Information Processing Letters 111 (2011) 671–673

Contents lists available at ScienceDirect

Information Processing Letters

www.elsevier.com/locate/ipl

A note on exact distance labeling

Oren Weimann *, David Peleg

The Weizmann Institute of Science, Israel

ARTICLE INFO

Article history:
Received 19 December 2010
Received in revised form 16 April 2011
Accepted 19 April 2011
Available online 22 April 2011
Communicated by A.A. Bertossi

Keywords:
Labeling schemes
Shortest paths
Graph algorithms
Distributed systems

ABSTRACT

We show that the vertices of an edge-weighted undirected graph can be labeled with labels of size $O(n)$ such that the exact distance between any two vertices can be inferred from their labels alone in $O(\log^* n)$ time. This improves the previous best exact distance labeling scheme that also requires $O(n)$-sized labels but $O(\log\log n)$ time to compute the distance. Our scheme is almost optimal as exact distance labeling is known to require labels of length $\Omega(n)$.

1. Introduction

A *distance labeling scheme* of a graph G is a way of assigning unique labels to the vertices of G so that the distance between any two vertices can be inferred from their labels alone. The scheme is composed of a marker algorithm for labeling the vertices with (hopefully short) labels, coupled with a (hopefully fast) decoder algorithm for extracting a distance from two labels. In this paper we focus on exact distance labeling for edge-weighted undirected general graphs.

One can clearly label every vertex with its vector of distances to all other vertices in G. For n-node graphs, this leads to $O(n\log n)$ bit labels with $O(1)$ time to decode the distance. Graham and Pollak [4] showed how to reduce the label length to $\Theta(n)$ at the prohibitive cost of $\Theta(n)$ query time to decode the distance. This follows from the *Squashed Cube Conjecture* that was proved by Winkler [7] and states that we can label each vertex by $n-1$ symbols such that the distance between two vertices corresponds to the Hamming distance of the two labels. Gavoille et al. [3] presented $\Theta(n)$-sized labels requiring only $O(\log\log n)$ time for decoding the distance. They also showed that general undirected graphs require labels of length $\Omega(n)$ (regardless of the distance decoding time). This lower bound holds even if we relax the requirement for *exact* distances and settle for an approximate distance with stretch bounded by 3. In this paper, we show how to improve the labeling scheme of Gavoille et al. to $\Theta(n)$-sized labels requiring $O(\log^* n)$ time for decoding the distance.

Our results are mainly of theoretical interest as $\Theta(n)$-sized labels are typically too long. In practice, there are two ways of overcoming the $\Theta(n)$ lower bound. The first is to design specific schemes for restricted graph families: Gavoille et al. presented $\Theta(\log^2 n)$-sized labels for trees, $O(\sqrt{n}\log n)$-sized labels for planar graphs, and $O(r(n)\log^2 n)$-sized labels for any graph with separator $r(n)$. In terms of lower bounds, they showed a lower bound of $\Omega(n^{1/3})$ for planar graphs and $\Omega(\sqrt{n})$ for bounded degree graphs. The other option of overcoming the lower bounds is to settle for approximate distances. There have been many results exhibiting short approximate distance labels, see for example [1,2,5,6].

2. Preliminaries

Our labeling scheme is a modification of the Gavoille et al. scheme. We now give a slightly simplified presentation

* Corresponding author.
E-mail addresses: oren.weimann@weizmann.ac.il (O. Weimann), david.peleg@weizmann.ac.il (D. Peleg).

0020-0190/$ – see front matter
doi:10.1016/j.ipl.2011.04.006

of their scheme which we can later modify. In particular, we first discuss their use of *ρ-dominating sets*.

A ρ-dominating set for a graph G is a set $S \subseteq V(G)$ such that for every vertex $u \in V(G)$ there is a vertex $v \in S$ at distance at most ρ from it. The vertex v is called the *dominator* of u and denote $v = dom_S(u)$. It is easy to show that for every n-vertex connected graph G and integer $\rho \geqslant 0$, there exists a ρ-dominating set of cardinality at most n/ρ.

Given an undirected connected graph G with positive edge-lengths, we denote by $d(x, y)$ the length of the shortest path between vertices x and y in G.

Lemma 1. *(See [3].) For every two nodes $x, y \in V(G)$:*

$$d\big(dom_S(x), dom_S(y)\big) - 2\rho \leqslant d(x, y) \leqslant d\big(dom_S(x), dom_S(y)\big) + 2\rho.$$

The above lemma implies that knowing ρ, $d(x, y) \bmod (4\rho + 1)$, and $d(dom_S(x), dom_S(y))$, one can compute $d(x, y)$. Indeed, the lemma shows there are $4\rho + 1$ consecutive possible values for $d(x, y)$, exactly one of which can be congruent to $d(x, y)$ modulo $4\rho + 1$. This fact is crucial in both the Gavoille et al. scheme as well as in ours. Gavoille et al. considered a collection of $k = \lceil \log\log n \rceil$ dominating sets $S_0, \ldots, S_k$ so that S_i is a 2^i-dominating set of G. In this way, $|S_i| \leqslant n/2^i$, and $S_0 = V(G)$.

Consider the vertices of S_k. There are only $O(n/\log n)$ such vertices, so each one can afford to store in its label the vector of distances to all other vertices in S_k. This would imply a label of length $\log n \cdot O(n/\log n) = O(n)$ for every vertex in S_k, and the distance between two vertices in S_k can be decoded in $O(1)$ time. The Gavoille et al. scheme proceeds inductively. The label of every vertex u in S_i is composed of two fields: The first field is a copy of the label of $dom_{S_{i+1}}(u)$, and the second field is the list $\{d(u, v) \bmod (4 \cdot 2^i + 1)\}_{v \in S_i}$. This way, when we want to compute the distance between two vertices u, u' that are both in S_i, we first recursively compute $d(dom_{S_{i+1}}(u), dom_{S_{i+1}}(u'))$ from the first field of u's label. We then compute $d(u, u') \bmod (4 \cdot 2^i + 1)$ from the second field of u's label in $O(1)$ time. From the above discussion, these two values are enough to decode $d(u, u')$. The correctness of the scheme follows from the fact that every two vertices belong to $S_0 = V(G)$. The time to decode a distance is bounded by $O(k) = O(\log\log n)$. The size of a label is bounded by

$$O(n) + \sum_{i=1}^{k} |S_i| \log\big(4 \cdot 2^i + 1\big) \leqslant O(n) + n \sum_{i=1}^{k} \frac{2 + i}{2^i} = O(n).$$

In the next section we show a simple modification of the label so its size remains $O(n)$ but decoding a distance reduces from $O(\log\log n)$ to $O(\log^* n)$.

3. Our labeling scheme

We take a closer look at the Gavoille et al. scheme. Consider a vertex $u \in S_i$ and let the jth *dominator* of u, denoted by $dom^j(u)$, be defined as follows: The first dominator of u is $dom^1(u) = dom_{S_{i+1}}(u)$, the second dominator is $dom^2(u) = dom_{S_{i+2}}(dom_{S_{i+1}}(u))$, the third is $dom^3(u) = dom_{S_{i+3}}(dom_{S_{i+2}}(dom_{S_{i+1}}(u)))$, and so on. The following is a simple extension of Lemma 1.

Lemma 2. *For every two nodes $x, y \in S_i$ and every $j \geqslant i$, one can compute $d(x, y)$ from $d(x, y) \bmod (4(2^i + 2^{i+1} + \cdots + 2^j) + 1)$ and $d(dom^j(x), dom^j(y))$.*

Proof. As we mentioned before, Lemma 1 shows that one can compute $d(x, y)$ out of $d(x, y) \bmod (4 \cdot 2^i + 1)$ and $d(dom_{S_{i+1}}(x), dom_{S_{i+1}}(y))$. This implies correctness for the case $j = i$. We prove correctness for the case $j = i + 1$, a similar argument works for larger values of j. By Lemma 1 we know that

$$d\big(dom_{S_{i+1}}(x), dom_{S_{i+1}}(y)\big) - 2 \cdot 2^i \leqslant d(x, y) \leqslant d\big(dom_{S_{i+1}}(x), dom_{S_{i+1}}(y)\big) + 2 \cdot 2^i.$$

Applying Lemma 1 on $dom_{S_{i+1}}(x)$ and $dom_{S_{i+1}}(y)$ (rather than x and y) we get that

$$d\big(dom^2(x), dom^2(y)\big) - 2 \cdot 2^{i+1} \leqslant d\big(dom_{S_{i+1}}(x), dom_{S_{i+1}}(x)\big) \leqslant d\big(dom^2(x), dom^2(y)\big) + 2 \cdot 2^{i+1}.$$

Combining these equations we get that

$$d\big(dom^2(x), dom^2(y)\big) - 2\big(2^{i+1} + 2^i\big) \leqslant d(x, y) \leqslant d\big(dom^2(x), dom^2(y)\big) + 2\big(2^{i+1} + 2^i\big).$$

This means that there are $4(2^{i+1} + 2^i) + 1$ consecutive possible values for $d(x, y)$, exactly one of which can be congruent to $d(x, y)$ modulo $4(2^{i+1} + 2^i) + 1$. $\square$

We modify the labeling scheme according to Lemma 2 in the following way.

The label of every vertex u in S_i is composed of two fields: The first field is a copy of the label of $dom^j(u)$ where $j = 2^{i/2}$, and the second field is the list $\{d(u, v) \bmod (4(2^i + 2^{i+1} + \cdots + 2^j) + 1)\}_{v \in S_i}$. This way, when we want to compute the distance between two vertices u, u' that are both in S_i, we first recursively compute $d(dom^j(u), dom^j(u'))$ from the first field of u's label. We then compute $d(u, u') \bmod (4(2^i + 2^{i+1} + \cdots + 2^j) + 1)$ from the second field of u's label in $O(1)$ time. By Lemma 2, these two values are enough to decode $d(u, u')$.

We are thus left with showing that the size of a label is $O(n)$ and that a distance can be decoded in $O(\log^* n)$ time. Letting $k = \lceil \log\log n \rceil$, the size of a label is bounded by

$$O(n) + \sum_{i=1}^{k} |S_i| \log\big(4\big(2^i + 2^{i+1} + \cdots + 2^{2^{i/2}}\big) + 1\big) \leqslant O(n) + n \sum_{i=1}^{k} \frac{4}{2^{i/2}} = O(n).$$

To establish the distance decoding time, define the function $f(i) = 2^{i/2}$. The time to decode a distance is proportional to the number of times we need to apply the function f starting with 0 in order to get to $\log\log n$. This is bounded by $O(\log^* n)$. In fact, it is even a bit better and bounded by $O(\log^*(\log\log n))$.

4. Conclusions

In this paper, we have improved the query-time in the labeling scheme of [3] for general graphs from $O(\log\log n)$ to $O(\log^*(\log\log n))$. The size of our labels is $\Theta(n)$ which is known to be optimal for exact distances [3]. There are no known lower bounds on the query-time and the gap between our query-time and $O(1)$ thus remains as an open problem.

References

[1] I. Abraham, C. Gavoille, Object location using path separators, in: Proceedings of the 25th ACM Symposium on Principles Of Distributed Computing (PODC), 2006, pp. 188–197.
[2] C. Gavoille, M. Katz, N.A. Katz, C. Paul, D. Peleg, Approximate distance labeling schemes, in: Proceedings of the 9th Annual European Symposium on Algorithms (ESA), 2001, pp. 476–487.
[3] C. Gavoille, D. Peleg, S. Perennes, R. Raz, Distance labeling in graphs, Journal of Algorithms 53 (2004) 85–112.
[4] R.L. Graham, H.O. Pollak, On embedding graphs in squashed cubes, Lecture Notes in Math. 303 (1972) 99–110.
[5] A. Gupta, A. Kumar, R. Rastogi, Traveling with a pez dispenser (or, routing issues in MPLS), SIAM Journal on Computing 32 (2) (2005) 453–474.
[6] D. Peleg, Proximity-preserving labeling schemes, Journal of Graph Theory 33 (2000) 167–176.
[7] P.M. Winkler, Proof of the squashed cube conjecture, Combinatorica 3 (1) (1983) 135–139.