}}$ , and we say that  $\Pi_{\text{pf}}$  is the *certificate for  $O(\log n)$  solvability*. We will show that  $\Pi_{\text{pf}}$  and hence also the original problem  $\Pi$  can be solved in  $O(\log n)$  rounds. Note that a minimal absorbing subgraph has the property that any labeling of the two endpoints of a sufficiently long path with labels from the subgraph admits an extension of the solution to the entire path with labels from the subgraph. This provides the intuition why reducing the labels to those of a minimal absorbing subgraph allows for an  $O(\log n)$ -round algorithm using the rake-and-compress approach explained in Sect. 5.3.

The procedure is described more formally in Algorithm 2, and an example of execution for a concrete problem can be seen in Fig. 2.

Now, let us prove some of the properties of Algorithm 2. First, we observe that this is indeed a polynomial-time algorithm.

**Lemma 5.4** *Algorithm 2 runs in polynomial time in the size of the description of  $\Pi$ .*

**Proof** When creating a successive restrictions of  $\Pi$  in Algorithm 2, we always remove at least one label. Hence we invoke Algorithm 1 at most  $|\Sigma(\Pi)|$ , and Algorithm 1 runs in polynomial time [17].  $\square$

**Algorithm 2:** findLogCertificate( $\Pi$ )

**Input:** LCL problem  $\Pi$   
**Output:** Certificate for  $O(\log n)$  solvability if it exists, or  $\epsilon$  otherwise

$\Pi_0 \leftarrow \Pi$   
 $i \leftarrow 0$   
**repeat**  
      $\Pi_{i+1} \leftarrow$  removePathInflexibleConfigurations( $\Pi_i$ )  $\triangleright$  See Algorithm 1.  
      $i \leftarrow i + 1$   
**until**  $\Pi_i = \Pi_{i-1}$   
**if**  $\Pi_i$  is empty **then**  
     **return**  $\epsilon$   $\triangleright$   $\Pi$  cannot be solved in  $O(\log n)$  rounds.  
**else**  
      $\Sigma' \leftarrow$  labels that induce a minimal absorbing subgraph of automaton  $\mathcal{M}(\Pi_i)$   $\triangleright$  See Definition 4.12.  
      $\Pi_{\text{pf}} \leftarrow$  the restriction of  $\Pi_i$  to  $\Sigma'$   
     **return**  $\Pi_{\text{pf}}$   $\triangleright$  Certificate for  $O(\log n)$  solvability.  
**end**

Then we prove that the step where we restrict to a minimal absorbing subgraph behaves well; in particular, it will preserve flexibility.

**Lemma 5.5** *Let  $\Pi$  be a non-empty LCL problem, such that all of its states are path-flexible. Let  $\Sigma'$  be a set of labels that induces a minimal absorbing subgraph of automaton  $\mathcal{M}(\Pi)$ , and let  $\Pi_{\text{pf}}$  be the restriction of  $\Pi$  to labels  $\Sigma'$ . Then all states of  $\Pi_{\text{pf}}$  are flexible, there is a walk between any two states of  $\mathcal{M}(\Pi_{\text{pf}})$ , and  $\mathcal{M}(\Pi_{\text{pf}})$  has at least one edge.*

**Proof** First, let us prove that the restriction will preserve the flexibility of the states that remain. Since for every state in a minimal absorbing subgraph all outgoing edges are connected to states in the same minimal absorbing subgraph by definition, then no configuration for these states will be removed, and all returning walks for a state will stay. Second, a walk between any two states of  $\mathcal{M}(\Pi_{\text{pf}})$  is implied by the fact that  $\mathcal{M}(\Pi_{\text{pf}})$  is strongly connected. Lastly,  $\mathcal{M}(\Pi_{\text{pf}})$  has at least one edge, as every node has returning walks, hence incoming and outgoing edges, and the minimal absorbing subgraph is non-empty.  $\square$

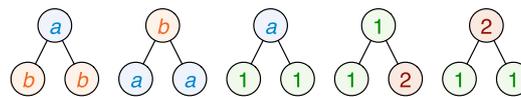
In the rest of the section, we will prove that our certificate for  $O(\log n)$  solvability indeed characterizes  $O(\log n)$  solvability in the following sense: if  $\Pi$  has a certificate for  $O(\log n)$  solvability, then  $\Pi$  can be solved in  $O(\log n)$  rounds, otherwise there is an  $n^{\Omega(1)}$  lower bound for  $\Pi$ . Hence Lemma 5.4 implies Theorem 5.3.

**5.3 Upper bound**

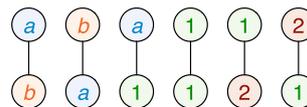
We prove that, if Algorithm 2 does not return  $\epsilon$ , then the original problem  $\Pi$  can be solved in  $O(\log n)$  rounds. Note that  $\Pi_{\text{pf}}$ , the result of Algorithm 2, is obtained by considering a subset of labels of  $\Pi$  and all constraints that use only

**Fig. 2** Certifying that problem  $\Pi_0$  is solvable in  $O(\log n)$  rounds (see Algorithm 2). The sample problem  $\Pi_0$  is a combination of the branch 2-coloring problem (5), using labels 1 and 2, and the normal 2-coloring problem (2), using labels  $a$  and  $b$ . The grayed out states in the automaton denote inflexible states. Problem  $\Pi_{pf}$  is the *path-flexible* form of problem  $\Pi_0$

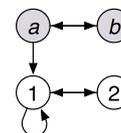
(a) Problem  $\Pi_0$ :



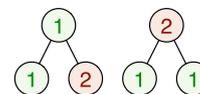
(b) Path-form  $\Pi_0^{path}$  of  $\Pi_0$ :



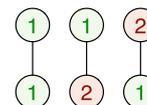
(c) Automaton  $\mathcal{M}(\Pi_0)$ :



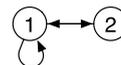
(d) Problem  $\Pi_1$ :



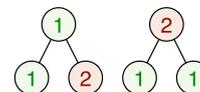
(e) Path-form  $\Pi_1^{path}$  of  $\Pi_1$ :



(f) Automaton  $\mathcal{M}(\Pi_1)$ :



(g) Problem  $\Pi_{pf}$ :



those labels, hence a solution for  $\Pi_{pf}$  is also a valid solution for  $\Pi$ . Hence, we prove our claim by providing an algorithm solving  $\Pi_{pf}$  in  $O(\log n)$  rounds. For this purpose, we start by providing a procedure that is a modified version of the rake and compress procedures of Miller and Reif [22], where, informally, we remove degree-2 nodes only if they are contained in long enough paths. We start with some definitions. Note that in a rooted tree we assume that each edge  $\{u, v\}$  is oriented from  $u$  to  $v$  if  $v$  is the parent of  $u$ .

**Definition 5.6 (leaves)** Let  $G = (V, E)$  be a graph. We define that  $leaves(G) \subseteq V$  is the set of all nodes with indegree 0.

**Definition 5.7 (long-paths)** Let  $G = (V, E)$  be a graph and  $p$  be a constant. Let  $X \subseteq V$  consist of all nodes of indegree 1. We define that  $long-path-nodes(G, p) \subseteq X$  consists of the

set of all nodes that belong to a connected component of size at least  $p$  in the subgraph of  $G$  induced by  $X$ .

We now define our variant of the rake-and-compress procedure. Note that a similar variant, for unrooted trees, appeared in [2].

**Definition 5.8 (RCP)** Let  $p \in \{1, 2, \dots\}$ . Procedure  $RCP(p)$  iteratively partitions the set of nodes  $V$  into non-empty sets  $V_1, V_2, \dots, V_L$  for some  $L$  as follows:

$$G_0 = G, \quad V_{i+1} = leaves(G_i) \cup long-path-nodes(G_i, p) \\ G_{i+1} = G_i \setminus V_{i+1}.$$

Note that the graphs  $G_i$  can be disconnected.

We now prove an upper bound on the highest possible layer obtained by the procedure. In particular, we prove that there is some layer  $L = O(\log n)$  such that  $G_{L+1}$  is empty.

For this purpose, we now prove that the number of nodes of  $G_{i+1}$  is at least a  $(1 - \frac{1}{6p})$  factor smaller than the number of nodes of  $G_i$ , implying that after  $O(\log n)$  steps we obtain an empty graph.

**Lemma 5.9** *Let  $p$  be a constant and let  $G = (V, E)$  be a tree with  $n$  nodes. At least one of the following holds:*

$$|\text{leaves}(G)| \geq \frac{n}{6p} \quad \text{or} \quad |\text{long-path-nodes}(G, p)| \geq \frac{n}{3}.$$

**Proof** Let  $n_0, n_1, n_{2+}$  be the number of nodes of indegree 0, 1, and 2 or more, respectively. We have  $n = n_0 + n_1 + n_{2+}$ . The number of edges in a tree is  $m = n - 1$  or by counting using indegrees we get  $m \geq 0n_0 + 1n_1 + 2n_{2+}$ , from which we obtain  $n_0 + n_1 + n_{2+} - 1 \geq 0n_0 + 1n_1 + 2n_{2+}$ . Hence  $n_{2+} < n_0$ . We have  $|\text{leaves}(G)| = n_0$  so if  $n_0 \geq n/(6p)$ , the claim holds. In what follows, we assume that  $n_0 < n/(6p)$  which together with  $n_{2+} < n_0$  implies that  $n_0 + n_{2+} < n/(3p)$ . This implies that the total number of nodes of indegree 1 nodes is  $n_1 = n - (n_0 + n_{2+}) > n - n/(3p) \geq 2n/3$ . Consider the subgraph  $G_1$  induced by indegree-1 nodes of  $G$ . If we contract each connected component of  $G_1$  into an edge, we obtain a tree  $G'$  in which we have  $n' = n_{2+} + n_0$  nodes and  $m' = n_{2+} + n_0 - 1$  edges. As each edge represents at most one connected component of  $G_1$ , there are fewer than  $n/(3p)$  components in  $G_1$ . Hence we have  $n_1 > 2n/3$  indegree-1 nodes that are contained in less than  $n/(3p)$  connected components. Since components of size less than  $p$  can contain at most  $p \cdot n/(3p) = n/3$  nodes in total, then there have to be at least  $2n/3 - n/3 = n/3$  nodes in the components of size at least  $p$ , hence  $|\text{long-path-nodes}(G, p)| \geq n/3$ .  $\square$

We now prove an upper bound on the time required for all nodes  $v$  to know the layer  $i$  in which they belong, that is, the layer  $i$  satisfying  $v \in V_i$ .

**Lemma 5.10** *RCP( $p$ ) can be computed in  $O(\log n)$  rounds in the LOCAL and CONGEST models.*

**Proof** We build a virtual graph where we iteratively remove nodes for  $O(\log n)$  rounds. At each step, nodes can check in 1 round which neighbors have already been removed, and hence compute their indegree in the virtual graph. Nodes mark themselves as removed if their indegree is 0, or if their indegree is 1 and they are in paths of length at least  $p$ . The result of each node is the step in which they have been marked as removed. Notice that each step requires  $O(p)$  rounds, even in CONGEST, and since  $p$  is a constant, this procedure requires  $O(\log n)$  rounds in total.  $\square$

We are now ready to prove that, if Algorithm 2 returns some problem  $\Pi_{\text{pf}}$ , then  $\Pi_{\text{pf}}$  (and hence  $\Pi$ ) can be solved in  $O(\log n)$  rounds, proving Theorem 5.1.

**Proof of Theorem 5.1** Let  $\Pi$  be a problem having a certificate for  $O(\log n)$  solvability. Then we will show that  $\Pi$  is  $O(\log n)$  solvable in the CONGEST model.

Let  $\Pi_{\text{pf}}$  be the path-flexible form from Algorithm 2. Let

$$k = \max_{\sigma \in \Sigma(\Pi_{\text{pf}})} (\text{flexibility}(\sigma)) + |\Sigma(\Pi_{\text{pf}})|,$$

where  $\text{flexibility}(\sigma)$  is the flexibility of a state  $\sigma$  in  $\mathcal{M}(\Pi_{\text{pf}})$  as defined in Algorithm 4.8.

Given a tree  $T$ , we solve  $\Pi_{\text{pf}}$  as follows. We start by running the procedure RCP( $k$ ) on  $T$ . After this process, each node  $v$  knows the set  $V_i$ ,  $1 \leq i \leq L$ , which it belongs to. Then, we compute a distance- $k$  coloring by using a palette of  $O(1)$  colors, which can be done in  $O(\log^* n)$  rounds, even in CONGEST, since  $\delta$  is constant (using, e.g., Linial's algorithm [21] on power graphs).

We then process the layers one by one, from layer  $L$  to 1. For each layer  $i$ , we label all (unlabeled) nodes in  $V_i$  and all of their children. We need to deal with two cases, either we are labeling a long path or we are labeling a leaf node (both in  $G_{i-1}$ ).

If a node  $v_j \in V_i$  is a leaf node, then by definition its children were not processed yet so they are not labeled. Node  $v_j$  could be labeled with its parent, or it is unlabeled. But in both cases, we can complete this labeling (by labeling the descendants of  $v_j$  and possibly  $v_j$  itself) as every label has a continuation below.

If we need to label a long path  $P$ , then by construction all inner nodes have no fixed labels so far. The topmost node can be labeled (as we may have already processed its parent) and the bottom-most node has indegree one, and thus it is connected to exactly one node from an upper layer, and hence it will have exactly one child already labeled. To label all nodes of  $P$ , we proceed in several steps. First, we exploit the precomputed distance- $k$  coloring to compute a  $(k, k)$ -ruling set on each path in parallel in constant time, by iterating through the constantly many color classes and adding to the ruling set all nodes of the processed color for which no node in distance at most  $k - 1$  is already contained in the ruling set. The ruling set nodes split the path into constant-length chunks. Next, for each endpoint of the path, we remove the closest ruling set node from the ruling set. This ensures that all chunks are of length at least  $k$ . Then, we label all nodes that still remain in the ruling set with an arbitrarily chosen label from  $\Sigma(\Pi_{\text{pf}})$ . Finally, we label the nodes in the constant-length chunks (and their children) in a consistent manner. This is possible since each label used for the ruling set nodes is flexible and has a walk to any other label in  $\mathcal{M}(\Pi_{\text{pf}})$  (as proved by Lemma 5.5) and the ruling set nodes are far enough apart (more than the flexibility of any label in  $\Pi_{\text{pf}}$ ).

As all of these steps can be performed in constant time (provided the precomputed distance- $k$  coloring), we can

label the whole tree in  $L \cdot O(1) + O(\log^* n) = O(\log n)$  rounds.  $\square$

### 5.4 Lower bound

We prove that if Algorithm 2 returns  $\epsilon$ , then the original problem  $\Pi = (\delta, \Sigma, C)$  requires  $n^{\Omega(1)}$  rounds to solve.

*A sequence of labels.* If Algorithm 2 returns  $\epsilon$  after  $k$  iterations, then there is a sequence  $\Sigma_1, \Sigma_2, \dots, \Sigma_k$  of sets of labels meeting the following conditions and leading to a sequence  $\Pi_0, \Pi_1, \Pi_2, \dots, \Pi_k$  of LCL problems:

- $\Pi_0 = \Pi$ .
- For  $1 \leq i \leq k$ ,  $\Pi_i$  is the LCL problem that is the restriction of  $\Pi_{i-1}$  to the label set  $\Sigma(\Pi_{i-1}) \setminus \Sigma_i$ , or equivalently,  $\Pi_i$  is the restriction of the original LCL problem  $\Pi$  to the label set  $\Sigma \setminus (\Sigma_1 \cup \Sigma_2 \cup \dots \cup \Sigma_i)$ .
- For  $1 \leq i \leq k$ ,  $\Sigma_i$  is the set of path-inflexible labels in  $\Pi_{i-1}$ .
- $\Sigma_{\Pi_k} = \emptyset$ , so  $\Sigma = \Sigma_1 \cup \Sigma_2 \cup \dots \cup \Sigma_k$  is a partition of  $\Sigma$ .

The set of labels  $\Sigma_i \subseteq \Sigma$  consists of the labels removed during the  $i$ th iteration of Algorithm 2, as they are path-inflexible in  $\Pi_{i-1}$ . As Algorithm 2 returns  $\epsilon$  after  $k$  iterations,  $\Sigma = \Sigma_1 \cup \Sigma_2 \cup \dots \cup \Sigma_k$  is a partition of  $\Sigma$ . The goal of this section is to show that solving  $\Pi$  requires  $\Omega(n^{1/k})$  rounds.

*Centered graphs.* A radius- $t$  centered graph is a pair  $(G, v)$  where  $v \in V$  is a node in  $G = (V, E)$  so that all  $u \in V$  are within distance  $t$  to  $v$ , and each  $u \in V$  whose distance to  $v$  is exactly  $t$  is permitted to have incident edges of the form  $e = \{u, ?\}$ , indicating that  $e$  is an external edge that connects  $u$  to some unknown node outside of  $G$ . As we only consider rooted trees, we assume that all edges are oriented towards the root, so that each node has outdegree at most 1.

Observe that the view of a node  $v$  after  $t$  rounds of communication in LOCAL can be described by a radius- $t$  centered graph. Therefore, a  $t$ -round LOCAL algorithm on  $n$ -node graphs is simply an assignment of a label  $\sigma \in \Sigma$  to each radius- $t$  centered graph  $(G, v)$  where each node in  $G$  has a distinct  $O(\log n)$ -bit identifier.

*Terminology.* In this section, we use the term radius- $t$  view of  $v$  to denote the corresponding radius- $t$  centered graph, and the term radius- $t$  neighborhood of  $v$  to denote the set of nodes that are within distance  $t$  to  $v$ . Note that the radius- $t$  view of  $v$  contains more information than the subgraph induced by the radius- $t$  neighborhood of  $v$ , as the radius- $t$  view of  $v$  includes information about the external edges.

*Permissible labels.* From now on, we fix  $A$  to be any LOCAL algorithm that solves  $\Pi = (\delta, \Sigma, C)$  in  $t$  rounds on  $n$ -node graphs. Given such an algorithm  $A$ , we say that a label  $\sigma \in \Sigma$  is permissible for  $(G, v)$  if there exists some assignment of

distinct  $O(\log n)$ -bit identifiers to the nodes in  $G$  such that the output of  $v$  is  $\sigma$  when we run  $A$  on  $G$ .

Using the notion of permissible labels, to show that  $\Pi = (\delta, \Sigma, C)$  cannot be solved in  $t$  rounds on  $n$ -node graphs, it suffices to find a graph  $G = (V, E)$  that has at most  $n$  nodes such that there exists a node  $v \in V$  such that no label  $\sigma \in \Sigma$  is permissible for the radius- $t$  centered graph  $(H, v)$  corresponding to the radius- $t$  view of  $v$  in  $G$ . The following lemma is useful for showing that some label  $\sigma \in \Sigma$  is not permissible for some radius- $t$  centered graph  $(G, v)$ .

**Lemma 5.11** *Let  $A$  be a  $t$ -round LOCAL algorithm that solves  $\Pi = (\delta, \Sigma, C)$  for  $n$ -node rooted trees. Let  $(G, v)$  be a fixed radius- $t$  centered graph. Let  $\Sigma^\diamond \subseteq \Sigma$  be a subset of labels, and let  $\Pi'$  be the restriction of  $\Pi$  to  $\Sigma \setminus \Sigma^\diamond$ . Suppose there exists a number  $K$  such that for any  $d \geq K$ , we can construct a rooted tree  $T$  containing a directed path  $P = v_1 \leftarrow v_2 \leftarrow \dots \leftarrow v_{d+1}$  meeting the following conditions.*

- (1) *The radius- $t$  views of  $v_1$  and  $v_{d+1}$  are isomorphic to  $(G, v)$ .*
- (2) *Let  $S$  denote the set of nodes  $v_1, v_2, \dots, v_{d+1}$  and their children. Then each  $\sigma \in \Sigma^\diamond$  is not permissible for the radius- $t$  views of each  $u \in S$ .*
- (3) *The radius- $(t+2)$  neighborhood of  $v_i$  contains at most  $n$  nodes, for each  $1 \leq i \leq d+1$ .*

*Then, for each  $\sigma \in \Sigma \setminus \Sigma^\diamond$ , the following holds: if  $\sigma$  is path-inflexible in  $\Pi'$ , then  $\sigma$  is not permissible for  $(G, v)$ .*

See Fig. 3 for an illustration of Lemma 5.11. Before proving Lemma 5.11, let us give a brief, informal example of how we might apply it. We assume we have already established that algorithm  $A$  cannot output labels from  $\Sigma^\diamond$  in certain “tricky” radius- $t$  views. Tree  $T$  is then constructed so that nodes of  $S$  have tricky views, so algorithm  $A$  is forced to solve the restriction  $\Pi'$  of  $\Pi$  around path  $P$ . Now if  $\Pi'$  contains some path-inflexible labels, we can apply Lemma 5.11 to rule out the possibility of using path-inflexible labels along path  $P$ , so we learn that the view  $(G, v)$  is super-tricky, as it rules out not only  $\Sigma^\diamond$ , but also all path-inflexible labels of  $\Pi'$ . We can repeat this argument to discover many super-tricky views, by constructing different trees  $T$ .

This way we can start with a problem  $\Pi_0$ , and rule out the use of path-inflexible labels of  $\Pi_0$  at least in some family of tricky views. Hence in those views we are, in essence, solving problem  $\Pi_1$ , which is the restriction of  $\Pi_0$  to path-flexible labels. We repeat the argument, and rule out the use of path-inflexible labels of  $\Pi_1$  in at least some family of super-tricky views, etc.

If we eventually arrive at an empty problem, we have reached a contradiction: algorithm  $A$  cannot solve the original problem in some family of particularly tricky views. However, plenty of care will be needed to keep track of the

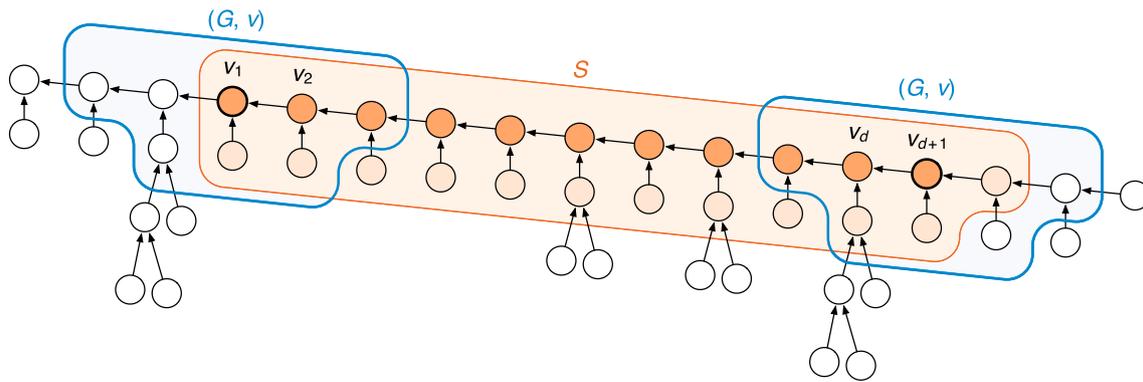


Fig. 3 An illustration of Lemma 5.11 with  $t = 2$

specific family of views, as well as to make sure that we can still construct a suitable tree  $T$  using only such views. We will get back to these soon, but this informal introduction will hopefully help to see why we first seek to prove this somewhat technical statement.

**Proof of Lemma 5.11** Fix a label  $\sigma \in \Sigma \setminus \Sigma^\diamond$  such that  $\sigma$  is path-inflexible in  $\Pi'$ . By the definition of path-flexibility, for any  $K$ , there exists an integer  $d \geq K$  such that the following statement holds:

- For any length- $d$  directed path  $P = v_1 \leftarrow v_2 \leftarrow \dots \leftarrow v_{d+1}$  such that each node  $v_i$  in  $P$  is assigned a label  $\lambda(v_i) \in \Sigma \setminus \Sigma^\diamond$ , if the two end points  $v_1$  and  $v_{d+1}$  are labeled with  $\sigma$ , then the labeling of  $P$ , interpreted as 1-ary tree, is not a valid solution for the path-form of  $\Pi'$ . More precisely, there must exist  $1 \leq i \leq d + 1$  such that  $(\lambda(v_i) : \lambda(v_{i+1}))$  is not an allowed configuration in the path-form of  $\Pi'$ .

For the rest of the proof, we pick  $d$  be a sufficiently large number such that the above statement holds. The precise choice of  $d$  is to be determined. We consider a rooted tree  $T$  that satisfies the lemma statement for this parameter  $d$ . We assume that  $\sigma$  is permissible for  $(G, v)$ , and then we will derive a contradiction.

Now consider the path  $P$  in the lemma statement. Here all nodes in  $P$  have exactly  $\delta$  children, and all nodes of  $P$  and their children can only be assigned labels from  $\Sigma \setminus \Sigma^\diamond$  by  $A$ . Again, if  $v_1$  and  $v_{d+1}$  are labeled with  $\sigma$ , then there must exist  $1 \leq i \leq d + 1$  such that the node configuration of  $v_i$  and its  $\delta$  children is not an allowed configuration of  $\Pi'$ . Furthermore, it cannot be an allowed configuration for  $\Pi$ , either, as  $\Pi'$  contains all configurations that consist of labels from  $\Sigma \setminus \Sigma^\diamond$ .

Consider any assignment of  $O(\log n)$ -bit identifiers to the nodes in  $(G, v)$  that makes  $A$  output  $\sigma$ , and apply this assignment to the radius- $t$  neighborhoods of  $v_1$  and  $v_{d+1}$  in  $T$ . Extend this identifier assignment to cover all nodes that are

within distance  $t + 2$  to some  $v_i$  such that the radius- $(t + 2)$  neighborhood of any  $v_i$  does not contain repeated identifiers. This is possible because we assume that  $T$  satisfies the property that the radius- $(t + 2)$  neighborhood of  $v_i$  contains at most  $n$  nodes, for each  $1 \leq i \leq d + 1$ , and because that we may choose  $d \gg t$  to be sufficiently large so that the radius- $(t + 2)$  neighborhood of each  $u \in S$  cannot simultaneously intersect the radius- $(t + 2)$  neighborhood of both  $v_1$  and  $v_{d+1}$ . Although some identifiers may appear several times in  $T$  and the total number of nodes in  $T$  may exceed  $n$ . As we will later see, they are not problematic.

Consider the output labels of  $v_i$  and its children, for  $1 \leq i \leq d + 1$ , resulting from simulating  $A$  on  $T$ . Note that  $A$  by definition cannot output labels that are not permissible, and hence all of these nodes receive labels from  $\Sigma \setminus \Sigma^\diamond$ . Our choice of  $d$  implies that there exists  $1 \leq i \leq d + 1$  such that the node configuration corresponding to  $v_i$  and its children is not in  $C$ . We take the subtree  $T'$  induced by the union of the radius- $(t + 1)$  neighborhood of  $v_i$  and its children. Since the radius- $(t + 2)$  neighborhood of  $v_i$  contains at most  $n$  nodes, the rooted tree  $T'$  also contains at most  $n$  nodes. The output labelings of  $v_i$  and its children due to simulating  $A$  are the same in both  $T$  and  $T'$ , as their radius- $t$  views are invariant of the underlying network being  $T$  or  $T'$ . This violates the correctness of  $A$ , as  $T'$  contains at most  $n$  nodes. Thus,  $\sigma$  cannot be permissible for  $(G, v)$ .  $\square$

*A hierarchical construction of rooted trees.* We first consider the following natural recursive construction of rooted trees. A *bipolar tree* is a tree with two distinguished nodes  $s$  and  $t$ , and it is also viewed as a rooted tree by setting  $s$  as the root. The unique path connecting  $s$  and  $t$  is called the *core path* of the bipolar tree. We consider the following operation  $\oplus^x$ .

- Given a rooted tree  $T$ , define  $\oplus^x T$  as the result of the following construction. Start with an  $x$ -node path  $(v_1, v_2, \dots, v_x)$ . Consider  $x(\delta - 1)$  copies of  $T$ , indexed

by two numbers  $i$  and  $j$ :

$$\{T'_{i,j}\}_{1 \leq i \leq x, 1 \leq j \leq \delta - 1}.$$

For  $1 \leq i \leq x$  and  $1 \leq j \leq \delta - 1$ , make the root of  $T'_{i,j}$  a child of  $v_i$  by adding an edge connecting them. Finally, set the two distinguished nodes of the resulting tree by  $s = v_1$  and  $t = v_x$ .

Based on this operation, we construct a sequence of bipolar trees  $T_0^x, T_1^x, \dots, T_k^x$ , where the nodes in  $T_i^x$  are partitioned into layers  $0, 1, \dots, i$ ; see Fig. 4:

- For  $i = 0$ , define  $T_0^x$  as the trivial bipolar tree consisting of only one isolated node  $v$  with  $s = v$  and  $t = v$ . We say that  $v$  is in layer 0.
- For  $1 \leq i \leq k$ , define  $T_i^x = \bigoplus^x T_{i-1}^x$ . We say that all nodes in the core path of  $T_i^x$  are in layer  $i$ .

For any constant  $\delta$ , it is straightforward to see that the number of nodes in  $T_k^x$  is  $n = O(x^k)$ , so  $x = \Omega(n^{1/k})$ . For each  $1 \leq i \leq k$ , the layer- $i$  nodes form paths consisting of exactly  $x$  nodes. We call such an  $x$ -node path a layer- $i$  path.

The tree  $T_k^x$  is analogous to the lower bound graph used in [16] for establishing the tight  $\Omega(n^{1/k})$  lower bound for some artificial LCL problem considered in [16]. The  $\Omega(n^{1/k})$  lower bound proof in [16] involves an argument showing that to solve the given LCL problem it is necessary that the two endpoints of a layer- $i$  path communicate with each other, and this costs at least  $x = \Omega(n^{1/k})$  rounds.

When  $x \geq 2$  and  $1 \leq j \leq k$ , there are three possible degrees in  $T_j^x$ :  $1$ ,  $\Delta - 1 = \delta$ , and  $\Delta = \delta + 1$ . A node has degree 1 if and only if it is in layer zero. A node has degree  $\delta$  if and only if it is the root or it is the last node  $v_x$  in some layer- $i$  path  $v_1 \leftarrow v_2 \leftarrow \dots \leftarrow v_x$ . All the remaining nodes have degree exactly  $\Delta = \delta + 1$ . Intuitively, the nodes with degree  $\delta$  are in the boundary of the graph.

*High-level ideas.* To prove the  $\Omega(n^{1/k})$  lower bound, we will construct a sequence  $S_1 \supseteq S_2 \supseteq \dots \supseteq S_k$  of non-empty sets of radius- $t$  centered graphs, where  $t = \Omega(n^{1/k})$ . Each radius- $t$  centered graph used in our construction is isomorphic to the radius- $t$  view of some node  $v$  in some graph of at most  $n$  nodes. By applying Lemma 5.11 inductively with  $\Sigma^\diamond = \Sigma_1 \cup \Sigma_2 \cup \dots \cup \Sigma_{i-1}$  and  $\sigma \in \Sigma_i$ , we will show that for any given  $t$ -round algorithm  $A$ , the labels in  $\Sigma_i$  are not permissible for the radius- $t$  centered graphs in  $S_i$ , for each  $1 \leq i \leq k$ . Therefore, the given LCL problem  $\Pi = (\delta, \Sigma, C)$  cannot be solved in  $t$  rounds.

The construction of  $S_1 \supseteq S_2 \supseteq \dots \supseteq S_k$  requires somewhat complicated definitions. To motivate these forthcoming definitions, we begin with describing a natural attempt to prove the  $\Omega(n^{1/k})$  lower bound directly using the trees  $T_k^x$  and see why it does not work.

Suppose that the given LCL problem  $\Pi = (\delta, \Sigma, C)$  can be solved in  $t = o(n^{1/k})$  rounds on  $n$ -node graphs by an algorithm  $A$ . We pick  $x$  to be a sufficiently large number such that  $x = \Theta(t)$  and the number of nodes in  $T_k^x$  is at most  $O(x^k) < n$ . Recall that  $\Sigma_1$  is the set of path-inflexible labels for the original LCL problem  $\Pi = (\delta, \Sigma, C)$ . Let  $P = v_1 \leftarrow v_2 \leftarrow \dots \leftarrow v_x$  be a layer-1 path in  $T_k^x$ . It is straightforward to see that for each  $1 + t < j < x - t$ , the radius- $t$  view of each  $v_j$  is identical. Let  $(G, v)$  denote the corresponding radius- $t$  centered subgraph. By the path inflexibility of the labels in  $\Sigma_1$ , all labels  $\sigma \in \Sigma_1$  are not permissible for  $(G, v)$ . Intuitively, this is because that we can find in  $T_k^x$  a path connecting two views isomorphic to  $(G, v)$  with a flexible path length. Similarly, we can apply the same argument for layer- $\ell$  paths for each  $1 \leq \ell \leq k$ , so we infer that the labels in  $\Sigma_1$  cannot be used to label nodes in layer 1 or above.

For the inductive step, suppose that we already know that  $\Sigma_1, \Sigma_2, \dots, \Sigma_{i-1}$  cannot be used to label nodes in layer  $i - 1$  or above. We consider the LCL problem that is the restriction of  $\Pi$  to the set of labels  $\Sigma_i \cup \Sigma_{i+1} \cup \dots \cup \Sigma_k$ . The above argument still works if we replace  $\Sigma_1$  by  $\Sigma_i$  and only consider the layers  $i \leq \ell \leq k$ , as we recall that  $\Sigma_i$  is the set of path-inflexible labels when we restrict to the set of labels  $\Sigma_i \cup \Sigma_{i+1} \cup \dots \cup \Sigma_k$ .

It appears that this approach allows us to show that for each  $1 \leq \ell \leq k$ , the layer- $\ell$  nodes cannot be labeled using  $\Sigma_1, \Sigma_2, \dots, \Sigma_\ell$ , so the given algorithm  $A$  cannot produce any output label for the layer- $k$  nodes, contradicting the correctness of  $A$ . This approach, however, has one issue. Consider again the layer- $\ell$  path  $P = v_1 \leftarrow v_2 \leftarrow \dots \leftarrow v_x$  in  $T_k^x$  in the above discussion. We are only able to show that the labels in  $\Sigma_1$  cannot be used to label the nodes  $v_j$  for each  $1 + t < j < x - t$ , as the radius- $t$  view of the remaining nodes in  $P$  are different. This is problematic because in the next level of induction, when we try to show that the labels in  $\Sigma_2$  cannot be used to label some node  $v$  that is in layer 2 or above, the proof relies on the condition that the labels in  $\Sigma_1$  cannot be used to label  $v$  and its children; see Lemma 5.11 and its proof. In particular, showing that  $\Sigma_1$  cannot be used to label the middle nodes in  $P$  whose radius- $t$  views are identical is not enough.

To resolve this issue, we need to consider essentially all possible radius- $t$  centered graph  $(G, v)$  corresponding to a radius- $t$  view of a layer- $i$  node, and we have to make sure that for any sufficiently large number  $d$ , we can find a rooted tree  $T$  that contains a length- $d$  directed path  $P = v_1 \leftarrow v_2 \leftarrow \dots \leftarrow v_{d+1}$  such that the radius- $t$  views of the two endpoints  $v_1$  and  $v_{d+1}$  are isomorphic to  $(G, v)$  and all the intermediate nodes  $v_2, v_3, \dots, v_d$  are in layer  $i$  or above, so that Lemma 5.11 is applicable.

To deal with the views  $(G, v)$  that do not belong to the central part of the long paths, we will need to concatenate two

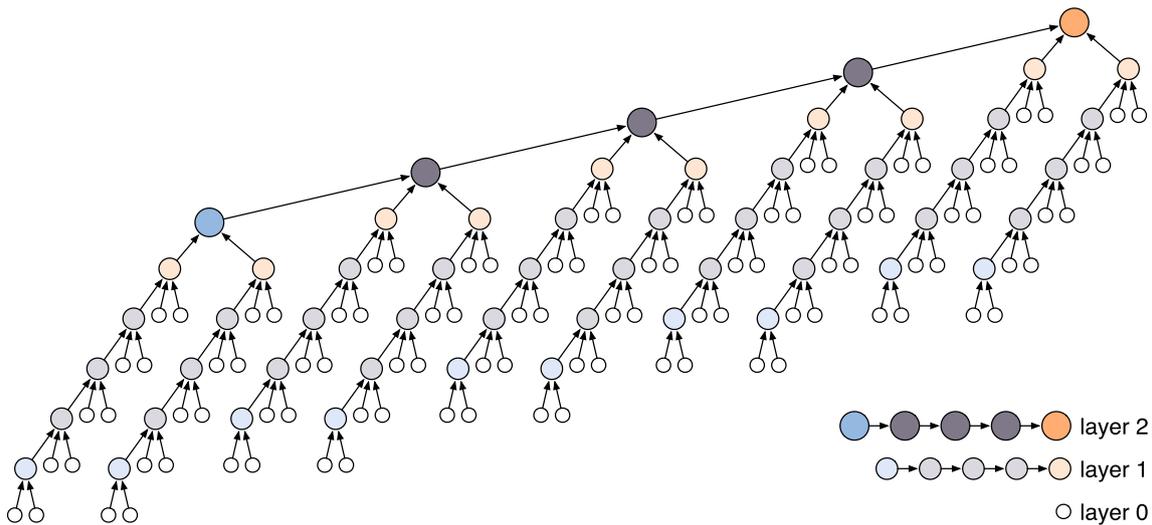


Fig. 4 Construction of the bipolar tree  $T_k^x$ , illustrated here for  $\delta = 3$ ,  $x = 5$ , and  $k = 2$

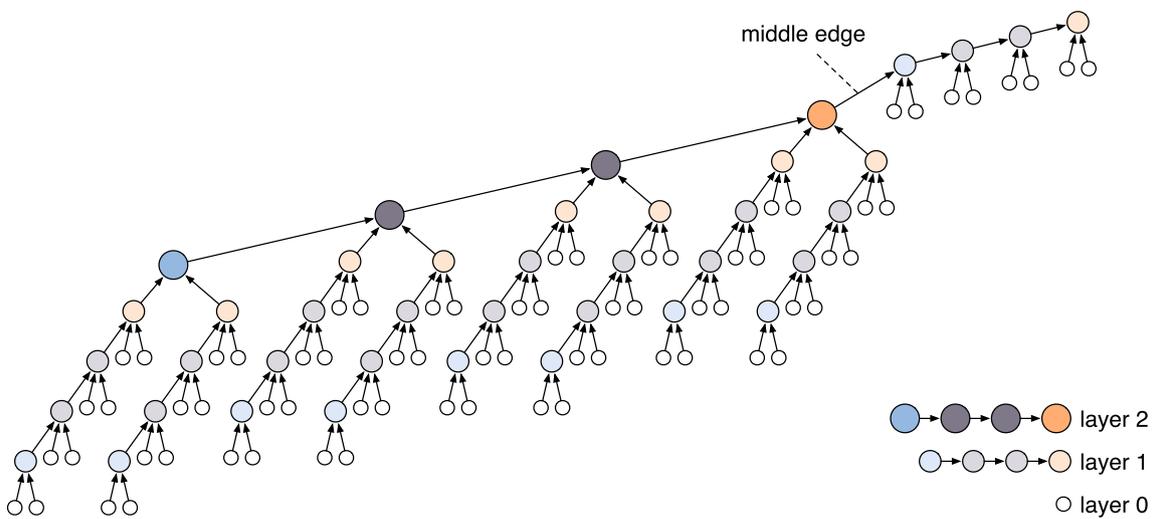


Fig. 5 Tree  $T_{i \leftarrow j}^x$ , for  $\delta = 3$ ,  $x = 4$ ,  $j = 2$ , and  $i = 1$

trees  $T_i^x$  and  $T_j^x$  for some  $1 \leq i \leq k$  and  $1 \leq j \leq k$  to obtain directed paths starting and ending with the same view  $(G, v)$ , so that we can apply Lemma 5.11. Such a concatenation will create new views that did not exist before in  $T_k^x$ . In order to capture all such views, we will consider the following definition  $T_{i \leftarrow j}^x$  and build the argument around it; see Fig. 5.

- For  $1 \leq i \leq k$  and  $1 \leq j \leq k$ , define  $T_{i \leftarrow j}^x$  as the result of the following construction. Let  $T_1' = T_i^x$  (distinguished nodes are  $s_1$  and  $t_1$ ) and  $T_2' = T_j^x$  (distinguished nodes are  $s_2$  and  $t_2$ ). Concatenate these two bipolar trees into a new bipolar tree by adding an edge  $\{t_1, s_2\}$  and setting  $s = s_1$  and  $t = t_2$ . We call  $e = \{t_1, s_2\}$  the *middle edge*. The layer numbers of the nodes are kept when  $T_i^x$  and  $T_j^x$  are linked together into  $T_{i \leftarrow j}^x$ .

We make the following two observations. For the special case of  $i = j$ ,  $T_{i \leftarrow i}^x$  is simply  $\bigoplus^{2x} T_{i-1}^x$ . For any number  $t$ , the number of nodes in the radius- $t$  neighborhood of any node in  $T_{i \leftarrow j}^x$  is  $O(t^{\max\{i,j\}}) = O(t^k)$ , regardless of  $x$ .

*A sequence of sets of radius- $t$  centered graphs.* Now, we are ready to define the set of radius- $t$  centered graphs  $S_i$ , for each  $1 \leq i \leq k$ . In the definition of  $S_i$ , we let  $x$  be any integer such that  $x \geq 2t + 2$ . It will be clear from the construction of  $S_i$  that the definition of  $S_i$  is invariant of the choice of  $x$ , as long as  $x$  is sufficiently large comparing with  $t$ .

The set  $S_i$  consists of all radius- $t$  centered graphs  $(G, v)$  such that there exists a node  $u$  in the rooted tree  $T_{a \leftarrow b}^x$  for some  $a$  and  $b$  meeting the following conditions.

- $i \leq a \leq k$  and  $i \leq b \leq k$ .
- The radius- $t$  view of  $u$  is isomorphic to  $(G, v)$ .

- The radius- $t$  view of  $u$  contains at least one node in the middle edge  $e$  of  $T_{a \leftarrow b}^x$ .
- The layer number of  $u$  is at least  $i$ .

Note that the threshold  $x \geq 2t + 2$  is chosen to make sure that for each node  $u'$  in the radius- $t$  neighborhood of  $u$ , if  $u'$  is not in layer zero, then its degree is exactly  $\Delta = \delta + 1$ , that is,  $u'$  has one parent and  $\delta$  children.

It is clear from the above definition of  $S_i$  that we have  $S_1 \supseteq S_2 \supseteq \dots \supseteq S_k \neq \emptyset$ . Before we proceed, we prove a result showing that  $S_i$  includes essentially all radius- $t$  view for layer- $i$  nodes in  $T_i^x$ . Formally, for each  $1 \leq i \leq k$ , we define  $S_i^*$  as the set of all radius- $t$  centered graphs  $(G, v)$  meeting the following conditions.

- There exist  $x \geq 1, i \leq j \leq k$ , and a layer- $i$  node  $u$  in  $T_j^x$  such that the radius- $t$  view of  $u$  in  $T_j^x$  is isomorphic to  $(G, v)$ . Furthermore, for each node  $u'$  in the radius- $t$  neighborhood of  $u$ , if  $u'$  is not in layer zero, then its degree is exactly  $\Delta = \delta + 1$ .

Intuitively,  $S_i^*$  is the set of all possible radius- $t$  views for layer- $i$  nodes, excluding those near the boundary. We exclude the views involving boundary nodes because we want to focus on the interior part of the graph where all nodes have the same degree  $\Delta = \delta + 1$ , except the layer-0 nodes whose degree is always one.

**Lemma 5.12** *For each  $1 \leq i \leq k$ , we have  $S_i^* \subseteq S_i$ .*

**Proof** Consider the node  $u$  in the graph  $T_j^x$  in the definition of  $S_i^*$ . Since the radius- $t$  neighborhood of  $u$  does not include any non-leaf node whose degree smaller than  $\Delta = \delta + 1$ , we may assume that  $x$  is an arbitrarily large number. Let  $u^*$  be any node in the radius- $t$  neighborhood of  $u$  that has the highest layer number. Let  $i^*$  be the layer number of  $u^*$ . We have  $i \leq i^* \leq j \leq k$ . The radius- $t$  neighborhood of  $u$  is confined to some subgraph  $T_{i^*}^x$  of  $T_j^x$  where  $u^*$  lies on the core path of  $T_{i^*}^x$ . The graph  $T_{i^*}^x$  can be viewed as a subgraph of  $T_{i^* \leftarrow i^*}^x$  such that  $u^*$  is a node in the middle edge of  $T_{i^* \leftarrow i^*}^x$ . As  $u$  is within distance  $t$  to  $u^*$  and the radius- $t$  view of  $u$  in  $T_{i^* \leftarrow i^*}^x, T_{i^*}^x$ , and the original graph  $T_j^x$  are identical, we conclude that the radius- $t$  view of  $u$  is isomorphic to some member in  $S_i$  by considering the graph  $T_{i^* \leftarrow i^*}^x$ .  $\square$

*The lower bound proof.* For any given integer  $t$ , we pick  $n$  to be the maximum number of nodes in the radius- $(t + 2)$  neighborhood of any node in  $T_{i \leftarrow j}^x$ , over all choices of  $i, j$ , and  $x$  such that  $1 \leq i \leq k, 1 \leq j \leq k$ , and  $x \geq 1$ . It is clear that  $n = O(t^k)$ , or equivalently  $t = \Omega(n^{1/k})$ . Therefore, to prove an  $\Omega(n^{1/k})$  lower bound for the given problem  $\Pi$ , it suffices to show the non-existence of a  $t$ -round algorithm  $A$  that solves  $\Pi$  on  $n$ -node graphs.

Suppose such an algorithm  $A$  exists. In Lemma 5.13, whose proof is deferred, we will prove by induction that all labels in  $\Sigma_i$  are not permissible for all centered graphs in  $S_i$ , for each  $1 \leq i \leq k$ . In particular, this means that all labels in  $\Sigma$  are not permissible for all centered graphs in  $S_k$ , as  $\Sigma = \Sigma_1 \cup \Sigma_2 \cup \dots \cup \Sigma_k$  and  $S_1 \supseteq S_2 \supseteq \dots \supseteq S_k$ .

**Lemma 5.13** *For each  $1 \leq j \leq k$ , all labels in  $\Sigma_j$  are not permissible for all centered graphs in  $S_j$ .*

We now prove the main result of this section assuming Lemma 5.13.

**Lemma 5.14** *If Algorithm 2 returns  $\epsilon$  after  $k$  iterations, then  $\Pi$  requires  $\Omega(n^{1/k})$  rounds to solve.*

**Proof** In view of the above discussion, it suffices to show that the algorithm  $A$  considered above does not exist. Recall that  $S_k \neq \emptyset$ , and each  $(G, v) \in S_k$  is isomorphic to the radius- $t$  view of some node  $u$  in  $T_{a \leftarrow b}^x$  with  $1 \leq i \leq k, 1 \leq j \leq k$ , and  $x \geq 2t + 2$ . Furthermore, the radius- $(t + 2)$  neighborhood of  $u$  contains at most  $n$  nodes. If we run  $A$  on the subgraph induced by the radius- $(t + 2)$  neighborhood of  $u$  in  $T_{a \leftarrow b}^x$ , then according to Lemma 5.13 the algorithm  $A$  does not output any label for  $u$ , violating the correctness of  $A$ , so such an algorithm  $A$  does not exist.  $\square$

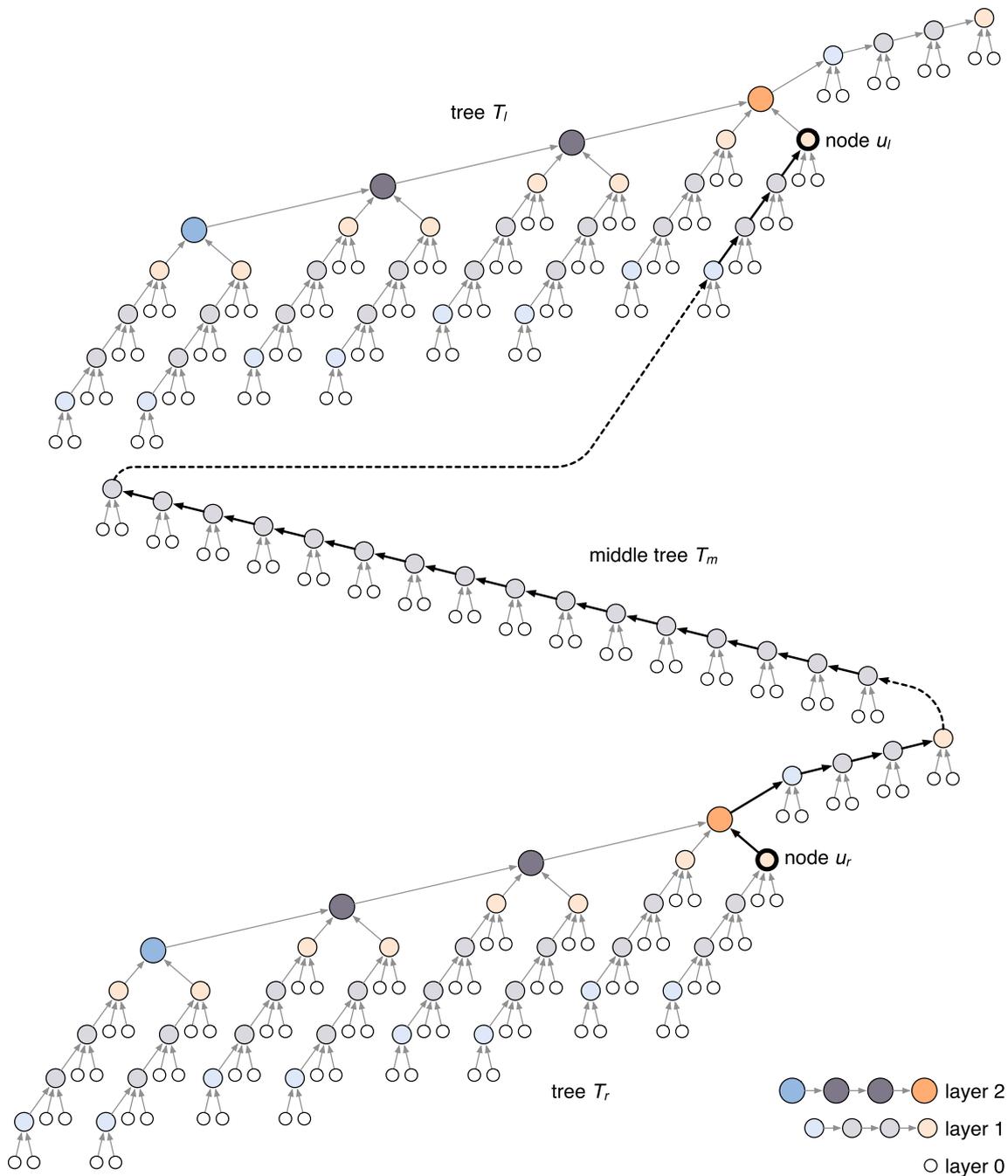
It is clear that Lemma 5.14 implies Theorem 5.2. *Constructing a rooted tree  $T$  for applying Lemma 5.11.* For the rest of the section, we prove Lemma 5.13. We begin with describing the construction of the rooted tree  $T$  needed for applying Lemma 5.11 in the proof of Lemma 5.13; see Fig. 6 for an illustration.

The construction of  $T$  is parameterized by any  $(G, v) \in S_j$ , for any  $1 \leq j \leq k$ . Recall from the definition of  $S_j$  that  $(G, v)$  is isomorphic to the radius- $t$  view of some layer- $i$  node  $u$  in  $T_{a \leftarrow b}^x$  such that  $1 \leq a \leq k, 1 \leq b \leq k$ , and  $j \leq i \leq \min\{a, b\}$ , and this radius- $t$  neighborhood contains at least one node in the middle edge  $e$  of  $T_{a \leftarrow b}^x$ .

From now on we fix  $x = 2t + 4$ . Then  $D = (k + 1)x - 1$  is an upper bound on the length of any root-to-leaf path in  $T_{a \leftarrow b}^x$ , for any  $1 \leq a \leq k$  and  $1 \leq b \leq k$ . We define  $K = 2D + x + 1$ .

The construction of  $T$  is also parameterized by a distance parameter  $d$  such that  $d \geq K$ . In the rooted tree  $T$  that we construct, there will be a length- $d$  directed path  $P = v_1 \leftarrow v_2 \leftarrow \dots \leftarrow v_{d+1}$  satisfying some good properties to make Lemma 5.11 applicable.

Intuitively,  $T$  will be the result of concatenating two copies  $T_l$  and  $T_r$  of  $T_{a \leftarrow b}^x$  via a middle tree  $T_m = \bigoplus^y T_{i-1}^x$ , and then  $P$  will be the unique directed path in  $T$  connecting the two copies of  $u$  in  $T_l$  and  $T_r$ . Note that  $T_m = \bigoplus^y T_{i-1}^x$  is simply a variant of  $T_i^x = \bigoplus^x T_{i-1}^x$  such that the length of the core path is  $y$  instead of  $x$ . We select  $y$  to make the length of  $P$  equals  $d$ . The points of concatenation will be selected to



**Fig. 6** Constructing a rooted tree  $T$  for applying Lemma 5.11. Here  $(G, v)$  was isomorphic to the view of some node  $u$  in layer 1 of  $T_{2^{i-1}}^x$ . Therefore we construct two copies of  $T_{2^{i-1}}^x$ , one of them is called  $T_l$  and the other one is  $T_r$ , and we identify the nodes  $u_l$  and  $u_r$  that have

views isomorphic to  $(G, v)$ . We identify the unique path  $P_r$  from  $u_r$  to the root of  $T_r$  and the unique layer-1 path  $P_l$  that takes us to  $u_l$  (dark arrows). Finally, we connect  $P_r$  through the middle tree to  $P_l$ . Note that the resulting path  $P = P_l \leftarrow P_m \leftarrow P_r$  does not use layer-0 nodes

ensure that all nodes in  $P$  are in layer  $i$  or above. Formally, the construction of the rooted tree  $T$  and its length- $d$  path  $P$  is as follows.

**The two trees  $T_l$  and  $T_r$ .** Recall that  $u$  is a layer- $i$  node in  $T_{a \leftarrow b}^x$  whose radius- $t$  neighborhood contains at least

one node in the middle edge. We consider two copies of  $T_{a \leftarrow b}^x$ , called  $T_l$  and  $T_r$ . The two copies of  $u$  in  $T_l$  and  $T_r$  are called  $u_l$  and  $u_r$ . Similarly, we write  $(s_l, t_l)$  and  $(s_r, t_r)$  to denote the two distinguished nodes of  $T_l$  and  $T_r$ .

**The path  $P_l$ .** If  $u_l$  is on the core path of  $T_l$ , then we define  $P_l$  to be the unique directed path  $u_l \leftarrow \dots \leftarrow t_l$ . Otherwise, then consider the unique layer- $i$  path  $w_1 \leftarrow \dots \leftarrow w_x$  that contains  $u_l$ , and then we define  $P_l$  to be the unique directed path  $u_l \leftarrow \dots \leftarrow w_x$ . Observe that all nodes in  $P_l$  are in layer  $i$  or above.

**The path  $P_r$ .** We define  $P_r$  to be the unique directed path  $s_r \leftarrow \dots \leftarrow u_r$  in  $T_r$ . Observe that all nodes in  $P_r$  are in layer  $i$  or above.

**The middle tree  $T_m$  and its path  $P_m$ .** Let  $D_l$  and  $D_r$  denote the lengths of  $P_l$  and  $P_r$ . Note that  $D_l \leq D$  and  $D_r \leq D$ . We define  $T_m = \bigoplus^y T_{i-1}^x$ , where  $y = d - D_l - D_r - 1$ , and we define  $P_m$  as the core path of  $T_m$ . Note that we must have  $y \geq x$ , due to the assumption  $d \geq K$  and our choice of  $K = 2D + x + 1$ . Clearly, all nodes in  $P_m$  are in layer  $i$ .

**Concatenation.** Now, we are ready to define the rooted tree  $T$  and its associated length- $d$  directed path  $P$ . We construct the directed path  $P$  by adding two edges to concatenate the three paths  $P_l, P_m,$  and  $P_r$  together:  $P = P_l \leftarrow P_m \leftarrow P_r$ . The length of  $P$  is exactly  $d$  due to our choice of  $y = d - D_l - D_r - 1$ . The rooted tree  $T$  is the result of this concatenation of  $T_l, T_m,$  and  $T_r$ .

The radius- $t$  views of  $u_l$  is isomorphic to  $(G, v)$ , regardless of the underlying graph being  $T$  or  $T_l$ . Similarly, the radius- $t$  views of  $u_r$  is isomorphic to  $(G, v)$ , regardless of the underlying graph being  $T$  or  $T_r$ . Hence the radius- $t$  neighborhoods of the two endpoints of  $P$  in  $T$  are isomorphic to the given radius- $t$  centered graph  $(G, v)$ . We also note that all nodes in  $P$  are in layer  $i$  or above, so all children of nodes in  $P$  are in layer  $i - 1$  or above.

Next, we will prove some additional properties of  $T$  and  $P$ . We begin with Lemma 5.15 and Lemma 5.16. Informally, in these lemmas we show that the local view seen from an edge connecting  $T_l, T_m,$  and  $T_r$  is isomorphic to the local view seen from the middle edge  $e$  of  $T_{a' \leftarrow b'}^x$ , for some choices of  $i \leq a' \leq k$  and  $i \leq b' \leq k$ .

**Lemma 5.15** *Let  $e_l = u' \leftarrow v'$  be the edge connecting  $T_l$  and  $T_m$ . Let  $U_l$  be the union of the radius- $(x - 1)$  neighborhood of  $u'$  and  $v'$  in  $T$ . There is a subgraph  $T'_l$  of  $T$  isomorphic to  $T_{a' \leftarrow b'}^x$  for some  $i \leq a' \leq k$  and  $b' = i$  such that  $T'_l$  contains all nodes in  $U_l$ . In the isomorphism, the edge  $e_l$  is mapped to the middle edge  $e$  of  $T_{a' \leftarrow b'}^x$ .*

**Proof** Let  $a'$  be the layer number of  $u'$ . Note that we have either  $a' = i$  or  $a' = b$ . In any case,  $i \leq a' \leq k$ . Consider the  $2x$ -node path  $u'_1 \leftarrow \dots \leftarrow u'_x = u' \leftarrow v' = v'_1 \leftarrow \dots \leftarrow v'_x$  in  $T$  defined as follows.

- $e_l = u' \leftarrow v'$  is the edge connecting  $T_l$  and  $T_m$ .
- $u'_1 \leftarrow \dots \leftarrow u'_x$  is the unique layer- $a'$  path in  $T_l$  containing  $u'$ .

- $v'_1 \leftarrow \dots \leftarrow v'_x$  is the path formed by the first  $x$  nodes in  $P_m$ .

We consider the subgraph  $T_{a'}^x$  induced by the nodes  $u'_1 \leftarrow \dots \leftarrow u'_x$  and their descendants in  $T_l$ . As  $v'_1 \leftarrow \dots \leftarrow v'_x$  are the first  $x$  nodes in the  $y$ -node core path of  $T_m = \bigoplus^y T_{i-1}^x$ , the nodes  $v'_1 \leftarrow \dots \leftarrow v'_x$  and their descendants induce a subgraph  $T_{b'}^x$  with  $b' = i$ . We choose  $T'_l$  to be the union of these two subgraphs  $T_{a'}^x$  and  $T_{b'}^x$ , together with the edge  $e_l = u' \leftarrow v'$ . It is clear that  $T'_l$  is isomorphic to  $T_{a' \leftarrow b'}^x$  and contains all nodes in  $U_l$ .  $\square$

**Lemma 5.16** *Let  $e_r = u' \leftarrow v'$  be the edge connecting  $T_m$  and  $T_r$ . Let  $U_r$  be the union of the radius- $(x - 1)$  neighborhood of  $u'$  and  $v'$  in  $T$ . There is a subgraph  $T'_r$  of  $T$  isomorphic to  $T_{a' \leftarrow b'}^x$  for  $a' = i$  and  $b' = a$  such that  $T'_r$  contains all nodes in  $U_r$ . In the isomorphism, the edge  $e_r$  is mapped to the middle edge  $e$  of  $T_{a' \leftarrow b'}^x$ .*

**Proof** Recall that  $T_m = \bigoplus^y T_{i-1}^x$  with  $y \geq x$  and  $T_r = T_{a \leftarrow b}^x$  is formed by connecting  $T_a^x$  and  $T_b^x$ . We write  $v'_1 \leftarrow \dots \leftarrow v'_y$  to denote the core path of  $T_m$ , and we let  $T' = \bigoplus^x T_{i-1}^x = T_i^x$  be a subtree of  $T_m$  induced by the  $x$ -node subpath  $v'_{y-x+1} \leftarrow \dots \leftarrow v'_y$  and the descendants of the nodes in this subpath.

The edge  $e_r = u' \leftarrow v'$  connects the two trees  $T' = T_i^x$  and  $T_a^x$ , as  $u'$  is the distinguished node  $t$  of  $T' = T_i^x$  and  $v'$  is the distinguished node  $s$  of  $T_a^x$ . Therefore, we may take  $T'_r$  to be the union of  $T' = T_i^x$  and  $T_a^x$ , together with the edge  $e_r = u' \leftarrow v'$ . The tree  $T'_r$  is isomorphic to  $T_{i \leftarrow a}^x$  and contains all nodes in  $U_r$ .  $\square$

Combining Lemma 5.15 and Lemma 5.16, in Lemma 5.17 we show that the local neighborhood of any node in  $T$  is isomorphic to the local neighborhood of some node in  $T_{a' \leftarrow b'}^x$ , for some choices of  $1 \leq a' \leq k, 1 \leq b' \leq k$  and  $x' \geq 1$ . In the proof of Lemma 5.17 we utilizes the fact that  $x = 2t + 4$ .

**Lemma 5.17** *For each node  $w$  in  $T$ , the subgraph induced by its radius- $(t+2)$  neighborhood is isomorphic to the subgraph induced by the radius- $(t+2)$  neighborhood of some node  $w'$  in  $T_{a' \leftarrow b'}^x$  for some  $1 \leq a' \leq k, 1 \leq b' \leq k$  and  $x' \geq 1$ .*

**Proof** The proof is done by a case analysis. We write  $U$  to denote the set of nodes within the radius- $(t+2)$  neighborhood of  $w$  in  $T$ . If  $U$  is completely confined in one of  $T_l$  or  $T_r$ , then the lemma holds with  $T_{a' \leftarrow b'}^x = T_{a \leftarrow b}^x$ , as both  $T_l$  and  $T_r$  are isomorphic to  $T_{a \leftarrow b}^x$ . If  $U$  is completely confined in  $T_m$ , then the lemma holds with  $T_{a' \leftarrow b'}^x = T_{i-1 \leftarrow i-1}^y$ , as  $T_m = \bigoplus^y T_{i-1}^x$  is a subgraph of  $T_{i-1 \leftarrow i-1}^y$ , as we recall that  $y \geq x$ .

If  $U$  contains the edge  $e_l$  connecting  $T_l$  and  $T_m$ , then  $U \subseteq U_l$ , where  $U_l$  is defined in Lemma 5.15. Note that the fact that  $x = 2t + 4$  is used to show that  $U \subseteq U_l$ . Therefore, the lemma holds with the tree  $T_{a' \leftarrow b'}^x$  considered in Lemma 5.15.

Finally, the remaining case is that  $U$  contains the edge  $e_r$  connecting  $T_m$  and  $T_r$ . Similar to the previous case, using Lemma 5.16 we obtain that  $U \subseteq U_r$ , so the lemma holds with the tree  $T_{a' \leftarrow b'}^x$  considered in Lemma 5.16.  $\square$

Same as the notation used in Lemma 5.11, for the rest of the section, we write  $S$  to denote the set of the nodes in  $P$  and their children. Using Lemma 5.12, Lemma 5.15, and Lemma 5.16, we prove Lemma 5.18, which shows that the radius- $t$  view of each node in  $S$  belongs to  $S_{i-1}$ .

**Lemma 5.18** *If  $i > 1$ , then the radius- $t$  view of each node in  $S$  belongs to  $S_{i-1}$ .*

**Proof** Let  $w \in S$ . Let  $i'$  be the layer number of  $w$ . From the construction of  $P$  we already know that all nodes on the path  $P$  has layer number at least  $i$ , so their children have layer number at least  $i - 1$ , and so  $i' \geq i - 1$ .

We first consider the case where the radius- $t$  neighborhood of  $w$  contains a node in the edge  $e_l$  connecting  $T_l$  and  $T_m$ . Then  $w$  has the same radius- $t$  view in both  $T$  and the graph  $T_{a' \leftarrow b'}^x$  considered in Lemma 5.15. Since  $i' \geq i - 1$ ,  $a' \geq i > i - 1$ ,  $b' \geq i > i - 1$ , and  $w$  is within distance  $t$  to a node in the middle edge of  $T_{a' \leftarrow b'}^x$ , this radius- $t$  view belongs to  $S_{i-1}$  by its definition. The case of where the radius- $t$  neighborhood of  $w$  contains a node in the edge  $e_r$  connecting  $T_m$  and  $T_r$  can be handled using Lemma 5.16 similarly.

From now on, we assume that the radius- $t$  neighborhood of  $w$  does not contain any node in  $e_l$  and  $e_r$ . There are three cases depending on whether the radius- $t$  neighborhood of  $w$  is confined to  $T_l$ ,  $T_m$ , or  $T_r$ .

Consider the case where the radius- $t$  neighborhood of  $w$  is confined to  $T_m$ . Since  $x$  is sufficiently large, the radius- $t$  neighborhood of  $w$  does not contain any non-leaf node whose degree is not  $\Delta = \delta + 1$ . Observe that  $T_m = \bigoplus^y T_{i-1}^x$  is a subgraph of  $T_i^y$  as  $y \geq x$ , so the radius- $t$  view of  $w$  is the same in  $T$ ,  $T_m$ , and  $T_i^y$ , and so this radius- $t$  view belongs to  $S_{i'}$ . By Lemma 5.12, we have  $S_{i'}^* \subseteq S_{i'}$ . We also have  $S_{i'} \subseteq S_{i-1}$  because  $i' \geq i - 1$ . Hence we conclude that this radius- $t$  view belongs to  $S_{i-1}$ , as desired.

For the rest of the proof, we consider the case where the radius- $t$  neighborhood of  $w$  is confined to  $T_l$ , as the case of  $T_r$  is similar. Recall that  $T_l = T_{a \leftarrow b}^x$  is constructed by concatenating  $T_a^x$  and  $T_b^x$  by a middle edge  $e$ . If the radius- $t$  neighborhood of  $w$  contains a node of  $e$ , then we know that this radius- $t$  view belongs to  $S_{i-1}$ , as we have  $a \geq i > i - 1$ ,  $b \geq i > i - 1$ , and  $i' \geq i - 1$ . Otherwise, the radius- $t$  neighborhood of  $w$  is confined to either  $T_a^x$  or  $T_b^x$ . Similarly, we may use Lemma 5.12 to show that the radius- $t$  view of  $w$  is in  $S_{i-1}$ .  $\square$

Using Lemma 5.17 and Lemma 5.18, we are now ready to prove Lemma 5.13.

**Proof of Lemma 5.13** By induction hypothesis, suppose that the lemma statement holds for smaller  $j$ -values. Fix any  $(G, v) \in S_j$ . Then  $(G, v)$  is isomorphic to the radius- $t$  neighborhood of a layer- $i$  node  $u$  in  $T_{a \leftarrow b}^x$  such that  $j \leq i \leq \min\{a, b\}$  and this radius- $t$  neighborhood contains at least one node in the middle edge  $e$  of  $T_{a \leftarrow b}^x$ .

Given  $T_{a \leftarrow b}^x$  and  $u$ , construct the rooted tree  $T$  and its directed path  $P$  as we discuss above. Remember in our construction there is a number  $K$  such that for each  $d \geq K$ , we are able to make  $d$  the length of  $P$ .

Consider  $\Sigma^\diamond = \Sigma_1 \cup \Sigma_2 \cup \dots \cup \Sigma_{j-1}$ . Fix any  $\sigma \in \Sigma_j$ . Recall that  $\Sigma_j$  is the set of path-inflexible labels for the restriction of  $\Pi$  to  $\Sigma \setminus \Sigma^\diamond$ . To prove the lemma, it suffices to show that  $\sigma$  is not permissible for  $(G, v)$ .

We apply Lemma 5.11 with the rooted tree  $T$  and its directed path  $P$  with  $\Sigma^\diamond$ . We will see that the properties of  $T$  and  $P$  that we discuss above imply that the three conditions of Lemma 5.11 are met. Condition (1) follows immediately from the construction of  $T$ . For Condition (2), if  $j = 1$ , then  $\Sigma^\diamond = \emptyset$ , so Condition (2) trivially holds; if  $j > 1$ , then  $i \geq j > 1$ , so we may apply Lemma 5.18 to obtain that for each node  $w \in S$ , its radius- $t$  neighborhood in  $T$  is in  $S_{i-1} \subseteq S_{j-1}$ . Therefore, by induction hypothesis, we know that each  $\sigma' \in \Sigma^\diamond$  is not permissible for the radius- $t$  view of each  $w \in S$ , so Condition (2) holds. Condition (3) follows from Lemma 5.17 that the radius- $(t + 2)$  neighborhood of each node in  $P$  is isomorphic to the radius- $(t + 2)$  neighborhood of some node in  $T_{a' \leftarrow b'}^{x'}$  for some  $1 \leq a' \leq k$ ,  $1 \leq b' \leq k$  and  $x' \geq 1$ , and our choice of  $n$  guarantees that the radius- $(t + 2)$  neighborhood of any node in  $T_{a' \leftarrow b'}^{x'}$  cannot contain more than  $n$  nodes. Hence Lemma 5.11 is applicable, so  $\sigma$  is not permissible for  $(G, v)$ .  $\square$

## 6 Sublogarithmic region

In this section we prove that there is no LCL problem  $\Pi$  with distributed time complexity between  $\omega(\log^* n)$  and  $o(\log n)$ . Also, we prove that, given a problem  $\Pi$ , we can *decide* if its complexity is  $O(\log^* n)$  or  $\Omega(\log n)$ . Moreover, we prove that randomness cannot help: if a problem has *randomized* complexity  $O(\log^* n)$ , then it has the same *deterministic* complexity.

### 6.1 High-level idea

Informally, we prove that all problems that are  $O(\log^* n)$  solvable can be solved in a normalized way, that is the following:

- Split the rooted tree in constant size rooted subtrees, where each root has some minimum distance from the leaves. Note that each leaf is the root of another subtree.

- In each subtree, assign labels to the leaves, such that for any assignment to the root, the subtree can be completed with a valid labeling.
- Complete the labeling in each subtree independently.

Note that the only part requiring  $\Theta(\log^* n)$  is the first one, while the rest requires constant time. We then also prove that we can *decide* if there is a subset of labels, and an assignment for the leaves of the subtrees, that satisfies the second point.

### 6.2 Certificate

We start by defining what is a uniform certificate of  $O(\log^* n)$  solvability. Informally, it is a sequence of labeled trees having the same depth and the leaves labeled in the same way, such that for each label used in the trees there is a tree with the root labeled with that label. An example of such a certificate for the 3-coloring problem is depicted in Fig. 7.

**Definition 6.1** (*uniform certificate for  $O(\log^* n)$  solvability*) Let  $\Pi$  be an LCL problem. A uniform certificate of  $O(\log^* n)$  solvability for  $\Pi$  with labels  $\Sigma_{\mathcal{T}} = \{\sigma_0, \dots, \sigma_t\} \subseteq \Sigma(\Pi)$  and depth  $d$  is a sequence  $\mathcal{T}$  of  $t$  labeled trees (denoted by  $\mathcal{T}_i$ ) such that:

- (1) Each tree is a complete  $\delta$ -ary tree of depth  $d$  ( $d$  has to be at least one).
- (2) Each tree  $\mathcal{T}_i$  is labeled with labels from  $\Sigma_{\mathcal{T}}$  and correct w.r.t. configurations  $C(\Pi)$ .
- (3) Let  $\overline{\mathcal{T}}_i$  be the tree obtained by starting from  $\mathcal{T}_i$  and removing the labels of all non-leaf nodes. It must hold that all trees  $\overline{\mathcal{T}}_i$  are isomorphic, preserving the labeling.
- (4) Root of tree  $\mathcal{T}_i$  is labeled with label  $\sigma_i$ .

We will see that a problem  $\Pi$  can be solved in  $O(\log^* n)$  rounds if and only if a certificate of  $O(\log^* n)$  solvability for  $\Pi$  exists. We will later show that we can decide if such a certificate exists. We will now give an alternative definition of certificate, that we will later prove to be equivalent.

**Definition 6.2** (*coprime certificate for  $O(\log^* n)$  solvability*) Let  $\Pi$  be an LCL problem. A coprime certificate of  $O(\log^* n)$  solvability for  $\Pi$  with labels  $\Sigma_{\mathcal{T}} = \{\sigma_0, \dots, \sigma_t\} \subseteq \Sigma(\Pi)$  and depth pair  $(d_1, d_2)$  is a pair of sequences  $\mathcal{T}^1$  and  $\mathcal{T}^2$  of  $t$  labeled trees (denoted by  $\mathcal{T}_i^1$  and  $\mathcal{T}_i^2$ ) such that:

- (1) The depths  $d_1$  and  $d_2$  are coprime.
- (2) Each tree of  $\mathcal{T}^1$  (resp.  $\mathcal{T}^2$ ) is a complete  $\delta$ -ary tree of depth  $d_1 \geq 1$  (resp.  $d_2 \geq 1$ ).
- (3) Each tree is labeled with labels from  $\Sigma(\Pi)$  and correct w.r.t. configurations  $C(\Pi)$ .
- (4) Let  $\overline{\mathcal{T}}_i^1$  (resp.  $\overline{\mathcal{T}}_i^2$ ) be the tree obtained by starting from  $\mathcal{T}_i^1$  (resp.  $\mathcal{T}_i^2$ ) and removing the labels of all non-leaf

nodes. It must hold that all trees  $\overline{\mathcal{T}}_i^1$  (resp.  $\overline{\mathcal{T}}_i^2$ ) are isomorphic, preserving the labeling.

- (5) The root of the tree  $\mathcal{T}_i^1$  (resp.  $\mathcal{T}_i^2$ ) is labeled with label  $\sigma_i$ .

Note that the difference between a uniform certificate and a coprime certificate is that a coprime certificate requires two uniform certificates of coprime depth, but it allows internal nodes of the trees to be labeled from labels of  $\Sigma(\Pi)$  that are not in  $\Sigma_{\mathcal{T}}$ . In the following, we will sometimes omit the type of the certificate, and we will just talk about *certificate for  $O(\log^* n)$  solvability*. In this case, we will refer to a *uniform certificate*.

### 6.3 Upper bound

We now present an  $O(\log^* n)$ -round algorithm that is able to solve  $\Pi$  if there exists a certificate for  $O(\log^* n)$  solvability for  $\Pi$ .

**Theorem 6.3** *Assume that a uniform or coprime certificate for  $O(\log^* n)$  solvability for  $\Pi$  exists. Then  $\Pi$  can be solved in  $O(\log^* n)$  rounds in the CONGEST model.*

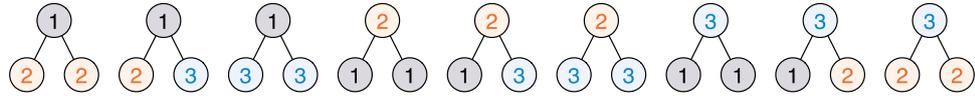
**Proof** We will prove our claim by describing an algorithm  $A$ . The algorithm will consist of two main phases. First, we split the tree into constant size subtrees in  $O(\log^* n)$  rounds. Then, we operate in a constant number of rounds on these subtrees in parallel. We assume that nodes are far enough from the root or the leaves of the tree, as we can imagine our tree to be embedded in a slightly larger tree. Since the leaves are unconstrained, this does not affect the validity of the solution that we compute.

Consider the following family of problems defined on directed paths. Let  $\alpha, \beta$  be two parameters. Labels are from  $\{1, a_2, \dots, a_\alpha\} \cup \{1, b_2, \dots, b_\beta\}$ . Allowed configurations are

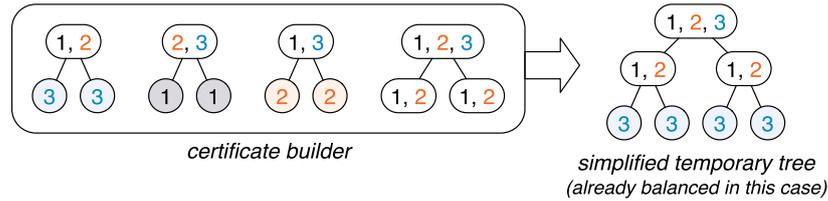
$$\{(a_i, a_{i+1}) \mid 2 \leq i < \alpha\} \cup \{(b_i, b_{i+1}) \mid 2 \leq i < \beta\} \cup \{(1, a_2), (1, b_2), (\alpha, 1), (\beta, 1)\}.$$

Essentially, this problem requires to label a directed path such that if a node is labeled 1, then its successor is either labeled  $a_2$  or  $b_2$ , and then in the first case we continue counting up to  $a_\alpha$  and then start again with 1, while in the second case we continue counting up to  $b_\beta$  and then start again with 1. If  $\alpha$  and  $\beta$  are coprime, then this problem can be solved in  $O(\log^* n)$  rounds, and by [17, Theorem 16] we can solve this problem in rooted trees, such that every root-to-leaf path is labeled with a valid labeling w.r.t. the definition of the problem on directed paths in  $O(\log^* n)$  rounds as well. Let us now modify the solution as follows: let  $\ell(v)$  be the labeling obtained on each node  $v$ . Each node  $v$  labels itself with  $\ell(u)$ , where  $u$  is the parent of  $v$ . We obtain that all siblings have the same

(a) Problem: 3-coloring in binary trees:



(b) Finding a certificate:



(c) Certificate for  $O(\log^* n)$ -round solvability:

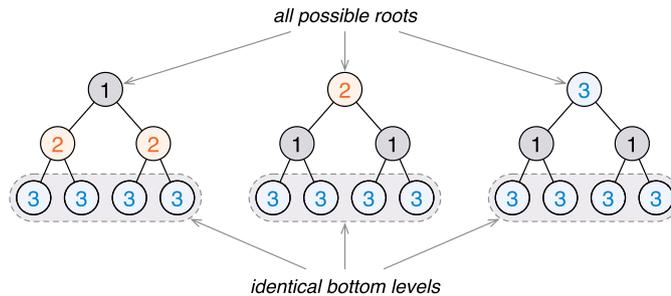


Fig. 7 Finding a uniform certificate for  $O(\log^* n)$  solvability (Definition 6.1) for the 3-coloring problem (Sect. 1.2)

labeling. Consider now the subtrees obtained by removing edges where endpoints are labeled  $(a_\alpha, 1)$  or  $(b_\beta, 1)$ . To each subtree, we add as new leaves the nodes on the other side of the edges that have been removed (that is, nodes labeled 1 are at the same time roots of their tree and leaves of the tree above). By how labels can propagate, we obtain that each obtained subtree is a perfect  $\delta$ -ary trees, where each tree has height either  $\alpha$  or  $\beta$ . If we are given a coprime certificate, we compute such a splitting with  $\alpha, \beta$  equal to the depth pair of the certificate, while if we are given a uniform certificate, we compute such a splitting with  $d, d + 1$ , where  $d$  is the depth of the certificate.

We now describe the second phase. In the following, we describe an algorithm that fixes, in constant time, the labeling of each subtree in parallel.

If we are given a coprime certificate, we proceed as follows. For every subtree of depth  $\alpha$  we label the leaves as the trees of  $\mathcal{T}^1$ , while for every tree of depth  $\beta$  we label the leaves as the trees of  $\mathcal{T}^2$ . Note that all the leaves are also the root of a tree below. Hence, for each subtree, we have fixed the labeling of the root and the leaves. Now, for each tree of depth  $\alpha$  (resp.  $\beta$ ) we complete the labeling as in  $\mathcal{T}_i^1$  (resp.  $\mathcal{T}_i^2$ ), where  $\sigma_i$  is the label assigned to the root. In this way, we obtain a valid labeling for the whole tree.

If we are given a uniform certificate, to each subtree, we assign to the nodes at depth  $d$  the labels of the trees of the cer-

tificate. On subtrees of depth  $d + 1$  we then assign a labeling to the nodes at depth  $d + 1$  by using only labels of the certificate. This is possible since each label of the certificate is a root of a certificate of the tree, and hence has a continuation below that only uses labels of the certificate. We now need to complete the labeling of trees of depth  $d$  where all roots and all leaves have labels of the certificate, and all internal nodes are unlabeled. This is possible by copying the labels assigned to the internal nodes of the trees of the certificate.

The round complexity of the described algorithm is  $O(\log^* n)$  for computing the subtrees, and  $O(1)$  for everything else, hence we have an algorithm that has a round complexity of  $O(\log^* n)$ .  $\square$

### 6.4 Lower bound

We now prove that, if there is no certificate for  $O(\log^* n)$  solvability, then the problem requires  $\Omega(\log n)$ , even for randomized algorithms.

We start by considering deterministic algorithms. We will prove that if there is a deterministic  $o(\log n)$  algorithm for  $\Pi$ , then we can construct an  $O(\log^* n)$  certificate for it. In the following lemma we will prove something stronger, that will be useful later when considering constant-time algorithms. This lemma essentially says that, if there exists a fast enough algorithm that uses some set of labels far enough from the

root and the leaves, then we can construct a certificate that uses the same set of labels. Moreover, we can force a leaf of the certificate to contain some specific label (this specific part will be used when considering constant-time algorithms).

**Lemma 6.4** *Assume that there exists a deterministic algorithm  $A$  solving  $\Pi$  in  $T(n) \in o(\log n)$  rounds on instances of size  $n$ . Let  $n_0$  be any integer satisfying  $n_0 > (1 + \delta)^{10T(n_0)}$ . Let  $S$  be the maximal set of labels satisfying that for each  $s \in S$  there exists an instance of  $\Pi$  of size  $n_0$  in which  $A$  outputs  $s$  on at least one node at distance strictly larger than  $T(n_0)$  from the root and from any leaf. Let  $\bar{s}$  be an arbitrary label in  $S$ . Then there exists a certificate of  $O(\log^* n)$  solvability that contains all labels in  $S$ , and where at least one leaf is labeled  $\bar{s}$ .*

**Proof** For all  $s \in S$ , let  $H_s$  be an instance of size  $n_0$  in which there exists a node  $v_s$  having distance strictly larger than  $T = T(n_0)$  from the root and any leaf, where  $A$  outputs  $s$ . Let  $B(v)$  be the radius- $T$  neighborhood of a node  $v$ .

We now consider a  $\delta$ -ary tree  $G$  of  $n_0$  nodes that is “as balanced as possible”. Note that the height of  $G$  is at least  $10T$ . Let  $r$  be an arbitrary node at distance  $T + 1$  from the root of  $G$ . Let  $L$  be the set of descendants of  $r$  that are at distance exactly  $5T$  from  $r$ . Since the tree is balanced and has height at least  $10T$ , then  $r$  and all nodes of  $L$  do not see the root of  $G$  or any leaf of  $G$ . Also,  $B(r)$  and  $B(\ell)$  are disjoint, for all  $\ell \in L$ . Moreover, note that for all nodes  $x$  that are on paths that connect  $r$  with nodes of  $L$ , it holds that either  $B(x)$  is disjoint with  $r$  or  $B(x)$  is disjoint with  $B(\ell)$ , for all  $\ell \in L$ .

We now pick an arbitrary node  $\ell \in L$  and fix the identifiers in its neighborhood to make  $B(\ell) = B(v_{\bar{s}})$  (that is the subgraph of  $H_{\bar{s}}$  in which node  $v_{\bar{s}}$  outputs  $\bar{s}$ ). Then, for each label  $s \in S$  we make a copy of  $G$  (copying the partial ID assignment as well), and we call it  $G_s$ . In each copy  $G_s$ , we additionally fix the neighborhood of  $r_s$ , that is the copy of  $r$ , to make it equal to  $B(v_s)$ . Then, we fix the identifiers of all other nodes by using unique identifiers not in  $B(v_s) \cup B(v_{\bar{s}})$ . Crucially, we assign the same identifier to all the nodes that are copies of the same node of  $G$ .

We argue that, by running  $A$  on the obtained trees, we must obtain a valid solution, even if some identifiers may not be unique. In fact, assume that there is a node in which the output does not satisfy the constraints of  $\Pi$ . The radius- $T$  neighborhood of this node must contain unique identifiers by construction, and we can hence construct a different instance of  $n_0$  nodes where all identifiers are unique and the same bad neighborhood is contained. This would imply that  $A$  fails in a valid instance, that is a contradiction.

Hence, we obtain that each tree  $G_s$  is properly labeled, and since  $S$  is by definition maximal, then also nodes that are between  $r_s$  and nodes in  $L_s$ , that are the copy of labels in  $L$ , must be labeled with only labels in  $S$ .

Consider now the  $|S|$  trees obtained by taking from each tree  $G_s$  the subtree induced by node  $r_s$ , nodes in  $L_s$ , and all nodes between them. We obtain  $|S|$  trees that have the same labeling for the leaves (that is, leaves that are copies of the same node of  $G$  have the same labeling), at least one leaf is labeled  $\bar{s}$ , each tree has a different label of  $S$  assigned to the root, and all nodes are only labeled with labels in  $S$ . Hence we obtained a certificate for  $O(\log^* n)$  solvability that contains all labels in  $S$ , and where at least one leaf is labeled  $\bar{s}$ .  $\square$

In particular, since for any algorithm running in  $o(\log n)$  rounds there exists some  $n_0$  satisfying  $n_0 > (1 + \delta)^{10T(n_0)}$ , and since in  $\delta$ -ary trees of size  $n_0$  there exist nodes at distance strictly larger than  $T$  from the root and any leaf, implying that  $S$  is non-empty, then Lemma 6.4 shows that if there is a deterministic  $o(\log n)$  algorithm for  $\Pi$ , then we can construct an  $O(\log^* n)$  certificate for it. Hence, we obtain the following corollary.

**Corollary 6.5** *If  $\Pi$  has deterministic complexity  $o(\log n)$  then there exists a certificate for  $O(\log^* n)$  solvability.*

We can now prove that uniform and coprime certificates are in some sense equivalent.

**Lemma 6.6** *A uniform certificate for  $O(\log^* n)$  solvability exists if and only if a coprime certificate of  $O(\log^* n)$  solvability exists.*

**Proof** We first show that, given a uniform certificate  $\mathcal{T}$ , we can construct a coprime certificate. Let  $d$  be the depth of the uniform certificate, we show how to construct a different certificate of depth  $d + 1$ . Since each leaf is also a root of some tree in  $\mathcal{T}$ , then each leaf has a continuation below. We can construct a certificate of depth  $d + 1$  by starting from the trees in  $\mathcal{T}$  and extending them to depth  $d + 1$  in a consistent manner by using the continuation below that is guaranteed to exist.

We now prove that if there exists a coprime certificate then there exists a uniform certificate. By Theorem 6.3 a coprime certificate implies a deterministic  $O(\log^* n)$  algorithm, and by Corollary 6.5 this implies the existence of a uniform certificate.  $\square$

We are now ready to extend Corollary 6.5 to randomized algorithms.

**Lemma 6.7** *Let  $\Pi$  be an LCL problem for which no certificate for  $O(\log^* n)$  solvability exists. Then, the randomized and deterministic complexity of  $\Pi$  in the LOCAL model is  $\Omega(\log n)$ .*

**Proof** We start by proving the lemma in the case where randomization is allowed and nodes have no identifiers assigned to them, and we prove the lemma by showing the contrapositive. To this end, let  $\Pi$  be an LCL problem with randomized

complexity  $T(n) \in o(\log n)$ . We will show that there exists a certificate for  $O(\log^* n)$  solvability for  $\Pi$ .

Let  $\mathcal{A}$  denote an optimal randomized algorithm for  $\Pi$ ; in particular the worst-case runtime of  $\mathcal{A}$  on  $n$ -node trees is  $T(n)$ . Let  $\Sigma(\Pi)$  be the output label set of  $\Pi$  and set  $k = |\Sigma(\Pi)|$ . Since  $T(n) \in o(\log n)$ , there exists an integer  $n_0 \geq 5k^2$  such that  $1 \leq T(n_0) \leq 1/10 \cdot \log_\delta(n_0)$ . Let  $G$  be a rooted tree with  $n_0$  nodes that is ‘‘as balanced as possible’’; in particular the  $\lfloor \log_\delta(n_0) \rfloor$ -hop neighborhood of the root  $r$  is a perfectly balanced tree. Moreover, for any node  $v$ , denote the set of descendants of  $v$  that are at distance precisely  $2T(n_0)+1$  from  $v$  by  $D_1(v)$ , and those at distance  $2T(n_0)+2$  from  $v$  by  $D_2(v)$ . Since  $T(n_0) \leq 1/10 \cdot \log_\delta(n_0)$ , there exists a node  $v \in V(G)$  such that the distance between  $r$  and  $v$  is at least  $T(n_0) + 1$ , and the distance between any node  $u \in D_1(v) \cup D_2(v)$  and the leaf closest to  $u$  is also at least  $T(n_0) + 1$ . In particular, for any node  $u \in D_1(v) \cup D_2(v)$  the views of  $u$  and  $v$  during an  $T(n_0)$ -round algorithm are disjoint, and both  $v$  and any node in  $D_1(v) \cup D_2(v)$  do not see any leaf or the root during an  $T(n_0)$ -round algorithm.

Let  $S \subseteq \Sigma(\Pi)$  be the set of all labels that  $\mathcal{A}$  outputs with probability at least  $1/(k\sqrt{n_0})$  if  $\mathcal{A}$  does not see a leaf or the root (note that the view of all nodes that do not see a leaf or the root is the same, since nodes have no ids). By the definition of  $D_1(v)$  and  $D_2(v)$ , we know that  $D_1(v) \cup D_2(v) \leq \sqrt{n_0}/2$ . Note that  $\mathcal{A}$  outputs a specific label not in  $S$  with probability strictly less than  $1/(k\sqrt{n_0})$ , and hence any label not in  $S$  with probability less than  $1/\sqrt{n_0}$  by a union bound, and hence the probability that at least one node in  $D_1(v) \cup D_2(v)$  outputs a label not in  $S$  is at most  $|D_1(v) \cup D_2(v)|/\sqrt{n_0}$ , by a union bound. Thus, the probability that all nodes in  $D_1(v) \cup D_2(v)$  output a label from  $S$  is at least

$$1 - \frac{|D_1(v) \cup D_2(v)|}{\sqrt{n_0}} \geq \frac{1}{2}.$$

Let us assume for a contradiction that there is no certificate for  $O(\log^* n)$  solvability for  $\Pi$ . By Lemma 6.6 this implies that there is no coprime certificate as well. In particular, then there is no such certificate with labels from  $S$  and depth pair  $2T(n_0)+1$  and  $2T(n_0)+2$ . This implies that there exists a  $D(v) \in \{D_1(v), D_2(v)\}$  such that for any labeling  $\ell: D(v) \rightarrow S$  of the nodes in  $D(v)$  with labels from  $S$ , there exists some label  $s \in S$  such that  $\ell$  is *incompatible* with  $s$ , i.e., such that there is no correct solution for  $\Pi$  where  $v$  is labeled with  $s$  and  $D(v)$  is labeled according to  $\ell$ .

Since we have  $|S| \leq k$ , and the view of  $v$  during  $\mathcal{A}$  and the union of the views of the nodes in  $D(v)$  during  $\mathcal{A}$  are disjoint, it follows that the probability that the labeling  $\ell$  of  $D(v)$  that  $\mathcal{A}$  outputs is incompatible with the label  $s$  that  $\mathcal{A}$  outputs at  $v$  is at least

$$\frac{1}{2} \cdot \frac{1}{k\sqrt{n_0}} > \frac{1}{n_0}$$

since  $n_0 \geq 5k^2$ . Hence,  $\mathcal{A}$  fails with probability larger than  $1/n_0$ , yielding a contradiction, and proving the lemma for the case in which identifiers are not provided to the nodes, but randomization is allowed.

We now prove that the same result holds even if we provide identifiers to the nodes. Assume that there exists a  $o(\log n)$  randomized algorithm for the case in which identifiers are provided. We can run it in the case in which identifiers are not provided in the same asymptotic running time by first generating unique random identifiers in  $\{1, \dots, n^3\}$ , that can be done with a randomized algorithm with high probability of success. But this would imply a  $o(\log n)$  randomized algorithm for the case in which identifiers are not provided, contradicting the lemma. Finally, since the existence of a deterministic algorithm implies the existence of a randomized algorithm, then the lemma follows.  $\square$

## 6.5 Decidability

We now prove that we can decide if a certificate for  $O(\log^* n)$  solvability exists. Algorithm 4 describes a procedure that returns a *certificate builder* if and only if a certificate exists. A certificate builder is an object that can be used to easily construct a certificate, and in Lemma 6.9 we will show that, given a certificate builder, we can indeed construct a certificate. This procedure uses another subroutine, Algorithm 3, to try to find a certificate builder that uses a specific subset of labels, for all possible subsets of labels. On a high level, Algorithm 3 works as follows. We start with singleton sets, one for each label. Then, we repeatedly try to build new sets. Each new set is obtained as follows. We consider all tuples of size  $\delta$  of existing sets, and we see which configurations exist where the label of each leaf  $\ell_i$  is contained in the  $i$ th sets of the tuple. The set of roots of such configurations defines a new set. We repeat this process until we obtain a fixed point. The algorithm that we describe is also able to find a certificate builder that contains a leaf with some specific label, if required and if it exists; we will need this in the next section.

We now prove that Algorithm 4 outputs a certificate builder if and only if a certificate of  $O(\log^* n)$  solvability exists.

**Theorem 6.8** *Given an LCL problem  $\Pi$ , Algorithm 4 outputs a certificate builder if and only if  $\Pi$  has an  $O(\log^* n)$  certificate satisfying the leaf requirement.*

**Proof** We will later show, in Lemma 6.9, that if Algorithm 4 outputs a certificate builder then a certificate exists, so we now prove the reverse implication, that is, if the problem has a certificate of  $O(\log^* n)$  solvability (satisfying the leaf requirement), then Algorithm 4 will find a certificate builder.

**Algorithm 3:** findUnrestrictedCertificate( $\Pi, a$ )

**Input:** LCL problem  $\Pi$ , label  $a \in \Sigma(\Pi)$  or  $a = \epsilon$   
**Output:**  $\epsilon$  if no certificate exists or a certificate builder  
 $i \leftarrow 0$   
 $R_0 \leftarrow \{(\{\sigma\}, \sigma = a) \mid \sigma \in \Sigma(\Pi)\} \triangleright R_i$  is a set of pairs. The first element of each pair is a set of possible labels of certificate roots. The second element is a Boolean that indicates whether such set of root labels can have label  $a$  as one of its leaves. For  $R_0$ , the Boolean is true if and only if  $\sigma$  is equal to  $a$ .  
 $C\mathcal{B} \leftarrow \emptyset \triangleright$  Certificate builder which will describe how we constructed individual elements in sets  $R_i$ .  
**repeat**  
     $i \leftarrow i + 1$   
     $R_i \leftarrow R_{i-1}$   
    **for every  $\delta$ -tuple of sets of root labels and indicators**  
     $(r_1, a_1), (r_2, a_2), \dots, (r_\delta, a_\delta)$  **from**  $R_{i-1}$  **do**  
         $r_n \leftarrow \{\sigma \mid (\sigma : c_1, c_2, \dots, c_\delta) \in C(\Pi), c_1 \in r_1, c_2 \in r_2, \dots, c_\delta \in r_\delta\}$   
         $a_n \leftarrow a_1$  or  $a_2$  or ... or  $a_\delta \triangleright$  Indicates whether  $r_n$  will have a leaf layer containing label  $a$ .  
        **if**  $r_n \neq \emptyset$  **and**  $(r_n, a_n) \notin R_i$  **then**  
             $C\mathcal{B} \leftarrow C\mathcal{B} \cup ((r_n, a_n), ((r_1, a_1), (r_2, a_2), \dots, (r_\delta, a_\delta)))$   
             $R_i \leftarrow R_i \cup (r_n, a_n)$   
        **end**  
    **end**  
**until**  $R_i = R_{i-1} \triangleright$  Until we do not enlarge the set  $R_i$ .  
**if**  $(\Sigma(\Pi), a \neq \epsilon) \in R_i$  **and**  $C(\Pi)$  **is non-empty then**  
    **return**  $C\mathcal{B}$   
**else**  
    **return**  $\epsilon$   
**end**

**Algorithm 4:** findCertificate( $\Pi$ )

**Input:** LCL problem  $\Pi$   
**Output:**  $\epsilon$  if no certificate exists or a certificate builder  
**for all elements  $\Sigma'$  of  $2^{\Sigma(\Pi)}$  do**  
     $\Pi' \leftarrow$  restriction of  $\Pi$  to labels from  $\Sigma'$   
    certificateBuilder  $\leftarrow$  findUnrestrictedCertificate( $\Pi', \epsilon$ )  
    **if** certificateBuilder  $\neq \epsilon$  **then**  
        **return** certificateBuilder  
    **end**  
**end**  
**return**  $\epsilon$

Let  $\mathcal{T} = (\mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_t)$  be a certificate with labels  $\Sigma_{\mathcal{T}}$  that satisfies the leaf requirement (that is, if  $a \neq \epsilon$  then at least one leaf is labeled  $a$ ). Let  $|\mathcal{T}| = t$ . Let  $\lambda$  be the labeling function of the certificate (that is, a function mapping each node of each tree of the certificate to its assigned label), and let  $t_{i,j,k}$  be the  $j$ th node on level  $i$  of the  $k$ th tree.

First, we define  $S_{i,j}$  as the set of all labels of the  $j$ th nodes on level  $i$ , that is,  $S_{i,j} = \bigcup_{k=1}^{|\mathcal{T}|} \lambda(t_{i,j,k})$ .  $S_{0,0}$  is by definition equal to  $\Sigma_{\mathcal{T}}$ , and for all nodes on level  $d$ ,  $S_{d,0}, S_{d,1}, \dots, S_{d,\delta^d}$  are singletons, by definition of certificate (recall that each  $\mathcal{T}_i$  has depth  $d$ ).

We will prove by induction on the depth of  $\mathcal{T}$  that, for all  $i$  and  $j$ , there exists a pair  $(S'_{i,j}, x)$  in the set  $R$  of Algorithm 3, where  $S'_{i,j} \supseteq S_{i,j}$ , and  $x$  is true if and only if  $t_{i,j,k}$  is an ancestor of leaves labeled  $a$ . This would imply that  $(\Sigma_{\mathcal{T}}, b)$  is also in  $R$ , where  $b$  is true if and only if  $a \neq \epsilon$ , and that Algorithm 3 outputs certificate builder which is what we want to prove.

In the base case, sets  $S_{d,i}$  are just singletons  $\{\sigma\} \subseteq \Sigma_{\mathcal{T}}$  and we add  $(S_{d,i}, \sigma = a)$  in the initialization of the set  $R$ .

For the induction hypothesis, let us assume that all  $(S'_{i+1,j}, x)$  for level  $i + 1$  are in  $R$ , where  $S'_{i+1,j} \supseteq S_{i+1,j}$ , and  $x$  is true if and only if  $t_{i+1,j,k}$  are ancestors of a leaf labeled  $a$ . We prove the statement for  $i$ . In the algorithm we loop over all  $\delta$ -tuples of elements from  $R$  to enlarge  $R$ , and hence also over the tuple  $((S'_{i+1,m}, a_1), (S'_{i+1,m+1}, a_2), \dots, (S'_{i+1,m+\delta-1}, a_\delta))$ , where  $m = j\delta$ , that is, a tuple containing supersets of the sets assigned to nodes that are children of nodes in position  $(i, j)$ . Since certificate trees are labeled correctly, this implies that, starting from this tuple, we compute  $S_{i,j}$  or a superset of it. Also, the Boolean that we put in the pair that we add to  $R$  is also correct, since we compute it as the or of the ones of the children. □

**Lemma 6.9** *Let  $C\mathcal{B}$  be a non-empty certificate builder obtained from Algorithm 3 for LCL problem  $\Pi$  and a label  $a$ . Then there exists a certificate of  $O(\log^* n)$  solvability  $\mathcal{T}$  with at least one leaf labeled with  $a$  (for  $a \neq \epsilon$ ) and without such restriction for  $a = \epsilon$ .*

**Proof** Before dealing with the general case, if  $\Sigma(\Pi)$  consists of only one label  $\sigma$  and we have a non-empty certificate builder  $C\mathcal{B}$ , then we also know that  $C(\Pi)$  is non-empty (see last part of Algorithm 3). Hence a certificate will be just a single tree of depth one labeled with  $\sigma$ . Hence for the rest of the proof, we assume that  $\Sigma(\Pi)$  has size at least two.

To convert a certificate builder to a certificate for  $O(\log^* n)$  solvability, we proceed in four phases. The first phase consists of creating a *verbose temporary* tree  $\mathcal{T}_{vt}$  which will be labeled with pairs with first element being sets of labels and second element being an indicator for where to find label  $a$  (that is, each node of  $\mathcal{T}_{vt}$  is labeled with an element of  $R_i$ ). Tree  $\mathcal{T}_{vt}$  is created recursively as follows:

- (1) root of  $\mathcal{T}_{vt}$  is labeled with  $(\Sigma(\Pi), a \neq \epsilon)$
- (2) each node labeled with  $(r, a)$ , where  $r$  contains at least two labels, will have  $\delta$  children labeled with  $(r_i, a_i)$  according to the (unique) pair  $((r, a), ((r_1, a_1), \dots, (r_\delta, a_\delta)))$  in  $C\mathcal{B}$  which contains pair  $(r, a)$  as its first element.
- (3) each node labeled with  $(r, a)$ , where  $r$  is a singleton set, is a leaf.

The recursive definition is legal as the labels for children will always be placed in the certificate builder earlier, so we cannot have any loop (see Algorithm 3). In the case when  $a \neq \epsilon$ , by following the indicators (second pair of label) from root of  $\mathcal{T}_{vt}$  which will be labeled with  $(\Sigma(\Pi), true)$ , we must be able to reach a leaf that is labeled  $(\{a\}, true)$ . This implies that we have a singleton  $a$  as one of the leaves.

As we don't need the second element of each label anymore, let us simplify the further analysis by creating a *simplified temporary* tree  $\mathcal{T}_{st}$  as a simplification of  $\mathcal{T}_{vt}$  where each node is labeled only by the first element from the pair. Examples of such trees are depicted in Fig. 7b and Fig. 8b.

The second phase considers the case when label  $a$  is not  $\epsilon$ . In this phase we want to “push down” a leaf node labeled with the singleton label  $a$  so it is a deepest node of the tree  $\mathcal{T}_{st}$ . We will do it as follows. Let  $n_a$  be a node in  $\mathcal{T}_{st}$  that is labeled with the singleton label  $a$ . Let  $d_a$  be its depth. Since the root of  $\mathcal{T}_{st}$  is labeled with  $\Sigma(\Pi)$ , we know that there exists a hairy path of length  $d_a$  labeled with  $\Sigma(\Pi)$  that has both of its endpoints labeled  $a$ . For convenience, let  $P_{aa}$  denote such a hairy path and replace all labels by their singleton labels (label  $\sigma$  will become  $\{\sigma\}$ ). We will use  $P_{aa}$  and replace node  $n_a$  with the path  $P_{aa}$ . We have now essentially “pushed down” a leaf node labeled with the singleton label  $a$  by  $d_a$  steps. We will repeat such “pushing down” until we have that  $n_a$  is the deepest node of the tree  $\mathcal{T}_{st}$ . To summarize, now we have a tree  $\mathcal{T}_{st}$  that has its deepest leaf labeled with the singleton  $a$ .

In the third phase, we want to make all leaves to be on the same level. We do it as follows. Again, observe that since  $\Sigma(\Pi)$  is the root of  $\mathcal{T}_{st}$ , we have a continuation below for every label from  $\Sigma(\Pi)$ . We can use such continuation to “push down” every leaf node of  $\mathcal{T}_{st}$  that is not the deepest by one step in the same manner as in the previous phase. We replace a leaf node  $n_l$  labeled with a singleton  $\sigma$  with a  $\delta$ -ary tree of depth one, labeled with singletons corresponding to a continuation below for  $\sigma$ . To summarize, now we have a tree  $\mathcal{T}_{st}$  that has all leaves at the same level. Examples of such trees are again depicted in Fig. 7b and Fig. 8b.

Finally, in the last phase, we use  $\mathcal{T}_{st}$  to build  $|\Sigma(\Pi)|$  individual labeled trees  $\mathcal{T}_{st}$  that would form a certificate. At the beginning, let each  $\mathcal{T}_i$  be labeled exactly as  $\mathcal{T}_{st}$ . Then, for each  $\mathcal{T}_{st}$ , we fix its root label to a distinct label  $\sigma_i$ . Then, we recursively fix the labels of the children such that the resulting configuration is in  $C(\Pi)$ . Such a configuration will always exist as it is how we have constructed the certificate builder (see Algorithm 3). Examples of the obtained trees are depicted in Fig. 7c and Fig. 8c.  $\square$

We now prove an upper bound on the running time of Algorithm 4.

**Theorem 6.10** *The running time of Algorithm 4 is at most exponential in the size of the LCL problem.*

**Proof** Observe that every iteration of the for loop in Algorithm 3 either adds an element to  $R_i$  or finishes the algorithm. Hence, we can upper bound the number of iterations by the maximum size of each set  $R_i$ , that is  $2^{|\Sigma|+1}$ . Also, each iteration requires at most exponential time in  $\Sigma$  and  $\delta$ . Hence, the total running time of Algorithm 3 is exponential in  $\Sigma$  and  $\delta$ . Since Algorithm 4 just calls Algorithm 3 for every choice over  $\Sigma$ , then we get one more exponential slowdown, hence the claim follows.  $\square$

Hence we conclude the following theorem.

**Theorem 6.11** *Whether an LCL problem  $\Pi$  has round complexity  $O(\log^* n)$  or  $\Omega(\log n)$  can be decided in time at most exponential in the size of the LCL problem.*

## 7 Sub-log-star region

In this section we prove that there is no LCL problem  $\Pi$  with distributed time complexity between  $\omega(1)$  and  $o(\log^* n)$ . Also, we prove that, given a problem  $\Pi$ , we can *decide* if its complexity is  $O(1)$  or  $\Omega(\log^* n)$ . Moreover, we prove that randomness cannot help: if a problem has *randomized* complexity  $O(1)$ , then it has the same *deterministic* complexity.

### 7.1 High-level idea

We prove that deciding if a problem  $\Pi$  can be solved in constant time is surprisingly simple: a problem is  $O(1)$  rounds solvable if and only if it can be solved in  $O(\log^* n)$  rounds and  $\Pi$  contains an allowed configuration of a specific form (this configuration will be called *special*). This configuration must allow a node to use the same label  $\ell$  that one of its children uses, the labels used by this configuration should be contained in the ones used by some certificate for  $O(\log^* n)$  solvability, and  $\ell$  should be used by at least one leaf of the certificate. If we consider the definition of the MIS problem given in Sect. 1.3, we can see that it allows the configuration  $(b : b1)$ , and informally this configuration is what makes the problem constant-time solvable. Note that, however, the algorithm that we can obtain by using this certificate, while still being constant time, may have a worse complexity compared to the one described in Sect. 1.3. On the other hand, we can see that in the definition of the 3-coloring problem given in Sect. 1.2 there is no configuration of this form, and this is what makes the problem  $\Omega(\log^* n)$ .

Informally, the reason is the following. The  $n$  appearing in  $O(\log^* n)$  complexities does not usually refer to the size of the graph, but to the range of the identifiers assigned to the nodes. In fact, in the proof of Theorem 6.3,  $O(\log^* n)$  is spent only to compute some ruling set, while the rest only requires constant time, and in order to compute such a ruling set, a

distance- $k$  coloring, for some large enough  $k$ , is sufficient. Unfortunately, it is not possible to compute a distance- $k$  coloring in constant time, but as we will show, some defective coloring (that is, a coloring that allows some neighbors of a node to use the same color of the node) will be sufficient for our purposes. We show that in constant time we can produce some defective distance- $k$  coloring, for some large enough constant  $k$ , such that:

- we can label defective nodes with the special configuration,
- unlabeled nodes are properly colored, and
- labeled nodes that are in different connected components are far enough from each other.

We can then complete the partial labeling in constant time with the help of the certificate, similarly to how we use the certificate of  $O(\log^* n)$  solvability to solve problems in  $O(\log^* n)$  rounds, but this time we can speed the computation of the ruling set up, and make it run in constant time by exploiting the defective distance- $k$  coloring. In the other direction, we show that if the special configuration does not exist, or if it does not satisfy the required properties, then any algorithm solving the problem can also be used to solve the coloring problem with a constant size palette, that is known to require  $\Omega(\log^* n)$  rounds.

### 7.2 Certificate

We start by defining what is a certificate for  $O(1)$  solvability, that is nothing else but a certificate for  $O(\log^* n)$  solvability and a configuration of some specific form. An example of such a certificate for the MIS problem is depicted in Fig. 8.

**Definition 7.1** Let  $\Pi$  be an LCL problem. A certificate for  $O(1)$  solvability for problem  $\Pi$  is a pair  $\mathcal{S}$  consisting of a certificate for  $O(\log^* n)$  solvability  $\mathcal{T}$  and a configuration  $(a : b_1, \dots, a, \dots, b_\delta) \in \mathcal{C}(\Pi)$  where  $a, b_i \in \Sigma_{\mathcal{T}}$  and at least one leaf of the trees in  $\mathcal{T}$  is labeled  $a$ .

### 7.3 Upper bound

We now prove that we can use a certificate for  $O(1)$  solvability to construct an algorithm that solves the problem  $\Pi$  in constant time. Informally, we first spend a constant number of rounds to try to construct some distance- $k$  coloring. This coloring cannot always be correct, since the coloring problem requires  $\Omega(\log^* n)$  rounds. We will use the special configuration to label nodes in which the coloring procedure failed. The coloring will also satisfy some desirable property, such as having improperly colored regions that are far enough from each other. This will give us a proper distance- $k$  coloring in the unlabeled regions, and we will use this col-

oring to complete the labeling in constant time. The proof of this theorem will use some useful lemmas that we will prove later.

**Theorem 7.2** Any LCL problem  $\Pi$  that has a certificate of  $O(1)$  solvability is constant-time solvable with a deterministic CONGEST algorithm.

**Proof** We prove the statement by constructing a constant-time algorithm with the help of a special configuration  $(a : b_1, \dots, a, \dots, b_\delta)$  and  $O(\log^* n)$  certificate  $\mathcal{T}$  of depth  $d$  where at least one leaf of the trees in  $\mathcal{T}$  is labeled  $a$ . Let  $k = 20d + 1$ .

For each node  $v$ , let  $p(v) \in \{1, \dots, \delta\}$  be the index of  $v$  in the sorted sequence containing the identifier of  $v$  and all its siblings (that is,  $p(\cdot)$  emulates port numbers). We start by assigning a (possibly non-proper) coloring  $c(v)$  to each node  $v$ , as follows. Consider the sequence  $(v_i, i \geq 0)$  of nodes obtained by starting from  $v_0 = v$  and following edges going up. The color  $c(v)$  is defined as  $(p(v_0), p(v_1), \dots, p(v_{10k-1}))$ . If  $p(v_i)$  is undefined because  $v_i$  does not exist, that is, while going up we found the root, then we complete the sequence with 1s (this is equivalent to imagine the rooted tree to be embedded into a larger rooted tree, where all nodes of the original are far enough from the root of the larger tree). This can be done in constant time in CONGEST, by first computing and sending the  $p(\cdot)$  value of all children, and then repeatedly propagating down the  $p(\cdot)$  value received from the parent. We say that a prefix of length  $x$  of a color  $c = c(v)$  has period  $r$  if and only if  $c[i] = c[i+r]$  for all  $i < x - r$ .

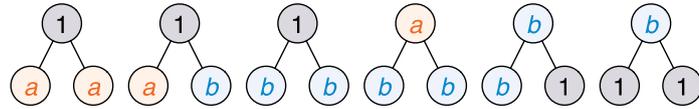
We define a *vertical path* to be a subpath of a root-to-leaf path. We mark all nodes  $v$  satisfying that the prefix of length  $9k$  of  $c(v)$  has period at most  $k$ . By Corollary 7.5, the connected components induced by marked nodes form vertical paths. We use the configuration  $(a : b_1, \dots, a, \dots, b_\delta)$  to label all marked nodes with  $a$ , and all their children with the other labels of the configuration in some arbitrary consistent manner.

By Corollary 7.4, all connected components of the marked nodes are at distance at least  $k$ , and by Lemma 7.6, all unmarked regions are distance- $k$  colored. We now show how to complete the partial labeling assigned to the nodes. We will start by splitting the unlabeled parts of the tree into constant size subtrees, by exploiting the distance coloring to perform this step in constant time. Then, we will operate in a constant number of rounds on these subtrees in parallel.

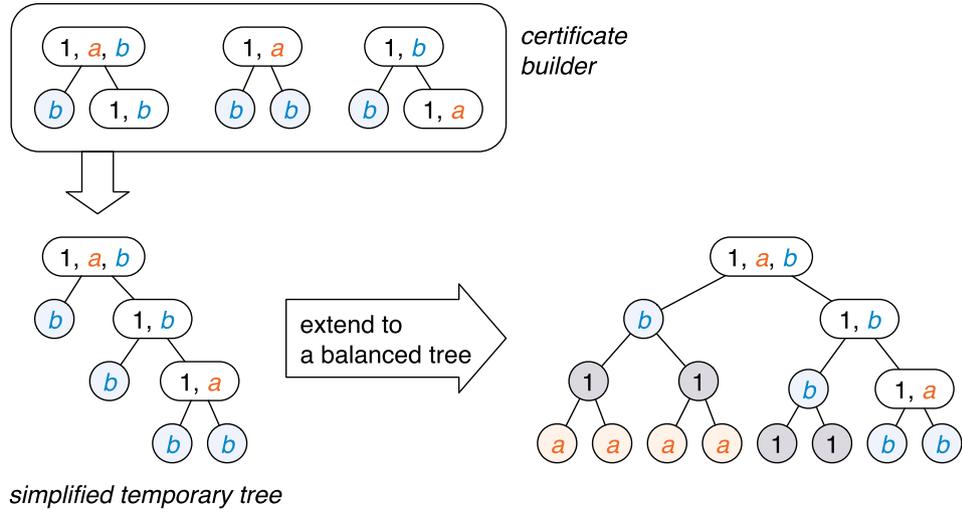
Let us now focus on the first phase, namely splitting the trees. We will compute a splitting of the unlabeled regions satisfying the following:

- Subtrees have overlapping *boundaries*, meaning that each leaf of a subtree is also the root of another subtree.
- The distance of each leaf from the root is in  $\{d, \dots, 10d\}$ .

(a) Problem: maximal independent set in binary trees:



(b) Finding a certificate where one of the leaf nodes is labeled with *b*:



(c) Certificate for  $O(1)$ -round solvability:

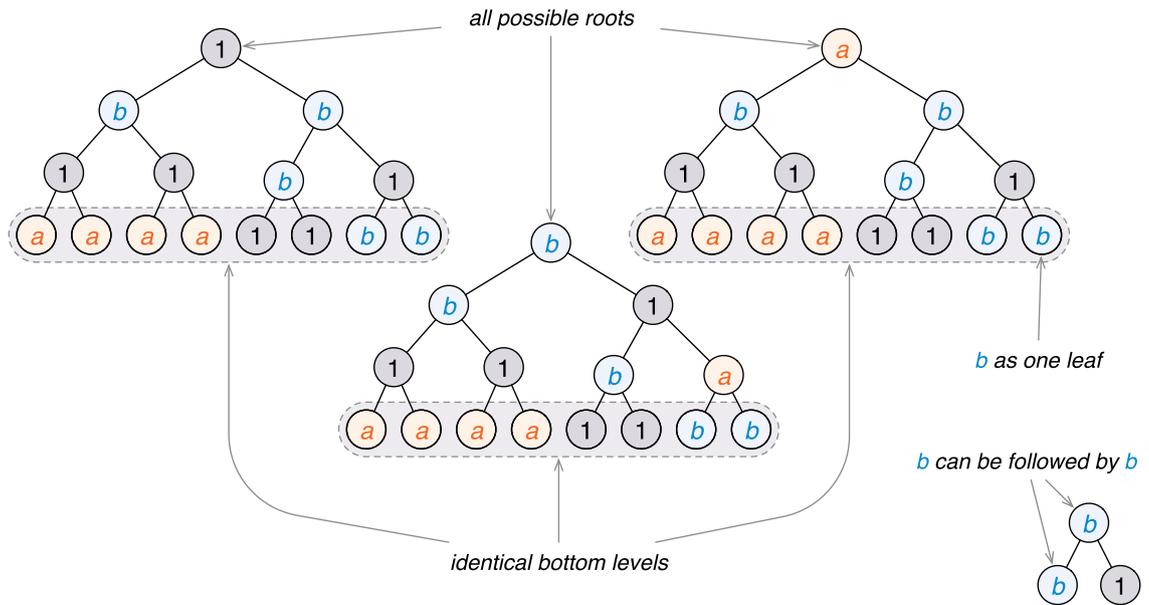


Fig. 8 Finding a certificate for  $O(1)$  solvability (Definition 7.1) for the maximal independent set problem (Sect. 1.3)

- All inner nodes of each subtree are unlabeled.

Consider the following problem, which we call *ruling set extension problem* (see Fig. 9a): We are given a directed path and some set  $S$  satisfying that the nodes in  $S$  are at distance at least  $2d$  from each other. The goal is to compute a set  $S'$  such that each node not in  $S \cup S'$  has a node in  $S \cup S'$  at distance at most  $4d$ , and for each node in  $S'$  it holds that the closest successor in  $S \cup S'$  is at distance at least  $2d$ . That is,  $S \cup S'$  is almost a  $(2d, 4d)$ -ruling set: nodes in the given set  $S$  can violate the ruling set requirements, since they could have a successor in  $S'$  at distance less than  $2d$ .

We show that the ruling set extension problem can be solved in  $O(1)$  rounds on directed paths by using a *one-sided* algorithm, provided that we are given a distance- $k$  coloring with  $O(1)$  colors for a sufficiently large  $k$ . A one-sided algorithm is an algorithm in which nodes only send messages to their predecessors (or equivalently, nodes only receive information from their successors). One-sided algorithms are convenient, as they are directly applicable in rooted trees [17]: if we have a one-sided algorithm that finds a ruling set extension in directed paths, we can apply the same algorithm in rooted trees and it will produce an output in which all root-to-leaf paths satisfy the constraints of the ruling set extension problem. Our one-sided algorithm works as follows (see Fig. 9b):

- (i) Compute  $S'$  by using a  $(2d, 2d)$ -ruling set algorithm (ignoring  $S$  entirely).
- (ii) Nodes in  $S'$  that have a successor in  $S$  within distance  $2d$  are removed from  $S'$ .

We obtain that nodes in  $S$  have no predecessors in  $S \cup S'$  within distance  $2d$ , nodes in  $S'$  are at distance at least  $2d$  from each other, and nodes not in  $S \cup S'$  have at least a node in  $S \cup S'$  within distance  $4d$ , and hence a solution for the ruling set extension problem. Here step (ii) is easy to implement with a one-sided algorithm. There is also a simple two-sided algorithm  $A$  for solving step (i): process the nodes by color classes; whenever we consider a particular node, check if there is already another node within distance  $2d$  that we have selected, and if not, select the node. Finally, we use the standard trick of “shifting the output” [17] to turn  $A$  into a one-sided algorithm  $A'$  that solves the same problem, as follows: Let  $T = O(1)$  be the running time of algorithm  $A$ , and let us label the nodes by  $v_1, v_2, \dots$  along the path. In algorithm  $A'$  node  $v_i$  will output whatever node  $v_{i+T}$  outputs in algorithm  $A$ . The output of node  $v_{i+T}$  in algorithm  $A$  only depends on the input colors of nodes  $v_i, v_{i+1}, \dots, v_{i+2T}$ , and hence one-sided information is sufficient for  $A'$  to simulate  $A$ . We have simply shifted the ruling set by  $T$  steps.

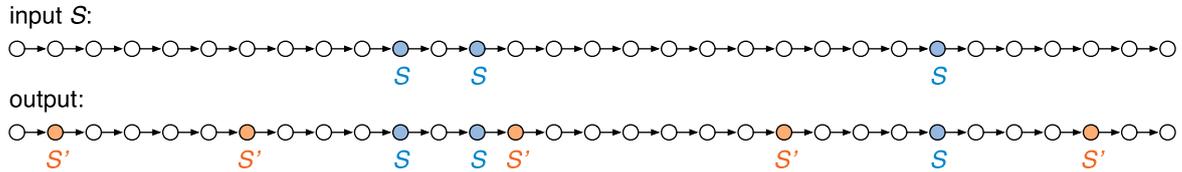
Let us now get back to the task of splitting trees. We proceed as follows. We add all nodes whose label is fixed to a

set  $S$ . Then, we run the one-sided algorithm for the ruling set extension problem on the subgraph induced by unlabeled nodes and their neighbors. Let  $S'$  be the output of the algorithm. As discussed, in any root-to-leaf path we obtain a solution for the problem described above, and observe that this implies that, in any root-to-leaf path, nodes in  $S \cup S'$  have at least one successor and one predecessor in  $S \cup S'$  at distance at most  $8d + 1 < 10d$ . What we obtained almost satisfies the requirements of the splitting, except that some nodes of  $S'$  may be too near to the nodes that were already in the set, since the minimum distance is guaranteed only while following successors (that is, by going up), and we now fix this issue.

Let  $P \subseteq S'$  be the set of nodes that have an already labeled (that is, nodes in  $S$ ) node as one of its descendants at distance less than  $d$ . For each such node  $n_p$  from  $P$  we do the following. Let  $d_c$  denote the distance from node  $n_p$  to its closest labeled node  $n_c$  below. We remove  $n_p$  from  $S'$  and add all nodes that are descendants of  $n_p$  at distance exactly  $d_c$  to the set  $S'$ . Note that, as the connected components of labeled nodes are strictly more than  $20d$  steps apart (by definition of  $k$ ), all nodes  $n_u$  just added to  $S'$  (except node  $n_c$ ) will *not* have any node of  $S$  below it that is closer than  $20d - \text{distance}(n_u, n_c) \geq 18d$  steps. The nodes added to  $S'$  will be closer to the nodes of  $S'$  below, but still at least  $2d - d = d$  far away, as distance between two nodes of  $S'$  on any root-to-leaf path was originally always at least  $2d$ . Similarly, they will be further away from a node in  $S'$  that is above them, but again at most distance  $8d + 1 + d = 9d + 1 \leq 10d$  far away. Now, observe that the nodes in  $S \cup S'$  partition the input tree in subtrees with the required properties.

We now describe the second phase. For that, we use the certificate  $\mathcal{T}$ , and for each subtree in parallel, we do the following. First, we check whether it has one of its leaves already fixed. If not, we directly fix labels at depth  $d$  exactly as a leaf layer of any tree from the certificate (they are by definition the same). Otherwise, more care is needed. Observe that each subtree can have at most one fixed leaf (labeled with  $a$ ). In fact, in any tree whose depth is upper bounded by  $10d$ , any two nodes are at most distance  $20d$  apart (the distance is upper bounded by the length of a walk that starts from a node, goes to the root, and then goes to other node); but as fixed nodes are strictly more than  $20d$  apart (by the definition of  $k$ ), having one fixed leaf means that all other leaves are unlabeled. Let such fixed leaf be denoted by  $n_f$  and also denote by  $n_d$  the node that is on the path connecting  $n_f$  to the root at distance  $d$  from the root. We fix labels at depth  $d$  exactly as a leaf layer of any tree from the certificate such that node  $n_d$  will be labeled with label  $a$ . This is possible as we can freely choose the ordering of the children. Then, we use the configuration  $(a : b_1, \dots, a, \dots, b_\delta)$  to label the nodes of the hairy path connecting  $n_d$  to  $n_f$ .

(a) Ruling set extension problem (for  $d = 1$ ):



(b) One-sided algorithm that solves the ruling set extension problem (for  $d = 1$ ):

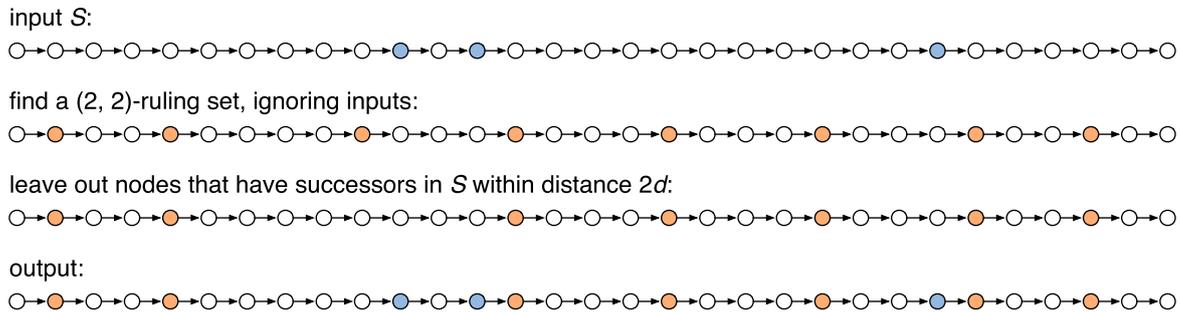


Fig. 9 Ruling set extension problem

Now, we have fixed layer  $d$  of all subtrees, and also some of the nodes below layer  $d$ . We proceed from layer  $d$  and towards lower layers and label all of its children arbitrarily as every label has a continuation below. This procedure will stop after constant time as our trees have constant depth.

The only remaining part is to fix labels for nodes that are between the roots and the nodes at layer  $d$ . For that, we use the certificate trees and for every subtree with root fixed to label  $\sigma_i$ , we use tree  $\mathcal{T}_i$  to label the upper layers.  $\square$

We now prove that if marked nodes are at distance at most  $k$ , then they must lie in the same vertical path. Intuitively, if two nodes  $x_1$  and  $x_2$  are siblings, then  $p(x_1) \neq p(x_2)$ , implying that they cannot both have a periodic color, and that the same must hold for the descendants of  $x_1$  and  $x_2$ , up to some distance. Hence, we can find nodes with a periodic color only by following vertical paths.

**Lemma 7.3** *If two marked nodes are at distance at most  $k$ , then they are in the same connected component of marked nodes, and each connected component forms a vertical path.*

**Proof** We will prove the statement by contradiction. Suppose that  $v_1$  and  $v_2$  are two marked nodes from different connected components that are at distance  $d \leq k$  and the prefix of length  $9k$  of their colors has period at most  $k$ . Let  $k_1$  and  $k_2$  be, respectively, the period of the colors of  $v_1$  and  $v_2$ . Since they are at distance  $d \leq k$  from each other, then their lowest common ancestor  $v_3$  is at distance at most  $k$  from both. Let  $d_1$  and  $d_2$  be, respectively, the distance of  $v_1$  and

$v_2$  from  $v_3$ . Since a prefix of length at least  $8k$  of  $c(v_3)$  is equal to suffixes of length at least  $8k$  of  $c(v_1)$  and  $c(v_2)$ , then the prefix of length  $8k$  of  $c(v_3)$  has also period  $k_3$  that is at most  $k$ , and that satisfies  $k_3 = k_1 = k_2 = k'$ . We now prove that either  $v_1 = v_3$  or  $v_2 = v_3$ . Assume it is not the case, then there must be two children  $x_1$  and  $x_2$  of  $v_3$  that lie in the two paths connecting  $v_3$  to  $v_1$  and  $v_2$ . Since  $v_3$  is the lowest common ancestor,  $x_1 \neq x_2$ , and since  $x_1$  and  $x_2$  are siblings, then  $p(x_1) \neq p(x_2)$ . Since  $p(x_1) = c(v_1)[d_1 - 1]$ , and  $p(x_2) = c(v_2)[d_2 - 1]$ , then by the period assumption  $c(v_1)[d_1 - 1 + k'] \neq c(v_2)[d_2 - 1 + k']$ , which is a contradiction since by going up from  $v_1$  for  $d_1 - 1 + k'$  steps we reach the same node that we reach by going up from  $v_2$  for  $d_2 - 1 + k'$  steps. Hence,  $v_1$  and  $v_2$  lie in the same vertical path. Also, note that all nodes in the vertical path between  $v_1$  and  $v_2$  will be marked as well, since the prefix of length  $9k$  of their color is contained in the union of the prefixes of  $v_1$  and  $v_2$ , which have period at most  $k$  in their  $9k$ -length prefixes.  $\square$

This lemma implies the following corollaries.

**Corollary 7.4** (Marked regions are far apart) *For all marked nodes  $v_1$  and  $v_2$  from different regions, their distance is strictly larger than  $k$ .*

**Corollary 7.5** (Marked region is a vertical path) *Every marked region forms a vertical path.*

We now prove that, if we consider the subgraph induced by unmarked nodes, the computed coloring  $c$  forms a proper distance- $k$  coloring.

**Lemma 7.6** (Unmarked regions form a distance- $k$  coloring) *Let  $C$  be a connected component of unmarked nodes. Then labels of nodes in  $C$  form a distance- $k$  coloring.*

**Proof** By contradiction. Suppose that  $v_1$  and  $v_2$  are two different unmarked nodes having the same color  $c$  and distance  $d < k$ . We will show that there is a marked node on the path between  $v_1$  and  $v_2$ , contradicting that  $v_1$  and  $v_2$  are from the same connected component. The distance to the lowest common ancestor (denoted by  $v_3$ ) is at most  $d$  for both nodes (the distance will be denoted by  $d_1$  and  $d_2$  for nodes  $v_1$  and  $v_2$  respectively). W.l.o.g. we assume that  $d_1 < d_2$  (because of symmetry and the fact that  $d_1 = d_2$  would contradict that nodes  $v_1$  and  $v_2$  have the same color and are different). As the path upwards from node  $v_3$  is the same for both nodes, we obtain that  $c[d_1 + i] = c[d_2 + i]$  for all  $0 \leq i < 10k - d_2$ . If we look at color  $c'$  of the lowest common ancestor, we obtain from the previous equalities that  $c'[i] = c'[d_2 - d_1 + i]$  for all  $0 \leq i < 10k - d_2$ , hence the prefix of length  $9k$  of  $c'$  has period at most  $d_2 - d_1 \leq k$ . Hence node  $v_3$  would be marked and on a path from  $v_1$  to  $v_2$  contradicting that they are from the same connected component.  $\square$

### 7.4 Lower bound

We now prove that, if a certificate for  $O(1)$  solvability does not exist, then the problem requires  $\Omega(\log^* n)$ , even for randomized algorithms. On a high level, we can prove that if there is no  $O(\log^* n)$  algorithm that can use the special configuration, then it means that we can convert any solution for  $\Pi$  into a proper coloring, implying that  $\Pi$  requires  $\Omega(\log^* n)$ .

**Theorem 7.7** *Let  $\Pi$  be an LCL problem for which no certificate for  $O(1)$  solvability exists. Then, the randomized and deterministic complexity of  $\Pi$  in the LOCAL model is  $\Omega(\log^* n)$ .*

**Proof** We consider two possible cases: either there is a configuration of the form  $(a : b_1, \dots, a, \dots, b_\delta)$  or not. In the latter case, each solution for  $\Pi$  is such that all nodes have a label that is different from the labels of the neighbors, meaning that we can interpret such a labeling as a coloring from a constant size palette. Since an algorithm for  $O(1)$ -coloring  $\delta$ -ary rooted trees could be simulated in directed paths, by imagining  $\delta - 1$  additional nodes connected to each node of the path, and since  $O(1)$ -coloring in paths is known to require  $O(\log^* n)$  rounds, even for randomized algorithms [21,23], then the claim follows.

Hence, assume that there is a configuration of the form  $(a : b_1, \dots, a, \dots, b_\delta)$ , but there does not exist a certificate for

$O(\log^* n)$  solvability that contains all the labels of the special configuration, and that  $a$  is the label of at least one leaf. We prove that for any algorithm  $A$  that solves  $\Pi$ , there exists some  $n_0$ , such that for any  $n > n_0$ , algorithm  $A$  running on any instance of size  $n$  must label all nodes that are at  $\omega(\log^* n)$  distance from the root and from any leaf such that they have a different label from all their neighbors. In other words, algorithm  $A$  computes a proper coloring in the intermediate layers.

Assume it is not the case, then there must exist an algorithm  $A$  such that, for any  $n_0$ , there exists some  $n > n_0$  such that on some instances of size  $n$  it labels at least one node that is at  $\omega(\log^* n)$  distance from the root and from any leaf by some configuration of the form  $(a : b_1, \dots, a, \dots, b_\delta)$ . This implies that also each label in  $\{b_1, \dots, a, \dots, b_\delta\}$  is used by at least one node that is at  $\omega(\log^* n)$  distance from the root and from any leaf, and hence, the requirements of Lemma 6.4 apply. By applying Lemma 6.4 with some  $n_0$  large enough guaranteed to exist by the running time of the algorithm, we get that there exists a certificate for  $O(\log^* n)$  solvability that contains all labels in  $\{b_1, \dots, a, \dots, b_\delta\}$  and that uses  $a$  in at least one leaf, contradicting the fact that there is no certificate for  $O(1)$  solvability.

Hence, in any solution for  $\Pi$  constructed by an algorithm running in  $O(\log^* n)$  rounds all nodes that are at  $\omega(\log^* n)$  distance from the root and from any leaf are labeled such that they have a different label from all their neighbors, and hence that all these nodes are properly colored, in any instance that is large enough. Hence we can use any  $O(\log^* n)$  algorithm for  $\Pi$  to solve  $O(1)$ -coloring in paths, by creating a virtual graph in which we connect large enough trees to each node of the path and extend the path on the endpoints, such that no node sees any root or leaf, and then running the algorithm. This implies that  $\Pi$  requires  $\Omega(\log^* n)$  rounds, and hence the claim follows.  $\square$

### 7.5 Decidability

The only additional requirement for a problem that is  $O(\log^* n)$  solvable to be constant-time solvable is the existence of configuration  $(a : b_1, \dots, a, \dots, b_\delta)$  where  $a, b_i$  are from  $\Sigma_{\mathcal{T}}$ , that are the labels used by the certificate  $\mathcal{T}$ , and at least one leaf in  $\mathcal{T}$  is labeled  $a$ .

Algorithm 3 allows us to search for a certificate builder containing a specific leaf, and by Lemma 6.9 a certificate builder of this form implies a certificate of the same form. Hence, we can just augment Algorithm 4 to additionally search only for a certificate builder that would satisfy having a configuration of the form  $(a : b_1, \dots, a, \dots, b_\delta)$  consisting of certificate labels, such that  $a$  appears in at least one leaf. This is done in Algorithm 5.

**Algorithm 5:** constantCertificate( $\Pi$ )

**Input:** LCL problem  $\Pi$   
**Output:**  $\epsilon$  if no certificate exists, or a certificate builder

**for** all subsets  $\Sigma'$  of  $\Sigma(\Pi)$  **do**  
     $\Pi' \leftarrow$  restriction of  $\Pi$  to labels from  $\Sigma'$   
    **for** all configurations of the form  $(a : b_1, \dots, a, \dots, b_\delta)$  in  $\Pi'$  **do**  
        certificateBuilder  $\leftarrow$   
        findUnrestrictedCertificate( $\Pi', a$ )  
        **if** certificateBuilder  $\neq \epsilon$  **then**  
            **return** certificateBuilder  
        **end**  
    **end**  
**end**  
**return**  $\epsilon$

**Theorem 7.8** *The running time of Algorithm 5 is exponential in the size of the LCL problem.*

**Proof** Follows by using the same arguments as for Algorithm 4 (Theorem 6.10).  $\square$

**Theorem 7.9** *Algorithm 5 outputs a certificate builder if and only if an LCL problem has a certificate for  $O(1)$  solvability.*

**Proof** Since Algorithm 5 tries to find a certificate builder for all subsets of labels for which a configuration of the form  $(a : b_1, \dots, a, \dots, b_\delta)$  exists, then the statement follows by using the same arguments as in Theorem 6.8.  $\square$

Hence we conclude the following theorem.

**Theorem 7.10** *Whether an LCL problem  $\Pi$  has round complexity  $O(1)$  or  $\Omega(\log^* n)$  can be decided in time at most exponential in the size of the LCL problem.*

**8 Polynomial region**

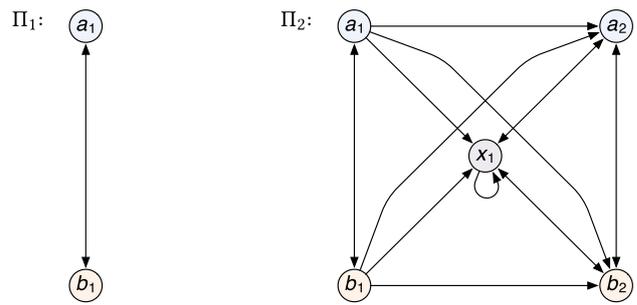
In this section, we describe an infinite sequence of LCL problems  $\Pi_1, \Pi_2, \dots$  with  $\delta = 2$  such that the complexity of  $\Pi_k = (2, \Sigma_k, C_k)$  is  $\Theta(n^{1/k})$  in both of the LOCAL and CONGEST models, as the lower bound applies to LOCAL and the upper bound applies to CONGEST.

The alphabet  $\Sigma_k$  for  $\Pi_k$  is

$$\Sigma_k = \{a_1, b_1, x_1, a_2, b_2, x_2, \dots, a_k, b_k\},$$

and the set of permitted configurations  $C_k$  for  $\Pi_k$  is defined as follows.

- For  $1 \leq i \leq k$ , add  $(a_i : \sigma, \sigma')$  to  $C_k$  for all  $\sigma, \sigma' \in \{a_1, b_1, x_1, a_2, b_2, x_2, \dots, a_{i-1}, b_{i-1}, x_{i-1}\} \cup \{b_i\}$ .
- For  $1 \leq i \leq k$ , add  $(b_i : \sigma, \sigma')$  to  $C_k$  for all  $\sigma, \sigma' \in \{a_1, b_1, x_1, a_2, b_2, x_2, \dots, a_{i-1}, b_{i-1}, x_{i-1}\} \cup \{a_i\}$ .



**Fig. 10** The automata associated with the path-forms of  $\Pi_1$  and  $\Pi_2$

- For  $1 \leq i \leq k - 1$ , add  $(x_i : \sigma, \sigma')$  to  $C_k$  for all  $\sigma \in \Sigma_k$  and  $\sigma' \in \{a_1, b_1, x_1, a_2, b_2, x_2, \dots, a_i, b_i\}$ .

When  $k = 1$ ,  $\Pi_1$  is exactly the proper 2-coloring problem with the two colors  $\Sigma_1 = \{a_1, b_1\}$ . When  $k = 2$ ,  $\Pi_2$  is a combination of two proper 2-coloring problems with the color sets  $\{a_1, b_1\}$  and  $\{a_2, b_2\}$  via the special label  $x_1$ . Whenever a node  $v$  is labeled  $x_1$ , it must have at least one child  $u$  such that the entire subtree rooted at  $u$  is properly 2-colored by  $\{a_1, b_1\}$ . For general  $k$ ,  $\Pi_k$  can be seen as a combination of  $k$  proper 2-coloring problems. See Fig. 10 for an illustration.

**Lemma 8.1** *For each positive integer  $k$ , the round complexity of  $\Pi_k$  is  $O(n^{1/k})$  in the CONGEST model.*

**Proof** Let  $V$  be the node set for a given  $n$ -node rooted tree  $T$ . We show that in  $O(n^{1/k})$  rounds, we can partition the set of nodes  $V$  into  $2k - 1$  parts

$$V = B_1 \cup X_1 \cup B_2 \cup X_2 \cup \dots \cup X_{k-1} \cup B_k$$

satisfying the following properties.

- (P1) For each  $1 \leq i \leq k$ , each connected component of  $B_i$  has at most  $O(n^{1/k})$  nodes.
- (P2) For each  $1 \leq i \leq k - 1$ , at least one child of each  $v \in X_i$  is in  $B_1 \cup X_1 \cup B_2 \cup X_2 \cup \dots \cup B_{i-1} \cup X_{i-1} \cup B_i$ .
- (P3) For each  $1 \leq i \leq k$ , the children of each  $v \in B_i$  are in  $B_1 \cup X_1 \cup B_2 \cup X_2 \cup \dots \cup X_{i-1} \cup B_i$ .

Once we have this partition,  $\Pi_k$  can be solved in  $O(n^{1/k})$  rounds by assigning  $x_i$  to each  $v \in X_i$  and labeling each connected component of  $B_i$  by an arbitrary proper 2-coloring with  $\{a_i, b_i\}$ .

The algorithm for computing the partition  $V = B_1 \cup X_1 \cup B_2 \cup X_2 \cup \dots \cup X_{k-1} \cup B_k$  has  $k$  iterations. We assume that at the beginning of iteration  $i$ , we have already computed the parts  $B_1, X_1, B_2, X_2, \dots, B_{i-1}, X_{i-1}$  in such a way that the set of remaining nodes  $U_i = V \setminus (B_1 \cup X_1 \cup B_2 \cup X_2 \cup \dots \cup B_{i-1} \cup X_{i-1})$  satisfies the following induction hypothesis.

- (IH) For each  $1 \leq i \leq k$ , we have  $|U_i| \leq n^{1-(i-1)/k}$ .

Note that (IH) holds initially for  $i = 1$ , as  $U_1 = V$  and  $|V| = n$ .

Suppose we are at the beginning of iteration  $i$ . Consider the subtree  $T_i$  induced by the remaining nodes  $U_i$ . For each  $v \in U_i$ , we write  $N_v$  to denote the number of nodes in the subtree of  $T_i$  rooted at  $v$ . We compute the two parts  $B_i$  and  $X_i$  as follows. It is clear that the computation takes  $O(n^{1/k})$  rounds.

- If  $i = k$ , then  $B_i = U_i$ .
- If  $1 \leq i < k$ , then  $B_i$  is the set of nodes  $v \in U_i$  with  $N_v \leq n^{1/k}$ .
- If  $1 \leq i < k$ , then  $X_i$  is the set of nodes  $v \in U_i$  with  $N_v > n^{1/k}$  that satisfy at least one of the following.
  - $N_u \leq n^{1/k}$  for at least one child  $u$  of  $v$ .
  - $v$  has exactly one child in  $T_i$ .

It is straightforward to verify that the properties (P1), (P2), and (P3) are satisfied for  $B_i$  and  $X_i$ .

- Consider the first property (P1). If  $i = k$ , then clearly  $|B_k| = |U_k| \leq n^{1/k}$  by (IH). If  $1 \leq i < k$ , then each  $v \in B_i$  can have at most  $n^{1/k}$  descendants in  $T_i$ , including  $v$  itself, so each connected component of  $B_i$  has at most  $n^{1/k}$  nodes.
- The second property (P2) follows from the definition of  $X_i$ . There are two cases for each  $v \in X_i$ . The first case is that there is a child  $u$  of  $v$  with  $N_u \leq n^{1/k}$ , so  $u \in B_i$ . The second case is that  $v$  has exactly one child in  $T_i$ , so the other child  $u$  of  $v$  is in  $V \setminus U_i = B_1 \cup X_1 \cup B_2 \cup X_2 \cup \dots \cup B_{i-1} \cup X_{i-1}$ . In both cases, (P2) is satisfied.
- For the third property (P3), consider any child  $u$  of a node  $v \in B_i$ . If  $u \in U_i$ , then the definition of  $B_i$  ensures that  $u \in B_i$ . Otherwise  $u \in B_1 \cup X_1 \cup B_2 \cup X_2 \cup \dots \cup B_{i-1} \cup X_{i-1}$ . In both cases, (P3) is satisfied.

For the rest of the proof, we consider the case  $1 \leq i < k$  and we will show that  $|U_i \setminus (B_i \cup X_i)| = |U_{i+1}| \leq n^{1-i/k}$ , so the induction hypothesis (IH) holds for  $U_{i+1}$ . Observe that each  $v \in U_{i+1}$  must have exactly two children in  $U_i$ , so the set  $W$  of nodes  $v \in U_i \setminus U_{i+1}$  whose parent belongs to  $U_{i+1}$  has size  $|W| \geq |U_{i+1}|$ . It is clear that  $W \subseteq X_i$ . Since each  $v \in X_i$  has  $N_v > n^{1/k}$ , we can lower bound the size of  $U_i$  by  $|U_i| \geq \sum_{v \in W} N_v > |W|n^{1/k} \geq |U_{i+1}|n^{1/k}$ , so  $|U_{i+1}| < |U_i|n^{-1/k} \leq n^{1-i/k}$ , as  $|U_i| \leq n^{1-(i-1)/k}$  by (IH) for  $U_i$ . □

**Lemma 8.2** *For each positive integer  $k$ , the round complexity of  $\Pi_k$  is  $\Omega(n^{1/k})$  in the LOCAL model.*

**Proof** Observe that, given the LCL problem  $\Pi_k$ , Algorithm 2 takes exactly  $k$  iterations to output  $\epsilon$ . For the first iteration, the labels  $\{a_1, b_1\}$  are path-inflexible in  $\Pi_k$ . For iteration  $1 <$

$i \leq k$ ,  $\{x_{i-1}, a_i, b_i\}$  are path-inflexible in  $\Pi_k$  restricted to the labels  $\Sigma_k \setminus \{a_1, b_1, x_1, a_2, b_2, x_2, \dots, a_{i-1}, b_{i-1}\}$ . Therefore, the round complexity of  $\Pi_k$  is  $\Omega(n^{1/k})$  by Lemma 5.14. □

Combining Lemma 8.1 and Lemma 8.2, we conclude the following theorem.

**Theorem 8.3** *For each positive integer  $k$ , the round complexity of  $\Pi_k$  is  $\Theta(n^{1/k})$  in both LOCAL and CONGEST.*

## 9 Future work

While we completely characterize all complexities for LCLs in rooted trees in both LOCAL and CONGEST, for both deterministic and randomized algorithms, and we show that we can decide what is the complexity of a given problem, there are many questions that are left open.

The first question regards the running time of the algorithm that tries to find a certificate for  $O(\log^* n)$  solvability. The current running time is exponential, and an open question is whether we can find such a certificate in polynomial time, or if we can prove that e.g. deciding the existence of a certificate is an NP-hard problem.

The second question regards the complexity class of  $n^{\Theta(1)}$ . While we present a practical algorithm that determines if the complexity is  $n^{\Theta(1)}$ , our algorithm does not determine the precise complexity. Whether there is an efficient algorithm for finding the precise value of  $k$  such that the complexity is  $\Theta(n^{1/k})$  remains open.

Another natural question regards extending our results to unrooted trees. While decidability is known in the  $\Omega(\log n)$  region, it is known to require exponential time [14]. In our setting, we can decide if a problem is  $O(\log n)$  or  $n^{\Omega(1)}$  in polynomial time via a characterization based on the existence of a minimal absorbing subgraph. Coincidentally, Brandt et al. [11, Section 6] also recently proved that essentially the same characterization characterizes whether a problem is  $O(\log n)$  or  $n^{\Omega(1)}$  for regular unrooted trees. With some minor modification, our polynomial-time algorithm that decides if a problem is  $O(\log n)$  or  $n^{\Omega(1)}$  can also be adapted to the setting of regular unrooted trees. However, deciding if a problem on regular trees requires  $O(1)$ ,  $\Theta(\log^* n)$ , or  $\Omega(\log n)$  rounds remains a major open question.

**Acknowledgements** We would like to thank Juho Hirvonen, Henrik Lievonen, Yannic Maus, and Mikael Rabie for discussions and comments, and anonymous reviewers for their helpful feedback on previous versions of this work. We also wish to acknowledge CSC – IT Center for Science, Finland, for computational resources.

**Funding** Open Access funding provided by Aalto University.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Balliu, A., Brandt, S., Chang, Y.-J., Olivetti, D., Rabie, M., Suomela, J.: The distributed complexity of locally checkable problems on paths is decidable. In: Proc. 38th ACM Symposium on Principles of Distributed Computing (PODC 2019). ACM Press, pp. 262–271. <https://doi.org/10.1145/3293611.3331606> arxiv: 1811.01672 (2019)
- Balliu, A., Brandt, S., Efron, Y., Hirvonen, J., Maus, Y., Olivetti, D., Suomela, J.: Classification of distributed binary labeling problems. In: Proc. 34th International Symposium on Distributed Computing (DISC 2020) (LIPIcs, Vol. 179). Schloss Dagstuhl–Leibniz-Zentrum für Informatik, pp. 17:1–17:17. <https://doi.org/10.4230/LIPIcs.DISC.2020.17> arxiv: 1911.13294
- Balliu, A., Brandt, S., Hirvonen, J., Olivetti, D., Rabie, M., Suomela, J.: Lower bounds for maximal matchings and maximal independent sets. In: Proc. 60th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2019). IEEE, pp. 481–497. <https://doi.org/10.1145/3461458> arxiv: 1901.02441 (2019)
- Balliu, A., Brandt, S., Olivetti, D., Suomela, J.: Almost global problems in the LOCAL model. Distributed Computing. <https://doi.org/10.1007/s00446-020-00375-2> arxiv: 1805.04776 (2020)
- Balliu, A., Brandt, S., Olivetti, D., Suomela, J.: How much does randomness help with locally checkable problems?. In: Proc. 39th ACM Symposium on Principles of Distributed Computing (PODC 2020). ACM Press, pp. 299–308. <https://doi.org/10.1145/3382734.3405715> arxiv: 1902.06803 (2020)
- Balliu, A., Censor-Hillel, K., Yannic Maus, Dennis Olivetti, and Jukka Suomela. 2021. Locally Checkable Labelings with Small Messages. <https://doi.org/10.4230/LIPIcs.DISC.2021.8> arxiv: 2105.05574
- Balliu, A., Hirvonen, J., Korhonen, J. H., Lempäinen, T., Olivetti, D., Suomela, J.: New classes of distributed time complexity. In: Proc. 50th ACM Symposium on Theory of Computing (STOC 2018). ACM Press, pp. 1307–1318. <https://doi.org/10.1145/3188745.3188860> arxiv: 1711.01871 (2018)
- Balliu, A., Hirvonen, J., Olivetti, D., Suomela, J.: Hardness of minimal symmetry breaking in distributed computing. In: Proc. 38th ACM Symposium on Principles of Distributed Computing (PODC 2019). ACM Press, pp. 369–378. <https://doi.org/10.1145/3293611.3331605> arxiv: 1811.01643 (2019)
- Barenboim, L., Elkin, M.: Distributed Graph Coloring: Fundamentals and Recent Developments. Morgan & Claypool. <https://doi.org/10.2200/S00520ED1V01Y201307DCT011> (2013)
- Brandt, S.: An Automatic Speedup Theorem for Distributed Problems. In: Proc. 38th ACM Symposium on Principles of Distributed Computing (PODC 2019). ACM, pp. 379–388. <https://doi.org/10.2200/S00520ED1V01Y201307DCT011> (2019)
- Brandt, S., Chang, Y.-J., Grebík, J., Grunau, C., Rozhoň, V., Vidnyánszky, Z.: Local Problems on Trees from the Perspectives of Distributed Algorithms, Finitary Factors, and Descriptive Combinatorics. <https://doi.org/10.4230/LIPIcs.ITCS.2022.29> arxiv: 2106.02066 (2021)
- Brandt, S., Fischer, O., Hirvonen, J., Keller, B., Lempäinen, T., Rybicki, J., Suomela, J., Uitto, J.: A lower bound for the distributed Lovász local lemma. In: Proc. 48th ACM Symposium on Theory of Computing (STOC 2016). ACM Press, pp. 479–488. <https://doi.org/10.1145/2897518.2897570> arxiv: 1511.00900 (2016)
- Brandt, S., Hirvonen, J., Korhonen, J. H., Lempäinen, T., Östergård, P.R.J., Purcell, C., Rybicki, J., Suomela, J., Uznański, P.: LCL problems on grids. In: Proc. 36th ACM Symposium on Principles of Distributed Computing (PODC 2017). ACM Press, pp. 101–110. <https://doi.org/10.1145/3087801.3087833> arxiv: 1702.05456 (2017)
- Chang, Y.-J.: The Complexity Landscape of Distributed Locally Checkable Problems on Trees. In: Proc. 34th International Symposium on Distributed Computing (DISC 2020) (LIPIcs, Vol. 179). Schloss Dagstuhl–Leibniz-Zentrum für Informatik, pp. 18:1–18:17 (2020) <https://doi.org/10.4230/LIPIcs.DISC.2020.18>
- Chang, Y.-J., Kopelowitz, T., Pettie, S.: An Exponential Separation between Randomized and Deterministic Complexity in the LOCAL Model. SIAM J. Comput. **48**(1), 122–143 (2019). <https://doi.org/10.1137/17M1117537>
- Chang, Y.-J., Pettie, S.: A Time Hierarchy Theorem for the LOCAL Model. SIAM J. Comput. **48**(1), 33–69 (2019). <https://doi.org/10.1137/17M1157957>
- Chang, Y.-J., Studený, J., Suomela, J.: Distributed graph problems through an automata-theoretic lens. In: Proc. 28th International Colloquium on Structural Information and Communication Complexity (SIROCCO 2021) (LNCS). Springer. [https://doi.org/10.1007/978-3-030-79527-6\\_3](https://doi.org/10.1007/978-3-030-79527-6_3) arxiv: 2002.07659 (2021)
- Chung, K.-M., Pettie, S., Su, H.-H.: Distributed algorithms for the Lovász local lemma and graph coloring. Distributed Comput. **30**(4), 261–280 (2017). <https://doi.org/10.1007/s00446-016-0287-6>
- Cole, R., Vishkin, U.: Deterministic Coin Tossing with Applications to Optimal Parallel List Ranking. Inf. Control. **70**(1), 32–53 (1986). [https://doi.org/10.1016/S0019-9958\(86\)80023-7](https://doi.org/10.1016/S0019-9958(86)80023-7)
- Fischer, M., Ghaffari, M.: Sublogarithmic Distributed Algorithms for Lovász Local Lemma, and the Complexity Hierarchy. In: Proc. 31st International Symposium on Distributed Computing (DISC 2017) (LIPIcs, Vol. 91). Schloss Dagstuhl–Leibniz-Zentrum für Informatik, pp. 18:1–18:16. (2017) <https://doi.org/10.4230/LIPIcs.DISC.2017.18>
- Linial, N.: Locality in Distributed Graph Algorithms. SIAM J. Comput. **21**(1), 193–201 (1992). <https://doi.org/10.1137/0221015>
- Miller, G.L., Reif, J.H.: Parallel tree contraction and its application. In: Proc. 26th Annual Symposium on Foundations of Computer Science (FOCS 1985). IEEE, pp. 478–489. <https://doi.org/10.1109/SFCS.1985.43> (1985)
- Naor, M.: A Lower Bound on Probabilistic Algorithms for Distributive Ring Coloring. SIAM J. Discret. Math. **4**(3), 409–412 (1991). <https://doi.org/10.1137/0404036>

24. Naor, M., Stockmeyer, L.J.: What Can be Computed Locally? *SIAM J. Comput.* **24**(6), 1259–1277 (1995). <https://doi.org/10.1137/S0097539793254571>
25. Olivetti, D.: Round Eliminator: a tool for automatic speedup simulation. <https://github.com/olidennis/round-eliminator> (2020)
26. Rozhoň, V., Ghaffari, M.: Polylogarithmic-time deterministic network decomposition and distributed derandomization. In: Proc. 52nd Annual ACM SIGACT Symposium on Theory of Computing (STOC 2020). ACM, pp. 350–363. <https://doi.org/10.1145/3357713.3384298> (2020)
27. Studený J., Tereshchenko, A.: Rooted Tree Classifier. <https://github.com/jendas1/rooted-tree-classifier> (2021)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.