On approximation intractability of the path–distance–width problem

Koichi Yamazaki

Department of Computer Science Gunma University, 1-5-1 Tenjin-cho, Kiryu zip:376-8515,
Gunma, Japan

Received 21 April 1999; revised 25 February 2000; accepted 10 April 2000

Abstract

Path–distance–width of a graph $G = (V, E)$, denoted by $pdw(G)$, is the minimum integer $k$ satisfying that there is a nonempty subset of $S \subseteq V$ such that the number of the nodes with distance $i$ from $S$ is at most $k$ for any nonnegative integer $i$. It is known that given a positive integer $k$ and a graph $G$, the decision problem $pdw(G) \leq k$ is NP-complete even if $G$ is a tree (Yamazaki et al. Lecture Notes in Computer Science, vol. 1203, Springer, Berlin, 1997, pp. 276–287). In this paper, we show that it is NP-hard to approximate the path–distance–width of a graph within a ratio $\frac{4}{3} - \varepsilon$ for any $\varepsilon > 0$, even for trees. © 2001 Elsevier Science B.V. All rights reserved.

1. Introduction

Given a graph $G$ and a set of nodes $X_0 \subseteq V(G)$, we can determine easily and uniquely the decomposition $(X_0, \ldots, X_d)$ in which $X_i$ is the set of nodes of distance $i$ from $X_0$. Thus, we can also determine the path–distance–width for the root set $X_0$ of $G$ $pdw_{X_0}(G) = \max_{0 \leq i \leq d}|X_i|$. In this paper we will consider a graph problem in which we wish to minimize the path–distance–width by choosing a suitable root set. The minimum path–distance–width is called the path–distance–width of $G$, and is denoted by $pdw(G)$. This problem may not be as easy as it may initially appear; indeed, it is known that even for the trees $T$ the decision problem $pdw(T) \leq k$ is NP-complete [11]. It is not known whether or not the problem is fixed parameter tractable, i.e., whether there exists an algorithm which solves the problem with running time $O(f(k)n^c)$, where $c$ is a constant and independent of $k$, and $f$ is any function.

1 Supported by the Scientific Grant-in-Aid for Encouragement of Young Scientists from Ministry of Education, Science, Sports and Culture of Japan.

E-mail address: koichi@comp.cs.gunma-u.ac.jp (K. Yamazaki).
This problem is related to the following two problems which are NP-complete even for the trees and fixed parameter intractable (see [4,5]):

**BANDWIDTH**

**Instance:** A graph \( G \) and a positive integer \( k \).

**Question:** Is there a bijection \( f: V(G) \rightarrow \{1, \ldots, |V(G)|\} \) such that \( \{u,v\} \in E(G) \) implies \(|f(u) - f(v)| \leq k\)?

**UNIFORM EMULATION ON A PATH**

**Instance:** A graph \( G \) and a positive integer \( k \).

**Question:** Is there a decomposition \((V_0, V_1, \ldots, V_\ell)\) of \( G \) such that

1. \( \bigcup_{i=0}^{\ell} V_i = V(G) \), \( V_i \cap V_j = \emptyset \) for \( 0 \leq i < j \leq \ell \),
2. \( |V_i| \leq k \) for all \( 0 \leq i \leq \ell \), and
3. \( \{u,v\} \in E(G) \) implies \( u, v \in V_i \cup V_{i+1} \) for some \( 0 \leq i \leq \ell - 1 \).

We denote the minimum integer \( k \) which satisfies the condition in BANDWIDTH by \( bw(G) \), similarly by \( uep(G) \) for UNIFORM EMULATION ON A PATH. From the point of view of approximability, there is a close relationship between BANDWIDTH and UNIFORM EMULATION ON A PATH. It is easy to see that \( uep(G) \leq bw(G) \leq 2uep(G) - 1 \), so BANDWIDTH can be approximated within a constant factor if and only if UNIFORM EMULATION ON A PATH can be approximated within a constant factor. The path–distance–width problem can be thought as UNIFORM EMULATION ON A PATH with the restriction that (4) for every \( 1 \leq i \leq \ell \) and for every \( v \in V_i \), there exists a node \( u \in V_{i-1} \) which is a neighbor of \( v \).

Recently, Blache et al. have shown, using gadget reduction from 3SAT, that BANDWIDTH does not have a polynomial time approximation algorithm with approximation ratio \( \frac{4}{3} - \varepsilon \) for any \( \varepsilon > 0 \) unless \( P = NP \), even for the trees [2]. It seems to be natural to ask whether or not the path–distance–width problem has a PTAS. In this paper we will show that for the path–distance–width problem there is no polynomial time approximation algorithm with approximation ratio \( \frac{4}{3} - \varepsilon \) for any \( \varepsilon > 0 \) unless \( P = NP \), even for the trees by using gadget reduction from EXACT COVER BY 3-SETS.

This paper is organized as follows: In Section 2 we give the definitions and notation which are used in this paper. The main results of the paper are described in Section 3.

### 2. Definitions

The graphs we consider are simple, undirected and connected, and contain no self-loops and no multiple edges. For a graph \( G \), we denote its set of node by \( V(G) \) and its set of edges by \( E(G) \). For a graph \( G \) and two nodes \( u \) and \( v \in V(G) \), \( dist_G(u,v) \) denotes the distance between \( u \) and \( v \) which is the number of edges on a shortest path between \( u \) and \( v \). For a set \( S \subseteq V(G) \) and a node \( w \in V(G) \), \( dist_G(S,w) \) denotes \( \min_{v \in S} dist_G(v,w) \).

We recall the definition of the path–distance–width.
Definition. Let $G$ be a graph. A sequence of subsets of $V(G) (X_0, X_1, \ldots, X_d)$ is called the path–distance-decomposition of $G$ with root set $X_0$ if $X_i = \{ v : \text{dist}_G(X_0, v) = i \} \neq \emptyset$ for all $0 \leq i \leq d$ and $V(G) = \bigcup_{0 \leq i \leq d} X_i$. The path–distance–width of $G$ with root set $X_0$, denoted by $\text{pdw}_{X_0}(G)$, is $\max_{0 \leq i \leq d} |X_i|$. The path–distance–width of $G$, denoted by $\text{pdw}(G)$, is $\min_{S \subseteq V(G)} \text{pdw}_S(G)$.

In the proof of the main theorem, we will show a reduction from the following problem.

EXACT COVER BY 3-SETS (X3C)

Instance: A set $X$ with $|X| = 3i$ and a collection $\mathcal{C}$ of three-element subsets of $X$.

Question: Does $\mathcal{C}$ contain an exact cover for $X$, i.e., a subcollection $\mathcal{C}' \subseteq \mathcal{C}$ so that each element of $X$ occurs in exactly one member of $\mathcal{C}'$?

3. Result

The idea of our gadget reduction is as follows. In our gadget reduction, we construct a tree $T_{\mathcal{C}}$ with the following gap property from given an instance $\mathcal{C} = \{C_1, C_2, \ldots, C_n\}$ with $X = \{e_1, e_2, \ldots, e_m\}$ of X3C.

Gap property. There are some constants $\alpha$ and $\beta$ ($\alpha < \beta$) such that if $\mathcal{C}$ is a yes instance, then $\text{pdw}(T_{\mathcal{C}}) \leq \alpha$, otherwise $\text{pdw}(T_{\mathcal{C}}) \geq \beta$.

From this gap (i.e., between $\alpha$ and $\beta$), we can derive the inapproximability (for details see [1]).

Let us consider a modified problem instead of original problem in order to catch the idea. In the modified problem, we will consider a function $W$ instead of $\text{pdw}$ as objective function. Then we will construct a weighted tree $T_{\mathcal{C}}$ with the property that the objective value is at most $2s$ for a yes instance, and at least $3s$ for a no instance, where $s$ is a constant.

The construction of $T_{\mathcal{C}}$ is as follows. Without loss of generality, we can assume that in $\mathcal{C}$ no element occurs in more than three subsets (see [8]), and there exists an element in $X$ which occurs in exactly one subset in $\mathcal{C}$. Each $C_i \in \mathcal{C}$ corresponds to a path $(v^1_i, v^2_i, \ldots, v^m_i)$ of length $m$. There are two extra paths: left path $(v^1_i, v^2_i, \ldots, v^m_i)$ and right path $(v^1_i, v^2_i, \ldots, v^m_i)$. The root has children as $v^1_i$, $v^1_i$, and $v^1_i$ for all $1 \leq i \leq n$. All nodes in $T_{\mathcal{C}}$ are associated with weight 0, $s$ or $2s$. The root has weight $2s$, $v^j_i$ for $1 \leq j \leq m$ has weight $s$, node $v^j_i$ has weight $(3-i)s$ $(i \leq 3)$ if $e_j$ occurs exactly $i$ subsets, node $v^j_i$ has weight $s$ if $e_j$ occurs in $C_i$, and the other nodes have weight 0 (see Fig. 1).

Now we demonstrate that $T_{\mathcal{C}}$ has the gap property. Let us consider a path–distance-decomposition $D = (X_0, \ldots, X_d)$ with a root set $X_0$, and define $W(D)$ as the maximum
of the total weight of $X_i$ over all $0 \leq i \leq d$. It is easy to see that if the instance has an exact cover $\mathcal{C}'$ and $X_0 = \{v_{im} : C_i \in \mathcal{C}'\} \cup \{v_{m}^r\}$, then $W(D) = 2s$ (see Fig. 2). On the other hand, if there exists a path–distance-decomposition $D$ such that $W(D) = 2s$, then the instance has an exact cover. The reason is as follows. Since the total weight of nodes in $T_{\ell}$ is equal to $4sm + 2s$, if $W(D) = 2s$ then $2m \leq d$. Furthermore, as the diameter of $T_{\ell}$ is $2m$, $d = 2m$. Hence, the total weight of nodes in $X_i$ is equals to $2s$ for each $0 \leq i \leq d$. From $2m \leq d$, any internal node of $T_{\ell}$ cannot be in $X_0$. Thus $X_0 \subseteq \{v_{im}^l : 1 \leq i \leq n\} \cup \{v_{m}^l, v_{m}^r\}$. From the second assumption, there exists an index $j$ ($1 \leq j \leq m$) such that $e_j$ occurs in exactly one subset in $\mathcal{C}$ (in Fig. 1, 8 is such an index). If $v_{m}^l$ and $v_{m}^r$ are both in $X_0$, then the total weight of nodes in $X_{m-j}$ is at least $3s$. Hence, without loss of generality, we can assume that $X_0 \subseteq \{v_{im}^l : 1 \leq i \leq n\} \cup \{v_{m}^r\}$. Now it is easy to see that $\{C_i : v_{m}^r \in X_0\}$ is an exact cover.

Since $W(D) \geq 2s$ for any path–distance-decomposition $D$ and $W(D)$ is a multiple of $s$, $W(D) > 2s$ implies $W(D) \geq 3s$. Therefore, if the instance does not have any exact cover $\mathcal{C}'$, then $W(D) \geq 3s$. 

---

**Fig. 1.** Tree $T_{\ell}$.

**Fig. 2.** A optimal layout for Tree $T_{\ell}$.
Theorem 3.1. For a path–distance–width problem, there is no polynomial time-approximation algorithm with approximation ratio $\frac{2}{3} - \varepsilon$ for any $\varepsilon > 0$, unless \textbf{P} = \textbf{NP}.

Proof. We will show a gadget reduction from exact cover by three-set (cf. [2]). Let $X3C = (X, \mathcal{C})$ be an instance of exact cover by three-set in which no element occurs in more than three subsets. The restriction does not violate its \textbf{NP}-completeness (see [8]), and plays an important role in the proof. Put $p = m/3$ and $q = n - p$, where $m = |X|$ and $n = |\mathcal{C}|$ (note that $q \geq 0$), and let $s$ be some integer which is fixed later.

Construct tree $T$ from $X3C$ in the following way. Each subset $C_i$ $(1 \leq i \leq n)$ will be associated with a path of length $m$ in which a star of size $s$ is attached to the $j$th node iff $e_j$ occurs in $C_i$. $T$ has two other paths, $L$ and $R$, of length $m$, thus $T$ has $m + 2$ paths of length $m$. A star of size $s + q$ is attached to each node on the path $L$. Furthermore, two additional stars of size $p$ and $2s + q$ are attached to the first node and the last node on $L$, respectively. If $e_j$ occurs exactly in $i$ $(i \leq 3)$ subsets, then a star of size $(3 - i)s + p$ is attached to the $j$th node on path $R$, while moreover an additional star of size $q + 1$ is attached to the last node on $R$. $T$ has a center node called root and which has also an attached star of size $2s + p$. The first node of each path and root are linked by an edge (see Fig. 3).

In the following, we will refer to the node $v$ on path $L(R, C_j)$ so that $\text{dist}_T(\text{root}, v) = m$ as $l(r, c_j)$, respectively.

First, we demonstrate that if $X3C$ has an exact cover $\mathcal{C}'$, then $\text{pdw}(T) \leq 2s + n + 1$. Set $X_0 = \{c_i: C_i \in \mathcal{C}'\} \cup \{l\} \cup \{v_1, v_2, \ldots, v_{2r+q}: v_i \ (1 \leq i \leq 2s + q) \text{ is a child of } l\}$ (see Fig. 4). Then obviously $\text{pdw}(T) \leq 2s + n + 1$.

Second, we will show that if $X3C$ does not have an exact cover, then $\text{pdw}(G) \geq \frac{8}{3}s$. Let $(X_0, X_1, \ldots, X_d)$ be an optimal decomposition for $G$. Without loss of generality, we can assume that:

(A1) the subset $C_1$ satisfies that $C_1 = \{e_{m-2}, e_{m-1}, e_m\}$ and $C_1 \cap \mathcal{C}' = \emptyset$ for all $C' \in \mathcal{C} - C_1$ (if there is no such set $C_1$, then modify the instance by adding new elements $e_{m+1}, e_{m+2}$, $e_{m+3}$ to $X$ and adding new set $\{e_{m+1}, e_{m+2}, e_{m+3}\}$ to $\mathcal{C}'$), and

(A2) $e_{m-3}$ occurs in exactly three subsets (if there is no such element then the instance can be solved in polynomial time, see [8]).

Let us consider the following cases.

Case 1: $X_0 \not\subseteq \{l, r\} \cup \{v: \text{(the parent of } v\text{)} \in \{l, r, c_1\}\} \cup \{c_j: 1 \leq j \leq n\}$. From the assumption (A1), $c_j \ (2 \leq j \leq n)$ does not have children. Thus in this case $X_0$ has a node $v$ such that $\text{dist}_T(\text{root}, v) \leq m - 1$. Hence, we have $d \leq 2m$. Let us define the weight of a node $v$ as $k = |h/s| \times s$, where $h$ is the number of neighbors of $v$ with degree 1, and denote the total of weight of nodes in $S \subseteq V(G)$ by $W(S)$. It is easy to see that if for some $i \ (2 \leq i \leq d - 1)$ $W(X_i) \geq 3s$, then $|X_{i+1}| \geq 3s$. That is, $\text{pdw}(G) < 3s$ implies $W(X_i) \leq 2s$ for all $2 \leq i \leq d - 1$. From $d \leq 2m$, if $\text{pdw}(G) < 3s$ then $W(X_2 \cup X_3 \cup \cdots \cup X_{d-1})$ is at most $2s((d - 1) - 2 + 1) \leq 2s(2m - 2)$. Note that $W(X_d) = 0$ as $d > 2$. Since $W(V(G))$ is at least $s(4m + 4)$, $W(X_0 \cup X_1)$ is at least $8s$. Thus $|X_0 \cup X_1 \cup X_2|$ is at least $8s$. Therefore, $\text{pdw}(G) \geq \frac{8}{3}s.$
Fig. 3. Tree $T$.

Fig. 4. A layout satisfying $pdw(T) \leq 2s + n + 1$. 
Case 2: \( X_0 \subseteq \{ l, r \} \cup \{ v : (\text{the parent of } v) \in \{ l, r, c_1 \} \} \cup \{ c_j : 1 \leq j \leq n \} \) and \( l \notin X_0 \).

Let \( i \) be the index such that \( l \in X_i \). Then in this case \( i \geq 1 \). If \( i \geq 2 \) then the \((2s + q) + (s + q)\) children of \( l \) have to be in \( X_{i+1} \). Thus, \( p\text{d}w(T) \geq 3s \). If \( i = 1 \) then we have two subcases root \( \in X_0 \) and root \( \in X_{m+1} \). If root \( \in X_{m+1} \) then \( c_j \notin X_0 \) for all \((2 \leq j \leq n)\), which means \( p\text{d}w(G) \geq 3s \) (recall the assumption (A1) and (A2)). If root \( \in X_m \), then the \( 2s + p \) children of root have to be in \( X_{m+1} \). From \( i = 1 \), the \((s + q) + p\) children of the first node on path \( L \) have to be in \( X_{m+1} \). Thus, we have \( p\text{d}w(G) \geq ((s + q) + p) + (2s + p) \).

Case 3: \( X_0 \subseteq \{ l, r \} \cup \{ v : (\text{the parent of } v) \in \{ l, r, c_1 \} \} \cup \{ c_j : 1 \leq j \leq n \} \), \( l \in X_0 \), and \( \{ v : v \text{ is a child of } r \} \cap X_0 \neq \emptyset \).

From \( l \in X_0 \), there are at least \( s + q \) nodes of distance 3 from \( l \). In this case, as \( r \notin X_0 \), the parent of \( r \) in \( T \) is in \( X_2 \), so \( X_3 \) has other \( 2s + p \) nodes of distance 3. Thus, \(|X_3| \geq (s + q) + (2s + p)\) (see left-hand side in Fig. 5).

Case 4: \( X_0 \subseteq \{ l \} \cup \{ v : (\text{the parent of } v) \in \{ l, r, c_1 \} \} \cup \{ c_j : 1 \leq j \leq n \} \), \( l \in X_0 \), and \( \{ v : v \text{ is a child of } r \} \cap X_0 \neq \emptyset \).

Without loss of generality we can assume \( c_x \in X_0 \) and \( c_y, c_z \notin X_0 \) otherwise we easily get \( p\text{d}w(G) \geq 3s \). Since \( l, c_x \in X_0 \) and \( c_1 \in X_1 \), we have \(|X_4| \geq (s + 1) + (s + 1) + (s + q + 1)\) (see the right-hand side in Fig. 5).
Case 6: $X_0 \subseteq \{l\} \cup \{v: \text{(the parent of } v) \in \{l, c_1\}\} \cup \{c_j: 1 \leq j \leq n\}$ and $l, c_1 \in X_0$. We can assume that $2s + q$ children of $l$ are in $X_0$, other $s + q$ children of $l$ are in $X_1$, and the children of $c_1$ are in $X_1$, because each $c_j$ ($2 \leq j \leq n$) does not have children (i.e., leaves) and $l, c_1 \in X_0$. Now it is easy to see that if $\mathcal{X}_3C$ does not have exact set cover, then $pdw(G) \geq 3s$.

As a result, in all cases we have seen that $pdw(G)$ is at least $\frac{3}{8}s$. Therefore, if $\mathcal{X}_3C$ does not have an exact cover, then $pdw(G) \geq \frac{3}{8}s$.

Suppose that there exists a $\frac{4}{3} - \varepsilon$ polynomial time approximation algorithm $A_\varepsilon$ for some $\varepsilon > 0$. Then we set $s > (4/3 - \varepsilon)(n+1)/2\varepsilon$, which means $\frac{3}{8}s > (\frac{4}{3} - \varepsilon)(2s + n + 1)$. It is clear that if $\mathcal{X}_3C$ has an exact cover, then the output of $A_\varepsilon$ is at most $2s + n + 1$. On the other hand, if $\mathcal{X}_3C$ does not have an exact cover then the output of $A_\varepsilon$ is at least $\frac{3}{8}s$. This means that EXACT COVER BY 3-SETS can be solved in polynomial time. Hence, we have shown that approximating the path–distance–width problem on trees is NP-hard for a factor of $\frac{4}{3} - \varepsilon$. □

4. Conclusion

We have shown that the path–distance–width problem (for short PDWP) does not have a PTAS. PDWP is related to the bandwidth problem (for short BWP). Recently, polynomial time approximation algorithms with a polylogarithmic approximation factor have been presented for BWP [3,6]. The author conjectures that PDWP have a polynomial time approximation algorithm with a polylogarithmic approximation factor. The technique volume respecting Euclidean embeddings introduced in [6] might be useful to derive such an approximation algorithm.

On the other hand, it have been shown that for any constant $k$ there is no polynomial time approximation algorithm with a constant approximation factor of $k$ for BWP [10]. It is also known that BWP can be approximated within a constant for dense graphs [9]. So the author conjectures that there is no polynomial time approximation algorithm with a constant approximation factor for PDWP, and PDWP has a PTAS for dense graphs (cf. [7]).

As another interesting further work, there is a graph–theoretical problem: What is the path–distance–width of complete $k$-ary trees? We do not know even if $k = 2$.

References


