Dynamic and Multi-Functional Labeling Schemes

Søren Dahlgaard, Mathias Bæk Tejs Knudsen, and Noy Rotbart

Department of Computer Science, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen, Denmark
{soerend,knudsen,noyro}@di.ku.dk

Abstract. We investigate labeling schemes supporting adjacency, ancestry, sibling, and connectivity queries in forests. In the course of more than 20 years, the existence of $\log n + O(\log \log n)$ labeling schemes supporting each of these functions was proven, with the most recent being ancestry [Fraigniaud and Korman, STOC ’10]. Several multi-functional labeling schemes also enjoy lower or upper bounds of $\log n + \Omega(\log \log n)$ or $\log n + O(\log \log n)$ respectively. Notably an upper bound of $\log n + 2 \log \log n$ for adjacency+siblings and a lower bound of $\log n + \log \log n$ for each of the functions siblings, ancestry, and connectivity [Alstrup et al., SODA ’03]. We improve the constants hidden in the $O$-notation, where our main technical contribution is a $\log n + 2 \log \log n$ lower bound for connectivity +ancestry and connectivity+siblings.

In the context of dynamic labeling schemes it is known that ancestry requires $\Omega(n)$ bits [Cohen, et al. PODS ’02]. In contrast, we show upper and lower bounds on the label size for adjacency, siblings, and connectivity of $2 \log n$ bits, and $3 \log n$ to support all three functions. We also show that there exist no efficient dynamic adjacency labeling schemes for planar, bounded treewidth, bounded arboricity and bounded degree graphs.

1 Introduction

A labeling scheme is a method of distributing the information about the structure of a graph among its vertices by assigning short labels, such that a selected function on pairs of vertices can be computed using only their labels. The concept was introduced by Kannan, Naor and Rudich [1], and explored by a wealth of subsequent work [2–7].

Labeling schemes for trees have been studied extensively in the literature due to their practical applications in improving the performance of XML search engines. Indeed, XML documents can be viewed as labeled forests, and typical

Research partly supported by Mikkel Thorup’s Advanced Grant from the Danish Council for Independent Research under the Sapere Aude research carrier programme.

Research partly supported by the FNU project AlgoDisc - Discrete Mathematics, Algorithms, and Data Structures.

© Springer International Publishing Switzerland 2014
DOI: 10.1007/978-3-319-13075-0_12
queries over the documents amount to testing classic properties such as adjacency, ancestry, siblings and connectivity between such labeled tree nodes [8]. In their seminal paper, Kannan et. al. [1] introduced labeling schemes using at most $2 \log n$ bits\(^1\) for each of the functions adjacency, siblings and ancestry. Improving these results have been motivated heavily by the fact that a small improvement of the label size may contribute significantly to the performance of XML search engines. Alstrup, Bille and Rauhe [3] established a lower bound of $\log n + \log \log n$ for the functions siblings, connectivity and ancestry along with a matching upper bound for the first two. For adjacency, a $\log n + O(\log^{\ast} n)$ labeling scheme was presented in [2]. A $\log n + O(\log \log n)$ labeling scheme for ancestry was established only recently by Fraigniaud and Korman [4].

In most settings, it is the case that the structure of the graph to be labeled is not known in advance. In contrast to the static setting described above, a dynamic labeling scheme receives the tree as an online sequence of topological events. Cohen, Kaplan and Milo [9] considered dynamic labeling schemes where the encoder receives $n$ leaf insertions and assigns unique labels that must remain unchanged throughout the labeling process. In this context, they showed a tight bound of $\Theta(n)$ bits for any dynamic ancestry labeling scheme. We stress the importance of their lower bound by showing that it extends to routing, NCA, and distance as well. In light of this lower bound, Korman, Peleg and Rodeh [10] introduced dynamic labeling schemes where node re-label is permitted and performed by message passing. In this model they obtain a compact labeling scheme for ancestry, while keeping the number of messages small. Additional results in this setting include conversion methods for static labeling schemes [10,11], as well as specialized distance [11] and routing [12] labeling schemes. See [13] for experimental evaluation.

Considering the static setting, a natural question is to determine the label size required to support some, or all, of the functions. Simply concatenating the labels mentioned yield a $O(\log n)$ label size, which is clearly undesired. Labeling schemes supporting multiple functions\(^2\) were previously studied for adjacency and sibling queries. Alstrup et al. [3] proved a $\log n + 5 \log \log n$ label size which was improved by Gavoille and Labourel [14] to $\log n + 2 \log \log n$. See Table 1 for a summary of labeling schemes for forests including the results of this paper.

### 1.1 Our Contribution

We contribute several upper and lower bounds for both dynamic and multi-functional labeling schemes. First, we observe that the naïve $2 \log n$ adjacency, siblings and connectivity labeling schemes are suitable for the dynamic setting without the need of relabeling. We then present simple families of insertion sequences for which labels of size $2 \log n$ are required, showing that in the dynamic setting the naïve labeling schemes are in fact optimal. The result is in contrast to the static case, where adjacency labels requires strictly fewer bits

---

\(^1\) Throughout this paper we let $\log n = \lceil \log_2 n \rceil$ unless stated otherwise.

\(^2\) We refer to such labeling schemes as multi-functional labeling schemes.