Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 80 (2005)

Proving NP-hardness for clique-width I: non-approximability of sequential clique-width

Michael R. Fellows[*], Frances A. Rosamond[*],
Udi Rotics[†], and Stefan Szeider[‡]

August 24, 2005

Abstract

Clique-width is a graph parameter, defined by a composition mechanism for vertex-labeled graphs, which measures in a certain sense the complexity of a graph. Hard graph problems (e.g., problems expressible in Monadic Second Order Logic, that includes NP-hard problems) can be solved efficiently for graphs of certified small clique-width. It is widely believed that determining the clique-width of a graph is NP-hard; in spite of considerable efforts, no NP-hardness proof has been found so far.

In this paper we show a non-approximability result for restricted form of clique-width, termed "r-sequential clique-width", considering only such clique-width constructions where one of any two graphs put together by disjoint union must have r or fewer vertices. In particular, we show that for every positive integer r, the r-sequential clique-width cannot be absolutely approximated in polynomial time unless P = NP, and that given G and k the question of whether the r-sequential clique-width of G is at most k is NP-complete.

We show further that this non-approximability result holds even for graphs of a very particular structure: for graphs obtained from cobipartite graphs by replacing edges with induced paths. In part II of this series of papers we use this strengthened result to show that, unless P = NP, there is no polynomial-time absolute approximation algorithm for (unrestricted) clique-width, and that, given a graph G and an integer k, deciding whether the the clique-width G is at most k is NP-complete. This solves a problem that has been open since the introduction of clique-width in the early 1990s.

1 Introduction

Clique-width is a graph parameter that measures in a certain sense the complexity of a graph. This parameter was first considered by Courcelle, Engelfriet, and Rozenberg [5] (the term clique-width was introduced later). The clique-width of a graph is the smallest number of labels that suffices to construct the graph using the operations: creation of a new vertex v with label i, disjoint union, insertion of edges between vertices of certain labels, and relabeling of vertices. Such a construction of a graph by means of these four operations using at most k different labels can be represented by an algebraic expression called a *k-expression*. (More exact definitions are provided in Section 2.) By a general result of Courcelle, Makowsky, and

[*]School of Electrical Engineering and Computer Science, University of Newcastle, Callaghan 2308 NSW, Australia, `[mfellows|fran]@cs.newcastle.edu.au`.

[†]School of Computer Science and Mathematics, Netanya Academic College, Netanya, Israel, `rotics@mars.netanya.ac.il`.

[‡]Department of Computer Science, Durham University, Durham, England, UK, `stefan.szeider@durham.ac.uk`.

ISSN 1433-8092

Rotics [6], any graph problem that can be expressed in Monadic Second Order Logic with second-order quantification on vertex sets (that includes NP-hard problems) can be solved in linear time for graphs of clique-width bounded by some constant k if the k-expression is provided as input to the algorithm (albeit the running time involves a constant which is exponential in k).

A main limit for applications of this result is that first one has to find a k-expression for the given graph (or decide that the clique-width of the given graph exceeds a given limit). The question whether there is a polynomial-time algorithm for computing the clique-width of a graph was already raised by Courcelle, et al. [5] in 1990.

Clique-width can be considered to be more general than treewidth since (i) there are graphs of constant clique-width but arbitrarily high treewidth, e.g., complete graphs, but (ii) graphs of bounded clique-width have also bounded treewidth. The latter was shown by Courcelle and Olariu [7]; the upper bound on the clique-width of a graph in terms of its treewidth was improved by Corneil and Rotics [4]. Since the computation of treewidth is well-known to be NP-hard (Arnborg, Corneil, and Proskurowski [1]), it is obvious to assume that such a hardness result should also hold for the more general parameter clique-width. However, no hardness result for clique-width similar to to the treewidth result [1] is known!

With considerable efforts, polynomial-time algorithms could be developed for recognizing graphs of clique-width at most 3 (see Corneil, Habib, Lanlignel, Reed, and Rotics [3]). Recently, Oum and Seymour [16] obtained an algorithm that, for any fixed k, runs in time $O(n^9 \log n)$ and computes $(2^{3k+2} - 1)$-expressions for graphs of clique-width at most k. This result is interesting, as it makes the notion "graph class of bounded clique-width" feasible; however, since the running time of algorithms as suggested by Courcelle et al. [6] crucially depends on k, closer approximations are desirable.

The graph parameter "NLC-width", introduced by Wanke [18], is defined similarly as clique-width, however a single operation that combines disjoint union and insertion of edges is used. Recently Gurski and Wanke [11] have reported that computing the NLC-width is NP-hard. Since NLC-width and clique-width can differ by a factor of 2 (see Johansson [13]), non-approximability with an absolute error guarantee for one of the two parameters does not imply a similar result for the other parameter.

In the present paper we show that the computation of a restricted form of clique-width, termed *r-sequential clique-width* (or simply *sequential clique-width* for $r = 1$), is NP-hard. Here we consider only clique-width constructions with skew disjoint unions; that is, where at least one of any two k-graphs put together by disjoint union is of order r or less (r is an arbitrarily large constant).

Sequential clique-width as a special case of clique-width can be considered as an analog to pathwidth as a special case of treewidth; trees corresponding to sequential clique-width constructions are path-like. The natural clique-width constructions of complete graphs (see Section 2 for an example) are sequential.

Our main result can be stated as follows.

(1) For every $r \geq 1$, the r-sequential clique-width of a graph cannot be approximated in polynomial time with an absolute error guarantee, unless P = NP.

(2) The following decision problem is NP-complete for any $r \geq 1$:

> r-SEQUENTIAL CLIQUE-WIDTH MINIMIZATION
> *Instance:* A graph G and a positive integer k.
> *Question:* Is the r-sequential clique-width of G at most k?

Furthermore, we obtain structural results relating the parameters clique-width, sequential clique-width, and r-sequential clique-width:

(3) For every $r \geq 2$, the sequential clique-width of a graph exceeds the r-sequential clique-width at most by r.

(4) For every $r \geq 1$ there exist graphs of constant clique-width but arbitrarily high r-sequential clique-width.

We show that the hardness results (1) and (2) hold even for a special graph class $\mathcal{D}$ consisting of graphs obtained from cobipartite graphs (complements of bipartite graphs) by replacing edges by induced paths. In a companion paper [9], we show that—in contrast to (4)—clique-width and sequential clique-width of graphs in $\mathcal{D}$ differ at most by a small constant. This implies that the non-approximability of sequential clique-width carries over to (general) clique-width. Whence computing the clique-width of a graph is NP-hard; this solves the outstanding problem that has been open for 15 years.

The key idea of our approach is to show NP-hardness of one graph parameter ((sequential) clique-width) by means of the non-approximability of another graph parameter (pathwidth). Such an approach might be applicable for showing intractability of other graph parameters.

2 Notation and preliminaries

A *layout* of a graph G with n vertices is a bijection $\varphi : V(G) \to \{1, \ldots, n\}$. For a layout φ of G we define the sets of vertices

$$\begin{aligned} L_G(i,\varphi) &= \{\, u \in V(G) : \varphi(u) \leq i \,\}, \\ R_G(i,\varphi) &= \{\, u \in V(G) : \varphi(u) > i \,\}, \\ L^*_G(i,\varphi) &= \{\, v \in L_G(i,\varphi) : \exists u \in R_G(i,\varphi) \text{ such that } uv \in E(G) \,\}, \\ R^*_G(i,\varphi) &= \{\, v \in R_G(i,\varphi) : \exists u \in L_G(i,\varphi) \text{ such that } uv \in E(G) \,\}. \end{aligned}$$

We call the maximum cardinality of $R^*_G(i,\varphi)$ and the maximum cardinality of $L^*_G(i,\varphi)$ the *in-degree* and the *out-degree* of the layout φ, respectively. The *vertex separation number* $\mathrm{vsn}(G)$ of G is defined as the smallest in-degree over all layouts of G (which equals the smallest out-degree over all layouts of G).

Let T be a tree and χ a labeling of the vertices of T by sets of vertices of G. The pair (T, χ) is a *tree decomposition* of G if (i) every vertex of G belongs to $\chi(t)$ for some vertex $t \in V(T)$; (ii) for every edge $vw \in E(G)$ there is some $t \in V(T)$ with $v, w \in \chi(t)$; (iii) for any vertices $t_1, t_2, t_3 \in V(T)$, if t_2 lies on a path from t_1 to t_3, then $\chi(t_1) \cap \chi(t_3) \subseteq \chi(t_2)$. The *width* of (T, χ) is the maximum $|\chi(t)| - 1$ over all vertices t of T. The *treewidth* $\mathrm{twd}(G)$ of G is the minimum width over all tree-decompositions of G. The *pathwidth* $\mathrm{pwd}(G)$ of G is the minimum width over all tree-decompositions (T, χ) of G where T is a path.

It is well-known that pathwidth and vertex separation number of a graph agree (see Kinnersley [15]).

Let k be a positive integer. A *k-graph* is a graph whose vertices are labeled by integers from $\{1, \ldots, k\}$. We consider an arbitrary graph as a k-graph with all vertices labeled by 1. We call the k-graph consisting of exactly one vertex v (say, labeled by $i \in \{1, \ldots, k\}$) an *initial k-graph* and denote it by $i(v)$. If a vertex v of a k-graph G is the only vertex with label i then we call v a *singleton*.

The *clique-width* $\mathrm{cwd}(G)$ of a graph G is the smallest integer k such that G can be constructed from initial k-graphs by means of repeated application of the following three operations.

- *Disjoint union* (denoted by $\oplus$);
- *Relabeling*: changing all labels i to j (denoted by $\rho_{i \to j}$);
- *Edge insertion*: connecting all vertices labeled by i with all vertices labeled by j (denoted by $\eta_{i,j}$).

We call the construction of a k-graph using the above operations a *clique-width construction*. A clique-width construction can be represented by an algebraic term composed of $\oplus$, $\rho_{i\to j}$, and $\eta_{i,j}$, ($i, j \in \{1, \dots, k\}$, and $i \neq j$). Such a term is called a *k-expression* defining G.

For example, the complete graph on the vertices u, v, w, x is defined by the 2-expression

$$\rho_{2\to 1}(\eta_{1,2}(\rho_{2\to 1}(\eta_{1,2}(\rho_{2\to 1}(\eta_{1,2}(2(u) \oplus 1(v))) \oplus 2(w))) \oplus 2(x)))$$

In general, every complete graph K_n, $n \geq 2$, has clique-width 2.

3 The r-sequential clique-width

In the sequel we consider clique-width constructions where disjoint union of two k-graphs is only allowed if at least one of them has r or fewer vertices. We call such clique-width constructions and the corresponding k-expressions *r-sequential* (or *sequential* for $r = 1$). The *r-sequential clique-width* of a graph G, denoted by $\mathrm{cwd}_r(G)$, is defined as the smallest k such that G can be defined by an r-sequential k-expression[1]. For example, the above 2-expression defining K_4 is sequential. In general, we have $\mathrm{cwd}_1(K_n) = \mathrm{cwd}(K_n)$ for every $n \geq 1$.

It is convenient to consider a sequential k-construction as a process where to some initial k-graph a sequence of operations is applied, defining the addition of a new vertex as a single operation

$$\alpha_{i(v)}(G) = G \oplus i(v).$$

Thus we can rewrite the above sequential 2-expression for K_4 as the sequence

$$1(u),\ \alpha_{2(v)},\ \eta_{1,2},\ \rho_{2\to 1},\ \alpha_{2(w)},\ \eta_{1,2},\ \rho_{2\to 1},\ \alpha_{2(x)},\ \eta_{1,2},\ \rho_{2\to 1}.$$

The next lemma shows that by considering r-sequential clique-width instead of sequential clique-width we cannot save more than r labels.

Lemma 1. *$\mathrm{cwd}_r(G) \leq \mathrm{cwd}_1(G) \leq \mathrm{cwd}_r(G) + r$ holds for every graph G and every $r \geq 1$.*

Proof. Let G be a k-graph and X an r-sequential k-expression of G. We show by induction on $n = |V(G)|$ that G has a sequential $(k + r)$-expression Y. If $n = 1$ then we simply put $Y = X$, hence assume $n > 1$. It follows that X describes a clique-width construction where G is obtained by edge insertions and relabelings from $G' \oplus H$; G', H are k-graphs with $\mathrm{cwd}_r(G') \leq k$ and $|V(H)| \leq r$. Let $V(H) = \{u_1, \dots, u_p\}$ and $E(H) = \{u_{a_1}u_{b_1}, \dots, u_{a_q}u_{b_q}\}$, $p \leq r$ and $q \leq r^2$. We construct G from G' by means of the following three steps.

1. First we add the vertices of H to G' using new labels $k + 1, \dots, k + p$,

   $$G'' = G' \oplus (k + 1)(u_1) \oplus \cdots \oplus (k + p)(u_p).$$

   Note that the disjoint unions are in accordance with the requirements of sequential clique-width.

2. Next we apply to G'' the edge insertions $\eta_{k+a_i,k+b_i}$, $i = 1, \dots, q$, and obtain the $(k + r)$-graph G'''. We observe that G''' and $G' \oplus H$ only differ in the labeling of the vertices of H. Since the vertices of H are singletons in G''', we can apply relabelings (as described by X) to obtain $G' \oplus H$ from G'''.

3. By assumption we can obtain G from $G' \oplus H$ by edge insertions and relabelings.

[1]Sequential cliquewidth was also considered by Gurski and Wanke [12] under the term linear clique-width.

The induction hypothesis applies to G', hence there is a sequential k-expression Y' defining G'. According to the three construction steps described above we extend Y' to a sequential $(k+r)$-expression Y defining G. Hence the induction proof is completed and the lemma follows. $\square$

The upper bound of Lemma 1 can be improved to $\mathrm{cwd}_1(G) \le \mathrm{cwd}_r(G) + f(r)$ where $f(r)$ is the largest clique-width of graphs with r vertices. In particular, since the clique-width of a graph with $r > 2$ vertices is at most $r-1$, we have $\mathrm{cwd}_1(G) \le \mathrm{cwd}_r(G) + r - 1$ for $r > 2$.

4 Proof of the main results

This section is devoted to the proof of our main results, namely, the non-approximability of r-sequential clique-width and the NP-completeness of the minimization problem.

Construction 1. *Let G denote a fixed simple graph with $n \ge 2$ vertices. We obtain a graph G' from G by replacing each edge uv of G by three internally disjoint paths (u, x_i, y_i, v), $i = 1, 2, 3$, of length 3 (see Figure 1); we call such paths* bridges.

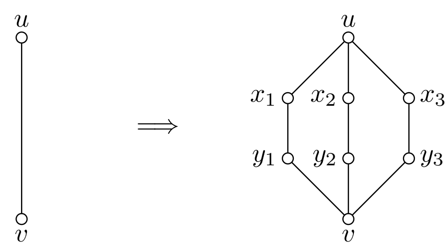

Figure 1: Construction of G'.

For the remainder of this section let G denote a fixed simple graph with $n \ge 2$ vertices and let G' denote the graph obtained from G by means of Construction 1.

Lemma 2. *Given a layout $\varphi : V(G) \to \{1, \ldots, n\}$ of G with out-degree k, we can construct in polynomial time a sequential $(k+4)$-expression defining G'. Consequently,* $\mathrm{cwd}_1(G') \le \mathrm{vsn}(G) + 4 = \mathrm{pwd}(G) + 4$.

Proof. For $i = 1, \ldots, n$ let Γ_i denote the set of vertices of G' that belong to $L_G(i, \varphi)$ or are of distance at most 2 apart from $L_G(i, \varphi)$. (Thus, if at least one end of a bridge b belongs to $L_G(i, \varphi)$, then both internal vertices of b belong to Γ_i.) Let Δ_i denote the subset of Γ_i consisting of vertices that are adjacent in G' with vertices outside of Γ_i. Furthermore, let G'_i denote the subgraph of G' induced by the set Γ_i.

We inductively obtain sequential $(k+4)$-expressions defining k-graphs G_i, $i = 1, \ldots, n$, such that the labeling of G_i satisfies the following conditions.

1. vertices in $\Gamma_i \setminus \Delta_i$ are labeled by 1;
2. vertices in Δ_i are labeled by integers from $5 \ldots, k+4$;
3. two vertices of Δ_i share the same label if and only if both vertices have a common neighbor in G'.

We construct G'_1 as follows. Let $f : R^*_G(1, \varphi) \to \{5, \ldots, k+4\}$ be an injective map (such map exists since $|R^*_G(1, \varphi)| \le k$). We introduce $u = \varphi^{-1}(1)$ as initial $(k+4)$-graph $2(u)$, and for every pair x, y of vertices that lie on a bridge between u and some $v \in R^*_G(1, \varphi)$ we apply the operations

$$\alpha_{3(x)},\ \eta_{2,3},\ \alpha_{4(y)},\ \eta_{3,4},\ \rho_{3\to 1},\ \rho_{4\to f(v)}.$$

Finally, we relabel u with 1, using the operation $\rho_{2\to 1}$. This gives a k-expression defining G'_1 and the claimed properties are evidently satisfied. Now assume that we have already a k-expression X_{i-1} defining G'_{i-1} for some $i \in \{2, \ldots, n\}$ with a labeling that satisfies the claimed properties. We extend X_{i-1} to a k-expression X_i defining G_i as follows. First we add a new vertex $u = \varphi^{-1}(i)$ labeled with 2 by disjoint union of the initial $(k+4)$-graph $2(u)$ and X_{i-1}. Note that u is the only vertex labeled with 2. For vertices $v \in R^*_G(i, \varphi)$ let $\Delta_{i-1}(v)$ denote the set of vertices in Δ_{i-1} that are adjacent to v in G'. By assumption, there is an injective map $f : R^*_G(i-1, \varphi) \to \{5, \ldots, k+4\}$ such that all vertices in $\Delta_{i-1}(u)$ have the same label $f(u)$ in G'_{i-1}, and no other vertex of G'_{i-1} is labeled with $f(u)$. Hence we can make the vertices in $\Delta_{i-1}(u)$ adjacent to u and relabel them with 1 using the operations $\eta_{f(u),2}$ and $\rho_{f(u)\to 1}$, respectively.

Since $|R^*_G(i, \varphi)| \leq k$, we can define an injective map $f' : R^*_G(i, \varphi) \to \{5, \ldots, k+4\}$ with $f'(v) = f(v)$ for $v \in R^*_G(i-1, \varphi) \cap R^*_G(i, \varphi)$. As above, for every pair x, y of vertices that lie on a bridge between u and some $v \in R^*_G(1, \varphi)$ we apply the operations

$$\alpha_{3(x)},\ \eta_{2,3},\ \alpha_{4(y)},\ \eta_{3,4},\ \rho_{3\to 1},\ \rho_{4\to f'(v)}.$$

Finally, we relabel u with 1, using the operation $\rho_{2\to 1}$.

It is straightforward to verify that after performing the described construction steps we are left with a k-graph G'_i that satisfies the claimed properties; the construction can be described by a sequential $(k+4)$-expression X_i. Since $G'_n = G'$, it follows that the sequential clique-width of G' is at most k. The $(k+4)$ -expression X_n can certainly be constructed in time proportional to $|E(G')| + |V(G')|$, hence the lemma is shown true. □

The next lemma will allow us to bound the vertex separation number of G in terms of the sequential clique-width of G', a result inverse to Lemma 2. To this end let us fix a sequential k-expression X defining G'. X gives rise to a sequence $G'_1, \ldots, G'_s$ of k-graphs such that G'_1 is an initial k-graph, $G'_s = G'$, and G'_i is obtained from G'_{i-1} by one of the operations η, ρ, and α $(i = 2, \ldots, s)$. For every edge $e \in E(G')$ let $j(e) := \min\{\, 1 \leq j \leq s : e \in E(G'_j) \,\}$. We call a bridge (u, x, y, v) *well-behaved* if u is a singleton in $G'_{j(ux)}$ and v is a singleton in $G'_{j(yv)}$.

Lemma 3. *At least one of any three parallel bridges of G' is well-behaved.*

Proof. For an edge $uv \in E(G)$ let $b_i = (u, x_i, y_i, v)$, $i = 1, 2, 3$, denote the parallel bridges of G'. For $i = 1, 2, 3$ we put $\alpha_i = \max(j(ux_i), j(y_iv))$.

Claim A: $j(ux_i)$ and $j(y_iv)$ must be distinct for $i = 1, 2, 3$. Otherwise, either u would have the same label as y_i or the same label as v in $G'_{j(ux_i)}$. In the first case, the addition of the edge y_iv causes the addition of the edge uv. In the second case, the addition of the edge y_iv causes the addition of the edge y_iu. However, neither uv nor y_iu is present in G'. Hence Claim A is shown.

Claim B: if $j(ux_i) < j(y_iv)$, then u is singleton in $G'_{j(ux_i)}$ for $i = 1, 2, 3$. Assume to the contrary that there is a vertex $w \in V(G'_{j(ux_i)}) \setminus \{u\}$ which shares the label with u. It follows that $wx_i \in E(G')$, hence $w = y_i$. This, however, implies that in $G'_{j(y_iv)}$ the edge uv is inserted, a contradiction. Hence Claim B is shown.

Now we proceed with the proof of the lemma. We consider two cases.

Case 1: $|\{\alpha_1, \alpha_2, \alpha_3\}| \leq 2$. We assume, w.l.o.g., $\alpha_1 = \alpha_2 = j(y_1v)$. Clearly $\alpha_2 = j(y_2v)$, since otherwise, if $\alpha_2 = j(ux_2)$, then some of the edges uy_1, uv were present in G'. Let w be a vertex of $G'_{j(y_1v)}$ that shares the label with v. It follows that $wy_1, wy_2 \in E(G')$, hence

$w = v$. Thus v is a singleton in $G'_{j(y_1v)}$. Since $j(ux_1) < j(y_1v)$, it follows from Claim B that u is a singleton in $G'_{j(ux_1)}$. Hence the bridge b_1 is well-behaved.

Case 2: $|\{\alpha_1, \alpha_2, \alpha_3\}| = 3$. We assume, w.l.o.g., that $j(y_1v) = \alpha_1 < \alpha_2 < \alpha_3$.

Subcase 2a: $j(y_2v) > j(y_1v)$ or $j(y_3, v) > j(y_1v)$. W.l.o.g., $j(y_2v) > j(y_1v)$. Similarly as above we conclude that for any vertex w of $G'_{j(y_1v)}$ that shares the label with v, the edges y_1w, y_2w are added in $G'_{j(y_1v)}$, $G'_{j(y_2v)}$, respectively. Hence $w = v$ and so v is a singleton in $G'_{j(y_iv)}$. Furthermore, since $j(ux_1) < j(y_1v)$, it follows by Claim B that u is a singleton in $G'_{j(ux_1)}$. Hence the bridge b_1 is well-behaved.

Subcase 2b: $j(y_2v) \leq j(y_1v)$ and $j(y_3, v) \leq j(y_1v)$. It follows that $\alpha_2 = j(ux_2)$ and $\alpha_3 = j(ux_3)$. We show that u is a singleton in $G'_{j(ux_2)}$. Let w be a vertex of $G'_{j(ux_2)}$ that shares the label with u. Consequently, the edges wx_2, wx_3 are added in $G'_{j(ux_2)}$ and $G'_{j(ux_3)}$, respectively. Thus $u = w$ and so u is indeed a singleton in $G'_{j(ux_2)}$. Using a symmetrical version of Claim B, we conclude from $j(y_2v) < j(ux_2)$ that v is a singleton in $G'_{j(y_2v)}$. Hence the bridge b_2 is well-behaved. □

Lemma 4. *From a sequential k-expression defining G' we can construct in polynomial time a layout for G with out-degree at most k. Consequently,* $\mathrm{vsn}(G) = \mathrm{pwd}(G) \leq \mathrm{cwd}_1(G')$.

Proof. For a vertex $v \in V(G)$ let $\beta(v)$ denote the smallest integer in $\{1, \ldots, s\}$ such that v is not a singleton of $G'_{\beta(v)}$. Note that $\beta(v)$ is defined for every v of $V(G)$, since we assume that G has more than one vertex and all vertices of the final G' have label 1. Note also that if $\beta(v) = \beta(v') = j$ holds for two vertices $v, v' \in V(G)$, then v and v' have the same label in G'_j, but no other vertex in G'_j shares its label with v and v' (either v and v' are singletons in G'_{j-1} and one of the two vertices is relabeled with the other's label in G'_j, or one of the two vertices is a singleton in G'_{j-1} and the other vertex is introduced in G'_j with the same label). Let $\varphi : V(G) \to \{1, \ldots, n\}$ be a layout satisfying $\varphi(v) < \varphi(v')$ whenever $\beta(v) < \beta(v')$.

It remains to show that the out-degree of the layout φ is at most k. Choose $i \in \{1, \ldots, n-1\}$ arbitrarily. We show that $|R^*_G(i, \varphi)| \leq k$. Let $w = \varphi^{-1}(i)$, $j = \beta(w)$, and consider the graph G'_j. By construction, the vertices of $L_G(i, \varphi)$ are not singletons of G'_j. We assign to every vertex $v \in R^*_G(i, \varphi)$ a label $f(v) \in \{1, \ldots, k\}$ as follows (it will turn out that f is an injective map). Choose arbitrarily a vertex $v \in R^*_G(i, \varphi)$. By definition, v is in G adjacent to a vertex $u \in L_G(i, \varphi)$. Thus u and v are joined by three parallel bridges in G'. By Lemma 3, at least one of the bridges between u and v, say $b = (u, x_v, y_v, v)$, is well-behaved. For vertices z of G'_j let $\ell(z)$ denote the label of z in G'_j. We put

$$f(v) = \begin{cases} \ell(v) & \text{if } v \in V(G'_j); & \text{(case 1)} \\ \ell(y_v) & \text{if } v \notin V(G'_j) \text{ and } y_v \in V(G'_j); & \text{(case 2)} \\ \ell(x_v) & \text{if } v, y_v \notin V(G'_j). & \text{(case 3)} \end{cases}$$

Since u is not a singleton in G'_j, the edge ux_v must already be present in G'_j as the bridge (u, x_v, y_v, v) is well-behaved. Consequently the above case distinction is exhaustive. We split the set $R^*_G(i, \varphi)$ into sets C_1, C_2, and C_3, such that a vertex v belongs to C_i if $f(v)$ is assigned by means of the above case i. We further split C_1 into sets $C_1^=$ and $C_1^<$ such that $v \in C_1$ belongs to $C_1^=$ if $\beta(w) = \beta(v)$ and v belongs to $C_1^<$ if $\beta(w) < \beta(v)$.

To show that f is an injective map, suppose to the contrary that $f(v) = f(v')$ for two distinct vertices $v, v' \in R^*_G(i, \varphi)$. Since the vertices of $C_1^<$ are singletons in G'_j, $v, v' \notin C_1^<$ follows. For any $v \in C_3$, the vertex x_v is a singleton in G'_j since the edge x_vy_v is still missing, hence $v, v' \notin C_3$. Furthermore, v and v' cannot both belong to $C_1^=$ since then both would share the label with w in G'_j, but as seen above, any $v \in C_1^=$ shares its label only with w. Similarly, if $v \in C_1^=$ and $v' \in C_2$, then v and $x_{v'}$ would share the label with w in G'_j, which is not possible for the same reason. Hence we are left with the case $v, v' \in C_2$.

Thus $f(v)$ is the label of y_v and $f(v')$ is the label of $y_{v'}$. The edges $y_v v, y_{v'} v'$ are not yet present in G'_j since the vertices v, v' are not yet present in G'_j either. If at a further step the edge $y_v v$ is added, also the edge $y_v v'$ is added, in contradiction to $y_v v' \notin E(G')$. Thus $f : R^*_G(i, \varphi) \to \{1, \ldots, k\}$ is indeed an injective map, and so $|R^*_G(i, \varphi)| \leq k$ follows. $\square$

In the proof of the next theorem we shall use a result of Bodlaender, Gilbert, Hafsteinsson, and Kloks [2], which states that, unless P = NP, there is no polynomial-time approximation algorithm for the pathwidth (i.e., the vertex separation number) of a graph G with an absolute error of at most $|V(G)|^\varepsilon$ for any $\varepsilon \in (0, 1)$. Moreover, Karpinski and Wirtgen [14] observed that this non-approximability result also holds for cobipartite graphs.

Construction 2. *Given a graph G, and an integer $q \geq 1$, we obtain a graph G^q by replacing each vertex v of G by q vertices $v_1, \ldots, v_q$, and by joining two vertices v_i, w_j by an edge if either $v = w$ and $i \neq j$ or $v \neq w$ and $vw \in E(G)$.*

Note that if G is cobipartite then so is G^q. Bodlaender, et al. [2] show the following equation (the proof stated in [2] for treewidth applies literally to pathwidth as well).

$$\mathrm{pwd}(G^q) = q(\mathrm{pwd}(G) + 1) - 1. \tag{1}$$

Let α be an integer-valued graph parameter. We consider the following decision problem.

α-MINIMIZATION

Instance: A graph G and a positive integer k.

question: Is $\alpha(G)$ at most k?

Arnborg, et al. [1] have shown that pwd-MINIMIZATION is NP-complete, even for cobipartite graphs.

Lemma 5. *Assume that there is a constant c such that $|\alpha(G) - \mathrm{pwd}(G)| \leq c$ holds for every cobipartite graph G with minimum degree at least 3. Then the following statements are true.*

1. *For a graph G with n vertices and minimum degree at least 3, $\alpha(G)$ cannot be approximated in polynomial-time with an absolute error guarantee of n^ε for any $\varepsilon \in (0, 1)$ unless* P = NP.

2. *α-MINIMIZATION is* NP*-hard.*

Proof. Part 1. Let $\varepsilon \in (0, 1)$ be a fixed constant, and assume that there exists a polynomial-time algorithm $\mathcal{A}$ that outputs for a given graph G with n vertices and minimum degree at least 3 an integer $\mathcal{A}(G)$ such that $|\mathcal{A}(G) - \alpha(G)| \leq n^\varepsilon$. Let G be any given cobipartite graph with n vertices. We form the graph G^3 using Construction 2. Note that G^3 is a cobipartite, of minimum degree at least 3, and the number of vertices of G^3 is $3n$. We apply algorithm $\mathcal{A}$ to G^3 and get $|\mathcal{A}(G^3) - \alpha(G^3)| \leq (3n)^\varepsilon$. The assumption of the lemma gives $|\mathcal{A}(G^3) - \mathrm{pwd}(G^3)| \leq (3n)^\varepsilon + c$. Using equation (1) we get $|\mathcal{A}(G^3)/3 - \mathrm{pwd}(G)| \leq 3^{\varepsilon - 1} n^\varepsilon + (c + 2)/3 \leq n^\varepsilon + (c + 2)/3$. For sufficiently large n, $|\mathcal{A}(G)/3 - \mathrm{pwd}(G)| \leq n^{\sqrt{\varepsilon}}$ follows, which, by the aforementioned result of Karpinski and Wirtgen [14], is not possible unless P = NP.

Part 2. We reduce pwd-MINIMIZATION to α-MINIMIZATION. Let (G, k) be an instance of pwd-MINIMIZATION. Since pwd-MINIMIZATION is NP-hard for cobipartite graphs, we may assume that G is cobipartite. We obtain in polynomial time an instance (G^*, k^*) of α-MINIMIZATION, putting $G^* = G^{2c+3}$ and $k^* = (2c + 3)(k + 1) + c - 1$. Observe that G^* is cobipartite and has minimum degree at least 3. We show that (G, k) is a yes-instance of pwd-MINIMIZATION if and only if (G^*, k^*) is a yes-instance of α-MINIMIZATION; that is,

$\mathrm{pwd}(G) \le k$ if and only if $\alpha(G^*) \le k^*$. First assume $\mathrm{pwd}(G) \le k$. Now by equation (1), $\alpha(G^*) \le \mathrm{pwd}(G^*) + c = (2c+3)(\mathrm{pwd}(G)+1) - 1 + c \le (2c+3)(k+1) + c - 1 = k^*$, thus $\alpha(G^*) \le k^*$ follows. Conversely, assume $\alpha(G^*) \le k^*$. We have $(2c+3)(\mathrm{pwd}(G)+1) - 1 = \mathrm{pwd}(G^*) \le \alpha(G^*) + c \le k^* + c = (2c+3)(k+1) + c - 1 + c$. Hence, $\mathrm{pwd}(G) \le k + 2c/(2c+3)$. Since $\mathrm{pwd}(G)$ and k are integers, $\mathrm{pwd}(G) \le k$ follows. $\square$

Theorem 1. *Let r be any positive integer.*

1. *The r-sequential clique-width of graphs with n vertices of degree greater than 2 cannot be approximated by a polynomial-time algorithm with an absolute error guarantee of n^ε for any $\varepsilon \in (0,1)$, unless* P = NP.

2. cwd_r-MINIMIZATION *is* NP*-complete.*

Proof. For a graph G we define the parameter $\alpha(G) = \mathrm{cwd}_r(G')$. That is, $\alpha(G)$ is the r-sequential clique-width of the graph G' obtained from G by means of Construction 1. We take the constant $c = r + 4$. By Lemmas 2 and 4, $|\alpha(G) - \mathrm{pwd}(G)| \le c$ follows; thus the assumption of Lemma 5 is met.

For showing the first part of the theorem, assume that for a constant $\varepsilon \in (0,1)$ there exists a polynomial-time algorithm $\mathcal{A}$ that outputs for a given graph G with n vertices of degree at least 3 an integer $\mathcal{A}(G)$ with $|A(G) - \mathrm{cwd}_r(G)| \le n^\varepsilon$. For a graph G with n vertices and minimum degree at least 3, G' has exactly n vertices of degree at least 3; applying $\mathcal{A}$ to G' gives now $|A(G') - \mathrm{cwd}_r(G')| = |A(G') - \alpha(G)| \le n^\varepsilon$. Hence, by the first part of Lemma 5 such algorithm $\mathcal{A}$ cannot exist unless P = NP.

By the second part of Lemma 5, α-MINIMIZATION is NP-hard. We reduce α-MINIMIZATION to cwd_r-MINIMIZATION by reducing an instance (G,k) of the former problem to the instance (G',k) of the latter problem; obviously $\alpha(G) \le k$ if and only if $\mathrm{cwd}_r(G') \le k$. Thus cwd_r-MINIMIZATION is NP-hard as well. We can guess an r-sequential k-expression for a given graph G and verify in polynomial time whether it is indeed an r-sequential k-expression for G, hence cwd_r-MINIMIZATION is in NP. Thus cwd_r-MINIMIZATION is NP-complete. $\square$

5 Clique-width versus r-sequential clique-width

In this final section we show that clique-width and r-sequential clique-width can differ significantly.

Theorem 2. *For every $r \ge 1$ there exist graphs of constant clique-width but arbitrarily high r-sequential clique-width.*

This result follows from the next two lemmas. In the remainder of this section, H denotes a ternary rooted tree. That is, every non-leaf of H has exactly three children. H' denotes the graph obtained from H by Construction 1.

We call a k-expression X to be 1-*terminal* if it does not contain the operations $\eta_{1,i}$, $\eta_{i,1}$, or $\rho_{1\to i}$. That is, the clique-width construction described by X has the property that after a vertex v has once received the label 1, no edges incident to v are inserted any more, and the label of v is not changed anymore.

Lemma 6. *H' has a* 1*-terminal* 4*-construction, thus* $\mathrm{cwd}(H') \le 4$.

Proof. We proceed by induction on the number n of vertices of H. The lemma holds by trivial reasons if $n = 1$. For $n = 4$, let u_1, u_2, u_3 denote the leaves and v the root of H. The edge $u_i v$ is replaced in H' by the bridges (u_i, x_i^j, y_i^j, v), $j = 1, 2, 3$. We put

$$\begin{aligned}
F_i^j &= \eta_{3,4}(4(x_i^j) \oplus 3(y_i^j)), \quad (1 \leq i, j \leq 3);\\
F^i &= \rho_{4\to 1}(\rho_{2\to 1}(\eta_{2,4}(F_1^i \oplus F_2^i \oplus F_3^i \oplus 2(u_i))));\\
F &= \rho_{3\to 1}(\eta_{2,3}(F^1 \oplus F^2 \oplus F^3 \oplus 2(v)));\\
H' &= \rho_{2\to 1}(F).
\end{aligned}$$

The corresponding 4-expression is evidently 1-terminal, and we have $\mathrm{cwd}(H') \leq 4$ for $n = 4$. Now assume $n > 4$. We can choose a vertex $v \in V(H)$ that is adjacent to three leaves u_1, u_2, u_3. We put $H_0 = H - (u_1, u_2, u_3)$. By induction hypothesis, H_0' has a 1-terminal 4-expression X_0. The vertex v is introduced in X_0 as initial 4-graph $i(v)$ with $2 \leq i \leq 4$; we assume, w.l.o.g., that $i = 2$. We obtain a 4-expression X from X_0 by replacing $2(v)$ with the 4-expression defining the 4-graph F as defined above. Since all vertices of F except v have the terminal label 1, and since X_0 is 1-terminal, we conclude that X is a 1-terminal 4-expression defining H'. □

Lemma 7. *If H is the complete ternary tree of height h then $\mathrm{cwd}_r(H') \geq h - r$ for any $r \geq 1$.*

Proof. From results of Schaeffler [17] and Ellis et al. [8] it follows that the pathwidth of H is h. Lemmas 4 and 1 yield $\mathrm{pwd}(H) \leq \mathrm{cwd}_1(H') \leq \mathrm{cwd}_r(H') + r$. □

6 Final remarks

In this paper we have established the first step for proving that computing the clique-width of a graph is NP-hard. For the second step [9] we consider the following simple construction: from a given graph G we obtain a graph G'' by replacing every edge of G by an induced path of length two. We show that

1. clique-width and sequential clique-width of G'' differ at most by a small constant if G is cobipartite, and

2. sequential clique-width of G' and sequential clique-width of G'' differ at most by a small constant.

This, together with Theorem 1, shows that, cwd-MINIMIZATION is NP-complete, and that the clique-width of a graph cannot be absolutely approximated in polynomial time unless P = NP.

Acknowledgement

We thank Derek Corneil for introducing us to the problem and for all the inspiring discussions. We also thank Frank Gurski for suggesting us the use of Construction 2 for the NP-completeness proof.

References

[1] S. Arnborg, D. G. Corneil, and A. Proskurowski. Complexity of finding embeddings in a k-tree. *SIAM J. Algebraic Discrete Methods*, 8(2):277–284, 1987.

[2] H. L. Bodlaender, J. R. Gilbert, H. Hafsteinsson, and T. Kloks. Approximating treewidth, pathwidth, frontsize, and shortest elimination tree. *J. Algorithms*, 18(2):238–255, 1995.

[3] D. G. Corneil, M. Habib, J.-M. Lanlignel, B. A. Reed, and U. Rotics. Polynomial time recognition of clique-width ≤ 3 graphs (extended abstract). In G. H. Gonnet, D. Panario, and A. Viola, editors, *Theoretical Informatics, 4th Latin American Symposium (LATIN 2000)*, volume 1776 of *Lecture Notes in Computer Science*, pages 126–134, 2000.

[4] D. G. Corneil and U. Rotics. On the relationship between clique-width and treewidth. *SIAM J. Comput.* To appear. An extended abstract appeared in the Proceedings of Graph-theoretic Concepts in Computer Science (WG 2001), LNCS 2204, pp. 78-90, Springer, 2001.

[5] B. Courcelle, J. Engelfriet, and G. Rozenberg. Context-free handle-rewriting hypergraph grammars. In H. Ehrig, H.-J. Kreowski, and G. Rozenberg, editors, *Graph-Grammars and their Application to Computer Science, 4th International Workshop, Bremen, Germany, March 5–9, 1990, Proceedings*, volume 532 of *Lecture Notes in Computer Science*, pages 253–268, 1991.

[6] B. Courcelle, J. A. Makowsky, and U. Rotics. Linear time solvable optimization problems on graphs of bounded clique-width. *Theory of Computing Systems*, 33:125–150, 2000.

[7] B. Courcelle and S. Olariu. Upper bounds to the clique width of graphs. *Discr. Appl. Math.*, 101(1-3):77–114, 2000.

[8] J. A. Ellis, I. H. Sudborough, and J. S. Turner. The vertex separation and search number of a graph. *Information and Computation*, 113(1):50–79, 1994.

[9] M. R. Fellows, F. A. Rosamond, U. Rotics, and S. Szeider. Proving NP-hardness for clique-width II: non-approximability of clique-width. Electronic Colloquium on Computational Complexity, Technical report TR05-080, **Revision 01**, 2005.

[10] F. V. Fomin, P. Heggernes, and J. A. Telle. Graph searching, elimination trees, and a generalization of bandwidth. In *Fundamentals of computation theory*, volume 2751 of *Lecture Notes in Computer Science*, pages 73–85. Springer Verlag, 2003.

[11] F. Gurski and E. Wanke. Minimizing NLC-width is NP-complete. Extended abstract accepted for WG 2005.

[12] F. Gurski and E. Wanke. On the relationship between NLC-width and linear NLC-with. Theoretical Computer Science, to appear.

[13] Ö. Johansson. Clique-decomposition, NLC-decomposition, and modular decomposition—relationships and results for random graphs. In *Proceedings of the Twenty-ninth Southeastern International Conference on Combinatorics, Graph Theory and Computing (Boca Raton, FL, 1998)*, volume 132 of *Congr. Numer.*, pages 39–60, 1998.

[14] M. Karpinski and J. Wirtgen. On approximation hardness of the bandwidth problem. Technical Report TR97-041, ECCC, Electronic Colloquium on Computational Complexity, 1997.

[15] N. G. Kinnersley. The vertex separation number of a graph equals its path-width. *Information Processing Letters*, 42(6):345–350, 1992.

[16] S. Oum and P. Seymour. Approximating clique-width and branch-width. Submitted, Oct. 2004.

[17] P. Scheffler. A linear algorithm for the pathwidth of trees. In *Topics in combinatorics and graph theory (Oberwolfach, 1990)*, pages 613–620. Physica, 1990.

[18] E. Wanke. k-NLC graphs and polynomial algorithms. *Discr. Appl. Math.*, 54(2-3):251–266, 1994. Efficient algorithms and partial k-trees.

ECCC ISSN 1433-8092
http://www.eccc.uni-trier.de/eccc
ftp://ftp.eccc.uni-trier.de/pub/eccc
ftpmail@ftp.eccc.uni-trier.de, subject 'help eccc'