Cliquewidth is NP-hard

Fellows, Rosamond, Rotics, and Szeider

May 2, 2005

Abstract

PENDING

1 Introduction

PENDING

2 Notation and preliminaries

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 *cliquewidth* $\mathrm{cwd}(G)$ 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 *cliquewidth construction*. A cliquewidth construction can be represented by an algebraic term composed of $\oplus$, $\rho_{i\to j}$, and $\eta_{i,j}$, ($i,j \in \{1,\ldots,k\}$, 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 cliquewidth 2.

3 Sequential cliquewidth

In the sequel we consider cliquewidth constructions where disjoint union of two k-graphs is only allowed if one of them has r or fewer vertices. We call such cliquewidth constructions and the corresponding k-expressions *r-sequential* (or *sequential* for $r = 1$). The *r-sequential cliquewidth* 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. 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

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

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 cliquewidth instead of sequential cliquewidth 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 $r \geq 1$.*

For the remainder of this section let G denote a fixed simple connected graph with $n \geq 2$ 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; we call such paths *bridges*.

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,* $\mathrm{cwd}_1(G') \leq \mathrm{vsn}(G) + 4$.

The next lemma will allow us to bound the vertex separation number of G in terms of the sequential cliquewidth 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.*

Theorem 1. *From a sequential k-expression defining G' we can construct in polynomial time a layout for G with out-degree at most k. Consequently,* $\mathrm{vsn}(G) = \mathrm{pwd}(G) \leq \mathrm{cwd}_1(G')$.

In the proof of the next theorem we shall use a result of Bodlaender, Gilbert, Hafsteinsson, and Kloks [BodlaenderGilbertHafsteinssonKloks95], which states that the pathwidth (i.e., the vertex separation number) of a graph G cannot be approximated by a polynomial-time algorithm with an absolute error guarantee of $|V(G)|^\varepsilon$ for any $\varepsilon \in (0, 1)$, unless P = NP.

Corollary 1. *The r-sequential cliquewidth of a graph G cannot be absolutely approximated by a polynomial-time algorithm, unless* $\mathrm{P} = \mathrm{NP}$.

The non-approximability result for pathwidth (Bodlaender, Gilbert, Hafsteinsson, and Kloks [BodlaenderGilbertHafsteinssonKloks95]) that we have used above, holds even for cobipartite graphs (graphs that are complements of bipartite graphs), as shown by Karpinski and Wirtgen [KarpinskiWirtgen97]. We will use this fact below.

4 Cwd-expressions for G''

Let G be a co-bipartite graph. Then the vertices of G can be partitioned into two cliques A and B. In the sequel we assume that G is connected (i.e., there is at least one edge between the two cliques. The case when G is not connected, is easy and will be mentioned at the end.

Let G'' be the graph obtained from G by replacing each edge uv of G with a path of length two. That is, the edge uv is replaced with new vertex denoted by $s_{u,v}$ which is adjacent just to u and v (and the edge uv is omitted). We call the vertices of G'' which are also vertices of G *regular vertices*. We call the vertices of G'' which are not vertices of G *special vertices*. The regular vertices which belong to A, B are called A-regular vertices, B-regular vertices, respectively. The special vertices of the form $s_{u,v}$ such that both u and v belong to A, B are called A-*special* vertices, B-*special* vertices, respectively. The special vertices of the form $s_{u,v}$ such that one of the vertices (say u) belongs to A and the other vertex (say v) belongs to B are called AB-*special* vertices.

We shall prove that:

Theorem 2. $\mathrm{cwd}_1(G'') \leq \mathrm{cwd}(G'') + 11$.

For the proof of Theorem 2 we shall use the following definitions and lemmas.

4.1 Definitions

Recall that for a cwd-expression t, we denote by $\mathsf{val}(t)$ the labeled graph defined by t. We denote a cwd-expression which uses k labels as a k-cwd-expression. Often when it is clear from the context we shall use the term expression instead of cwd-expression. We denote by $\mathsf{labels}(t)$ the number of labels used in $\mathsf{val}(t)$.

We say that an $\eta_{i,j}$-operation is *redundant* if it introduce no edges; i.e, if there are no vertices with label i, (or j) when the $\eta_{i,j}$ operation is established. We say that a $\rho_{i\to j}$-operation is *redundant* if there are no vertices with label i when the $\rho_{i\to j}$-operation is introduced.

We denote by $\mathsf{tree}(t)$ the cwd-expression tree corresponding to the cwd-expression t. For a node a of $\mathsf{tree}(t)$, we denote by $\mathsf{tree}(t)\langle a\rangle$ the subtree of $\mathsf{tree}(t)$ rooted at a. We denote by $t\langle a\rangle$ the cwd-expression corresponding to $\mathsf{tree}(t\langle a\rangle)$. Note that in $t\langle a\rangle$ (and similarly in $\mathsf{tree}(t\langle a\rangle)$) we assume that the operation a is already established. When a is an $\oplus$-operation we denote by $\mathsf{left}(a)$ and $\mathsf{right}(a)$ the left and right children of a in $\mathsf{tree}(t)$, respectively.

For a vertex x occurring in $t\langle a\rangle$, We say that x is *dead at* a (or *dead at* $\mathsf{val}(t\langle a\rangle)$) if all the edges incident with x in $\mathsf{val}(t)$ are included in $\mathsf{val}(t\langle a\rangle)$. Otherwise we say that x is *active at* a (or *active at* $\mathsf{val}(t\langle a\rangle)$).

We call any operation in $\mathsf{tree}(t)$ immediately below an $\oplus$-operation a *top operation*.

We say that a cwd-expression t is in *normal form* if its cwd-expression tree satisfies the following conditions:

1. There is no ρ-operation immediately above an $\oplus$-operation.
2. There is no η-operation immediately above a ρ-operation.
3. Each $\eta_{i,j}$-operation adds at least one edge of the form xy such that x and y are on the left and right hand sides of the first $\oplus$-operation below it.
4. Each $\eta_{i,j}$-operation does not add new edges of the form xy such that both x and y belong to the same side of the first $\oplus$-operation below it.
5. There are no redundant ρ-operations.
6. For each top operation a, all the dead vertices at a have label 1 at a.

Recall that we consider a non-labeled graph G as a labeled graph with all vertices labeled by the same label 1.

Note that the cwd-expression tree corresponding to a cwd-expression in normal form has the property that between any two consecutive $\oplus$-operations o_1 and o_2 (such that o_1 is below o_2 in the tree) there is a sequence of η-operations followed by a sequence of ρ-operations. In other words if we move up in the tree from o_1 to o_2, we meet a sequence of η-operations followed by a sequence of ρ-operations. If the sequence of η-operations is empty, then the sequence of ρ-operations should be empty, too.

Lemma 4. *Let t be a k-*cwd-*expression which defines a non-labeled graph G. Then there exists a k-*cwd-*expression t' in normal form which defines G.*

Proof. Starting from t repeat the following operations:

(1) If there is a ρ-operation above an $\oplus$-operation, then push it down below the $\oplus$-operation. That is, replace $\rho_{i\to j}(t_1 \oplus t_2)$ with $\rho_{i\to j}(t_1) \oplus \rho_{i\to j}(t_2)$.

(2) If there is an η-operation above a ρ-operation then push it down below the ρ-operation (possibly replacing it with two η-operations). For example, replace $\eta_{i,j}\rho_{\ell\to i}(t_1)$ with $\rho_{\ell\to i}\eta_{i,j}\eta_{\ell,j}(t_1)$.

(3) If there is an η-operation which does not add edges between the two sides of the first $\oplus$ below it, push it down below the $\oplus$-operation.

(4) If there is an η-operation which adds new edges of the form xy such that both x and y belong to the left (respectively, right) side of the first $\oplus$-operation below it, then add an η-operation to the left (respectively, to the right) of the $\oplus$-operation.

(5) Remove redundant η and ρ-operations.

(6) Suppose there is a top operation a such that some vertex (say x) is dead at a but does not have label 1 at a. Let h be the label of x at a. If there is no vertex labeled with 1 at a, or if there is an active vertex at a labeled with 1 at a, exchange the labels 1 and h everywhere in $t\langle a\rangle$. If there is already a dead vertex labeled with 1 at a, replace $t\langle a\rangle$ with $\rho_{h\to 1}(t\langle a\rangle)$.

When it will not be possible to do any of the above operations, the obtained cwd-expression defines G and is in normal form. This completes the proof of the lemma. ☐

For a k-cwd-expression t we denote by $\mathsf{norm}(t)$ an arbitrarily selected k-cwd-expression t in normal form which can be obtained from t by applying the transformations indicated the proof of Lemma 4. Note that the set of $\oplus$-operations in t and in $\mathsf{norm}(t)$ is the same. Note that when the norm-operation is applied to a sub-expression of t (say, to $t\langle a\rangle$ for some node a in $\mathsf{tree}(t)$), step (6) is applied only for vertices which are dead at a. In other words, vertices such that not all of their incident edges in $\mathsf{val}(t)$ exist in $\mathsf{val}(t\langle a\rangle)$ cannot be dead in $\mathsf{norm}(t\langle a\rangle)$.

Let a be an $\oplus$-operation of a cwd-expression t. Let b and c be the left and right children of a, respectively. We say that vertex x occurs on the *left (right) side* of a if it occurs at $t\langle b\rangle$ ($t\langle c\rangle$). We say that a is a 1-$\oplus$-operation if at least one side of a contains a single vertex. We say that a is a 2-$\oplus$-operation if it is not a 1-$\oplus$-operation. We say that t is a cwd$_1$-expression if all $\oplus$-operations in t are 1-$\oplus$-operations. We say that t is a k-cwd$_1$-expression if t is a cwd$_1$-expression which uses k labels.

Let t be a cwd-expression defining G''. Let a be a $\oplus$-operation of t. We say that there is a *separation* at a between the A-regular vertices and the B-regular vertices if all A-regular vertices of $\mathsf{val}(t\langle a\rangle)$ occur on one side of a (say, on the left side of a) and all the B-regular vertices of $\mathsf{val}(t\langle a\rangle)$ occur on the other side of a (say, on the right side of a).

Let $s_{x,y}$ be any special vertex and let a be a node of t. The label of $s_{x,y}$ at a is denoted as $\mathsf{conn}(x,y)$ if $\mathsf{val}(t\langle a\rangle)$ does not include the edges connecting $s_{x,y}$ to x and y. The label of $s_{x,y}$ at a is denoted $\mathsf{conn}(x)$ ($\mathsf{conn}(y)$) if $\mathsf{val}(t\langle a\rangle)$ includes the edge connecting $s_{x,y}$ to y (x) but does not include the edge connecting $s_{x,y}$ to x (y).

Proposition 1. *Let t be an expression which defines G'' and let a be any node in tree(t). Suppose that labels $\mathsf{conn}(y_1)$ and $\mathsf{conn}(y_2)$, for $y_1 \neq y_2$, are used in $\mathsf{val}(t\langle a\rangle)$. Then the labels $\mathsf{conn}(y_1)$ and $\mathsf{conn}(y_2)$ are different labels at $\mathsf{val}(t\langle a\rangle)$.*

Proof. Let vertices s_1 and s_2 having labels $\mathsf{conn}(y_1)$ and $\mathsf{conn}(y_2)$ at $\mathsf{val}(t\langle a\rangle)$, respectively. By definition, s_1 must be a special vertex of the form s_{x_1,y_1} where x_1 occurs in $\mathsf{val}(t\langle a\rangle)$ and s_1 is connected to x_1 at $\mathsf{val}(t\langle a\rangle)$ and y_1 is not in $\mathsf{val}(t\langle a\rangle)$. Similarly, s_2 must be a special vertex of the form s_{x_2,y_2} where x_2 occurs in $\mathsf{val}(t\langle a\rangle)$, s_2 is connected to x_2 at $\mathsf{val}(t\langle a\rangle)$, and y_2 is not in $\mathsf{val}(t\langle a\rangle)$. Since $y_1 \neq y_2$ and $x_1 \neq y_1$, it follows that $s_1 \neq s_2$. By definition, there is an η-operation above $t\langle a\rangle$ which connects s_1 to y_1. If the labels $\mathsf{conn}(y_1)$ and $\mathsf{conn}(y_2)$ are the same label at $\mathsf{val}(t\langle a\rangle)$, this η-operation will connect s_1 also to y_2, a contradiction. ☐

Proposition 2. *Let t be a* cwd*-expression defining G''. For each $\oplus$-operation a of t there is at most one pair of A-regular (B-regular) vertices which occur on different sides of a and have the same label at a.*

Proof. Suppose there are two different pairs of A-regular vertices (x_1, y_1) and (x_2, y_2) such that for $i = 1, 2$, x_i and y_i occur at different sides of a and have the same label at a. Assume without loss of generality that x_1 and x_2 occur on the left side of a and y_1 and y_2 occur on the right side of a. Clearly, either

$x_1 \neq x_2$ or $y_1 \neq y_2$. Assume without loss of generality that $x_1 \neq x_2$. Consider the special vertex s_{y_1,x_2}. If s_{y_1,x_2} is not in $t\langle a\rangle$, then when later on the edge connecting s_{y_1,x_2} to y_1 will be establish, also the edge connecting it to x_1 will be established, a contradiction. Thus s_{y_1,x_2} is in $t\langle a\rangle$. If s_{y_1,x_2} occurs on the left side of a then when the edge connecting it to y_1 will be established, it will be connected also to x_1, a contradiction. If s_{y_1,x_2} is on the right side of a, then when the edge connecting it to x_2 will be established, it will be connected also to y_2, a contradiction. The argument for two different pairs of B-regular vertices is symmetric. □

Proposition 3. *Let t be a* cwd-*expression defining G''. Let a be an $\oplus$-operation of t and let (x_1, y_1) be a pair of A-regular (B-regular) vertices which occur on different sides of a and have the same label at a. Then both x_1 and y_1 are active at a and for every other vertex (say z) occurring at a the label of z is different from the label of x_1 and y_1 at a.*

Proof. Since x_1 and y_1 have the same label at a, either they are both dead at a or they are both active at a. Suppose x_1 and y_1 are dead at a. Consider s_{x_1,y_1}. If s_{x_1,y_1} is not in $t\langle a\rangle$, then it is not possible to connect it to x_1 and y_1 (as they are dead at a), a contradiction. Assume without loss of generality that x_1 and s_{x_1,y_1} are on the same side of a. Since y_1 is on the other side of a, and y_1 is dead at a, it is not possible to connect s_{x_1,y_1} to y_1, a contradiction. We have shown that both x_1 and y_1 are active at a. If there is another vertex z with the same label as x_1 and y_1 at a, then, when the edges connecting some vertex of G'' (say, w) to x_1 and y_1 will be established (such edges must be established since x_1 and y_1 are active at a), also the edge connecting it to z will be established, a contradiction (no vertex of G'' is adjacent to x_1, y_1 and z). □

Proposition 4. *Let t be a* cwd-*expression defining G''. Let a be an $\oplus$-operation of t and let (x_1, y_1) be a pair of regular vertices which occur on different sides of a and have the same label at a. Then all the edges connecting x_1 (y_1) to its neighbors in $G'' - s_{x_1,y_1}$ exist in* $\mathsf{val}(t\langle a\rangle)$.

Proof. Let s be a vertex which is adjacent to x_1 in $G'' - \{s_{x_1,y_1}\}$. Clearly s must be a special vertex of the form $s_{x_1,z}$ for $z \neq y_1$. If s is not connected to x_1 in $\mathsf{val}(t\langle a\rangle)$, then it is not possible to connect s to x_1 without connecting it also to y_1, a contradiction. □

4.2 Property 1

Let t be a cwd-expression defining G''. We say that a pair of A-regular (B-regular) vertices (x_1, y_1) is a *bad pair* for t if x_1 and y_1 have the same label in $\mathsf{val}(t\langle a\rangle)$ where a is the unique $\oplus$-operation of t such that x_1 occurs at the left and y_1 occurs at the right side of a. In this case, we also say that a is the *bad $\oplus$-operation* for the pair (x_1, y_1).

We say that t has *Prperty 1* if the following conditions hold for t:

Condition 1.1: t is in normal form.

Condition 1.2: There are no bad pairs for t.

Lemma 5. *Let t be a k-*cwd-*expression defining G''. Then there exists a $(k+3)$-*cwd-*expression t' defining G'' such that t' has Property 1.*

Proof. Let t be a k-cwd-expression defining G''. Let $t_1 = \mathsf{norm}(t)$. If Condition 1.2 also holds for t_1, then we are done. Let (x_1, y_1) be a bad pair for t_1 and let a be the bad $\oplus$-operation for the pair (x_1, y_1). Let t_2' denote the expression obtained after removing s_{x_1,y_1} from t_1. Clearly, the $\oplus$-operation a still exists at t_2' and x_1, y_1 occur on different sides of a and have the same label (say ℓ) at $\mathsf{val}(t_2'\langle a\rangle)$.

Let b and c denote the left and right children of a in $\mathsf{tree}(t_3')$, respectively. Let t_2'' denote the expression obtained from t_2' by replacing $t_2'\langle b\rangle$ with $\rho_{l\to k+1}(t_2'\langle b\rangle)$. Thus in $t_2''\langle a\rangle$, vertex x_1 has label $k+1$ and vertex y_1 has label ℓ. Let t_2 denote the expression obtained from t_2'' by replacing $t_2''\langle a\rangle$ with

$$\rho_{l\to 1}(\rho_{k+1\to 1}(\rho_{k+3\to 1}(\eta_{k+1,k+3}(\eta_{k+3,l}(k+3(s_{x_1,y_1}) \oplus t_2''\langle a\rangle))))).$$

In words, the last expression is obtained by adding s_{x_1,y_1} to $t_2''\langle a\rangle$ with label $k+3$ connecting it to x_1 and y_1, and then renaming its label to 1 and renaming the labels $k+1$ and ℓ to 1. By Proposition 4, all edges incident with x_1 or y_1 in $G'' - s_{x_1,y_1}$ already exist at $\mathsf{val}(t_2''\langle a\rangle)$. Thus, after adding s_{x_1,y_1} and connecting it to x_1 and y_1, all the edges incident with x_1 and y_1 in G'' exists in $\mathsf{val}(t_2)$. It follows that t_2 defines G''. Clearly, t_2 has less bad pairs than t_1 (as (x_1, y_1) is not a bad pair in t_2). We can now repeat the above process for each bad pair of t_2 until we get an expression t_3 which defines G'' and does not have any bad pairs. Note that in the construction of t_3 we will use label $k+2$ instead of label $k+1$ in the above formula, if label $k+1$ is already used at the bad $\oplus$-operation a that we consider. By Proposition 2, each bad $\oplus$-operation a can be bad for at most two bad pairs. Hence after using labels $k+1$ and $k+2$ at a, we will not consider a later in the process. Finally, we set $t' = \mathsf{norm}(t_3)$ and t' satisfies the conditions of the lemma. □

4.3 Property 2

Let t be a k-cwd-expression defining G'' which has Property 1. We say that t has *Property 2*, if the following condition holds for each $\oplus$-operation a in t.

Condition 2.1: If there are two regular vertices (say, x and y) which are adjacent in G'', occur on different sides of a, and have unique labels at a, then the special vertex $s_{x,y}$ is introduced by a 1-$\oplus$-operation above a in $\mathsf{tree}(t)$.

Lemma 6. *Let t be a k-*cwd-*expression defining G'' such that t has Property 1. Then there exists a $(k+1)$-*cwd-*expression t' defining G'' such that t' has Property 2.*

Proof. Let a be an $\oplus$-operation in t such two regular vertices x and y which are adjacent in G'' occur on different sides of a. Let t_1 denote the expression obtained by removing $s_{x,y}$ from t and by adding it with a 1-$\oplus$-operation immediately above a with label $k+1$, connecting it to x and y and renaming its label to 1. Repeating this process for each two adjacent regular vertices x and y occurring on different sides of a we get an expression t_2 such that condition 2.1 holds for a. Repeating the process for each $\oplus$-operation in t, we get an expression t_3 such that condition 2.1 holds for all $\oplus$-operations in t_3. Finally, $t' = \mathsf{norm}(t_3)$, is a $(k+2)$-cwd-expression which defines G'' and has Property 2. □

4.4 Property 3

Let t be a cwd-expression defining G''. We say that an $\oplus$-operation a is *2-bad* in t if a is a 2-$\oplus$-operation and there exists a regular vertex having a unique label at a. We say that a regular vertex (say, x) is *bad in t* if there exists a 2-bad $\oplus$-operation a in t such that x has a unique label at a.

Let t be a k-cwd-expression defining G''. We say that t has *Property 3* if t has Property 2 and there are *no bad regular vertices in t*.

We will show below (Lemma 8) that if t is a k-cwd-expression defining G'' with Property 2, then there exists a $(k+2)$-cwd-expression t' defining G'' with Property 3. For showing this we need following definitions and lemma.

Let t be a $(k+2)$-cwd-expression defining the graph G''. We say that t has the *Extra Labels Property* if the following conditions hold:

1. For each 2-$\oplus$-operation a in t, labels $k+1$ and $k+2$ are not used in $\mathsf{val}(t\langle a\rangle)$. That is, there are no vertices with labels $k+1$ and $k+2$ in $\mathsf{val}(t\langle a\rangle)$.

2. There are no ρ-operations of the form $\rho_{l\to k+1}$ or $\rho_{l\to k+2}$ in t.

Lemma 7. *Let t be a* $(k+2)$-cwd-*expression defining* G'' *such that t has Property 2 and the Extra Labels Property, and there are bad regular vertices in t. Then there exists a* $(k+2)$-cwd-*expression* t' *defining* G'' *such that* t' *has Property 2 and the Extra Labels Property, and the number of bad regular vertices in* t' *is less than the number of bad regular vertices in t.*

Proof. Let x be a bad regular vertex in t. Let a be the highest 2-bad $\oplus$-operation in t such that x has a unique label at a. Note that we can assume that x keeps its initial label (say, j) in all nodes in $tree(t)$ occurring on the path from the node which creates x up to node a. Let S denote the set of all special vertices which are adjacent to x and occur in $t\langle a\rangle$. Let S_1 denote the set of all vertices in S of the form $s_{x,y}$ which are adjacent to x and not to y in $\mathsf{val}(t\langle a\rangle)$. Let S_2 denote the set of all vertices in S of the form $s_{x,y}$ which are adjacent to y and not to x in $\mathsf{val}(t\langle a\rangle)$. Let $S_3=\{s1,\ldots,s_m\}$ denote the set of all vertices in S of the form $s_{x,y}$ which are adjacent to x and y in $\mathsf{val}(t\langle a\rangle)$. Finally, let S_4 denote the set of all vertices in S of the form $s_{x,y}$ which are not adjacent to y and x in $\mathsf{val}(t\langle a\rangle)$.

Claim 1. $S_4=\emptyset$.

Proof of Claim 1. Let $s=s_{x,y}$ be a vertex in S_4. Suppose x and y occur on different sides of a. Since s is not adjacent to y in $\mathsf{val}(t\langle a\rangle)$, it follows that y must have a unique label at a. By definition, x has unique label at a. Since condition 2.1 holds for a, it follows that s must be introduced by a 1-$\oplus$-operation above a in $\mathsf{tree}(t)$, a contradiction (since s belong to $t\langle a\rangle$). Thus x and y occur on the same side (say, left side) of a. If s also occurs on the left side of a, then by the normality of t, s must be adjacent to both x and y in $\mathsf{val}(t\langle a\rangle)$, a contradiction. Thus x occurs on the right side of a. Let o denote the lowest $\oplus$-operation in $\mathsf{tree}(t)$ which contains both x and y. Clearly, o must be on the left side of a, and x and y occur on different sides of o. Since s is not adjacent to y in $\mathsf{val}(t\langle o\rangle)$, y must have a unique label at $\mathsf{val}(t\langle o\rangle)$. Since condition 2.1 holds for o, it follows that s must be introduced by a 1-$\oplus$-operation above o in $\mathsf{tree}(t)$. But since s

occurs on the right side of a, the 1-$\oplus$-operation which introduces s is not above o in $\mathsf{tree}(t)$, a contradiction. We conclude that S_4 must be empty. $\square$

By Claim 1, $S = S_1 \cup S_2 \cup S_3$. For $1 \leq i \leq m$, let $s_i = s_{x,y_i}$, let w_i denote the lowest η-operation in $\mathsf{tree}(t\langle a \rangle)$ which connects s_i to y_i, and let o_i denote the first $\oplus$-operation below w_i in $\mathsf{tree}(t\langle a \rangle)$.

Claim 2. *For $1 \leq ileqm$, if x occurs at $t\langle w_i \rangle$, then o_i is a 1-$\oplus$-operation such that one side of o_i is of the form $\ell(s_i)$ and the first two nodes above o_i are η-operations which connect s_i to x and y_i.*

Proof of Claim 2. Let o'_i be the lowest $\oplus$-operation which contains both x and y_i. Clearly o'_i occurs below w_i. Since y_i must have a unique label at $\mathsf{val}(t\langle w_i \rangle)$, it follows that that y_i and x have unique labels at $\mathsf{val}(t\langle o'_i \rangle)$ and occur on different sides of a. By condition 2.1, s_i is introduced by a 1-$\oplus$-operation (say, o''_i) above o'_i in $\mathsf{tree}(t)$. By the normality of t, the first two operations above o''_i must be η-operations which connect s_i to x and y_i. Hence one of these η-operations must be w_i which implies that o''_i is equal to o_i. This proves the claim. $\square$

We now define a sequence of expressions $e_0, \ldots, e_m$. Let e_0 denote the expression obtained from $t\langle a \rangle$ by omitting any η-operation which connects x to another vertex in $t\langle a \rangle$. Thus, x has no neighbors in $\mathsf{val}(e_0)$. For $1 \leq i \leq m$, e_i is defined according to the following cases.

Case 1: x is in $t\langle w_i \rangle$. By Claim 2, o_i is a 1-$\oplus$-operation of the form $\ell(s_i)$ and the first two operations above o_i are η-operations which connect s_i to x and y_i. Clearly, one of these η-operations is w_i and this operation exists in e_{j-1}, and the other operation does not exists in e_{j-1}. In this case, e_j is obtained from e_{j-1} by adding a $\rho_{\ell \to j}$-operation immediately above w_i, where ℓ is the label that s_i has at w_i. Since label j is already used in $t\langle w_i \rangle$, the number of labels used in e_i is not greater than the number of labels used in $t\langle a \rangle$.

Case 2: x is not in $t\langle w_i \rangle$. Let o'_i be the lowest $\oplus$-operation in $\mathsf{tree}(t)$ which contains both x and s_i. Since s_i is in S_3, o'_i is below a in $\mathsf{tree}(t)$. Thus o'_i belongs to e_{i-1}. Clearly, label j is used at o'_i. In this case e_i is obtained from e_{i-1} by adding a $\rho_{\ell \to j}$-operation immediately below o'_i at the side which contains s_i, where ℓ is the label of s_i at $\mathsf{val}(t\langle o'_i \rangle)$.

Let f_1 denote the expression obtained by omitting x and all the vertices in S_1 from e_m.

Since a is not a 1-$\oplus$-operation, we know that labels $k+1$ and $k+2$ are not used in $\mathsf{val}(t\langle a \rangle)$. By the construction of f_1, these labels are neither used in f_1. Let $f_2 = f_1 \oplus (k+2)(x)$. Let f_3 denote the expression obtained from f_2 by connecting all vertices which have label j at f_2 to x, and then renaming label j to 1. In symbols:

$$f_3 = \rho_{j \to 1}(\eta_{j,k+2}(f_2)).$$

Let $s = s_{x,y}$ be a vertex in S_1 and let ℓ denote the label of s at $\mathsf{val}(t\langle a \rangle)$. Let f_4 be the expression obtained from f_3 by adding s with label $k+1$ connecting it to x and then renaming its label to ℓ. In symbols:

$$f_4 = \rho_{k+1 \to l}(\eta_{k+1,k+2}(((k+1)(s) \oplus f_3))).$$

Let f_5 denote the expression obtained by repeating the above process for each vertex s in S_1. Let f_6 denote the expression obtained from f_4 by renaming the label of x to j. In other words, $f_5 = \rho_{k+2\to j}$. Finally, let t_1 denote the expression obtained from t by replacing $t\langle a\rangle$ with f_5.

Claim 3. *The expression t_1 defines the graph G''.*

Proof of Claim 3. We call an η-operation to be *legal* if it does not introduce edges which do not occur in G''. We first show that all η-operations of t_1 are legal. Consider the new η-operation added in the construction of t_1. The $\eta_{k+2,j}$-operation in f_3 connects x (having label $k+2$) to all the vertices having label j in $\mathsf{val}(f_2)$. By the construction of f_2 all vertices having label j in $\mathsf{val}(f_2)$ are in S_3. Thus this η-operation is legal. The $\eta_{k+1,k+2}$-operation in f_4 connects x to vertex s in S_1. Thus, this η-operation and all the new η-operations in f_5 are legal. The non new η-operations which occur in t_1 also occur in t and are legal since either they are redundant (i.e., introduce no edges) or they connect vertices which were already connected in t.

We show below that all the edges which occur in G'' also exist in $\mathsf{val}(t_1)$. Let uv be an edge in G''. If both u and v do not belong to $S \cup \{x\}$, then the η-operation which connects u and v in t also exists in t_1 and connects u and v in $\mathsf{val}(t_1)$. If both u and v belong to $S \cup \{x\}$, then one of them (say, u) must be equal to x and the other one (say, v) must be in S. If v is in S_1, then v is connected to x in the construction of f_4. If v is in S_2, then the label of v in $\mathsf{val}(f_5)$ is the same as the label of v in $\mathsf{val}(t\langle a\rangle)$, and the label of x in $\mathsf{val}(f_5)$ is the same as the label of x in $\mathsf{val}(t\langle a\rangle)$. Thus, the η-operation above a in $\mathsf{tree}(t)$ which connects v to x also exists in t_1 and connects v to x in $\mathsf{val}(t_1)$. If v is in S_3, then v has label j in $\mathsf{val}(f_2)$, and the $\eta_{j,k+2}$-operation in f_3 connects v to x.

Suppose that one vertex of u and v (say, u) belongs to $S \cup \{x\}$, and the other vertex (say, v) does not belong to $S \cup \{x\}$. If u is equal to x, then v does not occur in $t\langle a\rangle$, which implies that the η-operation above a in $\mathsf{tree}(t)$ which connects x to v in t also connects x to v in $\mathsf{val}(t_1)$. If u is not equal to x, then u is in S and is of the form $s_{x,y}$, and v is equal to y. If u is in S_1, then the label of u at $\mathsf{val}(f_5)$ is the same as the label of u at $t\langle a\rangle$. Thus, the η-operation above a which connects u to y in t also exists in t_1 and connects u to y in $\mathsf{val}(t_1)$. If u is in $S_2 \cup S_3$, then the η-operation which connects u to y in $\mathsf{val}(t\langle a\rangle)$ also exists in e_0 (and in f_5) and connects u to y in $\mathsf{val}(e_0)$ (and in $\mathsf{val}(f_5)$). Thus u and v are adjacent in $\mathsf{val}(t_1)$ too. □

Claim 4. *The number of bad regular vertices in t_1 is less than the number of bad regular vertices in t.*

Proof of Claim 4. Vertex x is introduced in f_5 with a 1-$\oplus$-operation (say, o) and all $\oplus$-operations above o in $\mathsf{tree}(f_5)$ are 1-$\oplus$-operations. Thus x is not a bad regular vertex in f_5. By the selection of a, x does not have a unique label in all bad $\oplus$-operations above a in $\mathsf{tree}(t)$. Thus x does not have a unique label in all bad $\oplus$-operations above the root of f_5 in t_1. We conclude that x is not a bad regular vertex in t_1. Since all regular vertices except x have the same labels and locations in t_1 and in t, it follows that the set of bad regular vertices in t_1 is equal to the set of bad regular vertices in t minus x. □

From the construction of f_5 it is clear that Conditions 1.2, 2.1, and the Extra Label Property hold for f_5. Thus, these conditions and the Extra Label Property hold also for t_1. Let $t' = \mathsf{norm}(t_1)$. Since the norm operation does not violate conditions 1.2, 2.1, and the Extra Labels Property, t' has Property 2 and the Extra Label Property. By Claims 3 and 4, t' defines G'' and has less bad regular vertices than t. This completes the proof of Lemma 7. $\square$

We are now ready to prove the following result.

Lemma 8. *Let t be a k-*cwd*-expression defining G'' such that t has Property 2. Then there exists a $(k+2)$-*cwd*-expression t' defining G'' such that t' has Property 3.*

Proof. Let t be a k-cwd-expression defining G'' such that t has Property 2. Since labels $k+1$ and $k+2$ are not used in t, t is also a $k+2$ expression which defines G'' and has the Extra Labels Property. If t does not have bad regular vertices we are done. Otherwise, replace t with the expression t' given by Lemma 7, and repeat this process until there are no bad regular vertices in t. $\square$

Proposition 5. *Let t be a k-*cwd*-expression defining G'' such that t has Property 3. Let a be a 2-$\oplus$-operation in t such that at least one regular vertex occurs on each side of a. Then there is a separation at a between the A-regular and the B-regular vertices.*

Proof. Let a be a 2-$\oplus$-operation in t and let x and y be two regular vertices occurring on different sides of a. Assume without loss of generality that x occurs on the left side of a and y occur on the right side of a. Suppose x and y are both A-regular vertices. Since Condition 1.2 holds for t, x and y do not have the same label in $\mathsf{val}(t\langle a\rangle)$. Since Property 3 holds for t, x and y do not have unique labels at $\mathsf{val}(t\langle a\rangle)$. Thus, there are two vertices w and z which have the same label as x and y in $\mathsf{val}(t\langle a\rangle)$, respectively. Let $s = s_{x,y}$. If s does not occur in $t\langle a\rangle$, then the η-operation in t which connects s to x, connects it also to w, a contradiction. If s occurs on the left side of a, then the η-operation which connects s to y connects it also to z, a contradiction. If s occurs on the right side of a, then the η-operation which connects s to x connects it also to w, a contradiction. Thus x and y can not be both A-regular vertices.

Similarly, x and y cannot be both B-regular vertices. Thus, one of x and y (say, x) must be A-regular and the other (say, y) must be B-regular. If there is a B-regular vertex (say, z) on the left side then there are two B-regular vertices (z and y) occurring on different sides of a, which is not possible by the above argument. Thus all the A-regular vertices occur on the left side of a and all the B-regular vertices occur on the right side of a. $\square$

4.5 Property 4

Let t be a k-cwd-expression defining G''. We say that t has *Property 4* if it has Property 3 and the following condition holds:

Condition 4: Either there are no 2-$\oplus$-operations in t or there is just one 2-$\oplus$-operation in t (say, a) and there is a separation at a between the A-regular and the B-regular vertices.

Lemma 9. *Let t be a k-*cwd*-expression defining G'' such that t has Property 3. Then there exists a k-*cwd*-expression t' which defines G'' and has Property 4.*

Proof. Let t be a k-cwd-expression which defines G'' and has Property 3. Let a be a 2-$\oplus$-operation in t such that one side of a (say, the left side) contains just special vertices (say, $s_1, \ldots, s_m$). Clearly, $s_1, \ldots, s_m$ are singletons in $\mathsf{val}(t\langle a\rangle)$ and have unique labels in $\mathsf{val}(t\langle a\rangle)$. Let $\ell_1, \ldots, \ell_m$ denote the labels of $s_1, \ldots, s_m$ in $\mathsf{val}(t\langle a\rangle)$, respectively. Let b be the right child of a. Let t_1 be the expression obtained from t by replacing $t\langle a\rangle$ with

$$t\langle b\rangle \oplus \ell_1(s_1) \oplus \cdots \oplus \ell_m(s_m).$$

Let $t_2 = \mathsf{norm}(t_1)$. It is easy to verify that t_2 also defines G'' and has Property 3. Let t' denote the expression obtained from t_2 by repeating the above process for each 2-$\oplus$-operation a in t_2 such that one side of a contains just special vertices. Let a be a 2-$\oplus$-operation in t'.

By the above construction, each side of a contains at least one regular vertex. By Proposition 5, since Property 3 holds for t', there is a separation at a in t' between the A-regular vertices and the B-regular vertices. Suppose there is another 2-$\oplus$-operation in t'. By the above argument each side of a' contains at least one regular vertex and there is a separation at a' in t' between the A-regular and the B-regular vertices. If a is a descendant of a' in $\mathsf{tree}(t')$, then there is no separation at a' between the A-regular and the B-regular vertices, a contradiction. Similarly, a' is not a descendant of a in $\mathsf{tree}(t')$. Let a'' be the lowest node in $\mathsf{tree}(t')$ which contains both a and a'. Clearly a'' must be a 2-$\oplus$-operation. By Proposition 5 there is a separation at a'' in t' between the A-regular and the B-regular vertices. Since a occurs on one side of a'', this side of a'' contains both A-regular and B-regular vertices, a contradiction. We conclude that a is a unique 2-$\oplus$-operation in t'. Thus t' is a k-cwd-expression which defines G'' and has Property 4. □

4.6 Property 5

Let t be a k-cwd-expression defining G''. We say that t has *Property 5* if it has *Property 4* and the following condition holds:

Condition 5: Let a be the unique 2-$\oplus$-operation in t. By Condition 4 there is a separation at a between the A-regular and the B-regular vertices. Then for each A-regular (B-regular) vertex x, which is active at a and occurs on one side (say left side) of a, there is a unique B-regular (A-regular) vertex y which is active at a and occurs on the other side (say right side) of a and has the same label as x in $\mathsf{val}(t\langle a\rangle)$.

Lemma 10. *Let t be a k-*cwd*-expression defining G'' such that t has Property 4. Then there exists a $(k+1)$-*cwd*-expression t' which defines G'' and has Property 5.*

Proof. Let a be the unique 2-$\oplus$-operation in t. Assume without loss of generality that all the A-regular vertices of $t\langle a\rangle$ occur on the left side of a and all the B-regular vertices of $t\langle a\rangle$ occur on the right side of a. Let x be a regular vertex which is active at a. Let ℓ denote the label of x at a. Since Condition 3 holds for t, the label of x at a is not unique. Suppose there are two vertices u and v which have label ℓ at a. Since x is active at a, there is an η-operation above a in $\mathsf{tree}(t)$ which connects some special vertex (say, s) to x. This η-operation connects s also to u and v, a contradiction (since s is adjacent in G to exactly two vertices). Thus, for each regular vertex x which is active at a there is a unique vertex (say y) which is active at a and has label ℓ in $\mathsf{val}(t\langle a\rangle)$. By a similar argument no η-operation above a in $\mathsf{tree}(t)$ connects a vertex other than $s_{x,y}$ to x or to y. Thus, the edges to all the neighbors of x and y in G besides $s_{x,y}$ already exists in $\mathsf{val}(t\langle a\rangle)$.

We now define the cwd-expression t_1 depending on the following cases:

Case 1: Both x and y are not A-regular and both x and y are not B-regular. Since Condition 5 holds in this case for x and y we set $t_1 = t$.

Case 2: Both x and y are A-regular. Let t_1 denote the expression obtained after removing $s_{x,y}$ from t. Let t_1'' denote the expression obtained from t_1 after adding, immediately below the left side of a, vertex $s_{x,y}$ with label $k+1$, connecting it to x and y and renaming its label and the labels of x and y to 1. In other words, let b be the left child of a in t_1, then t_1'' is obtained from t_1 by replacing $t_1\langle b\rangle$ with the expression

$$\rho_{l\to 1}(\rho_{k+1\to 1}(\eta_{k+1,l}(k+1(s_{x,y}) \oplus t_1\langle b\rangle)))).$$

In this case we set $t_1 = \mathsf{norm}(t_1'')$.

Case 3: Both x and y are B-regular. This case is symmetric to Case 2.

Let t' denote the expression obtained by repeating the above process to each regular vertex which is active at a. It is easy to see that t' defines G'' and has Property 5, as required. □

Lemma 11. *Let t be a k-cwd-expression defining G'' such that t has Property 5. Then there is a $(k+4)$-cwd_1-expression which defines G''.*

Proof. If there is no 2-$\oplus$-operation in t, the claim follows immediately. Let a be the unique 2-$\oplus$-operation in t. Let b and c denote the left and right children of a in $\mathsf{tree}(t)$. Assume without loss of generality that all the regular vertices in $t\langle b\rangle$ are A-regular and all regular vertices in $t\langle c\rangle$ are B-regular.

We first introduce the following notation. Let A_1 (B_1) denote the set of A-regular (B-regular) vertices occurring in $t\langle b\rangle$ ($t\langle c\rangle$). Let A_2 (B_2) denote the set of A-regular (B-regular) vertices which do not occur in in $t\langle b\rangle$ ($t\langle c\rangle$). In other words, $A_2 = A \setminus A_1$ and $B_2 = B \setminus B_1$. Let $\mathsf{Active}(A_1)$ ($\mathsf{Active}(B_1)$) denote the set of vertices of A_1 (B_1) which are active at a. Let $\mathsf{Dead}(A_1)$ ($\mathsf{Dead}(B_1)$) denote the set of vertices of A_1 (B_1) which are dead at a. Clearly, $A_1 = \mathsf{Active}(A_1) \cup \mathsf{Dead}(A_1)$ and $B_1 = \mathsf{Active}(B_1) \cup \mathsf{Dead}(B_1)$. By Condition 5, $|\mathsf{Active}(A_1)| = |\mathsf{Active}(B_1)|$. Let $|\mathsf{Dead}(A_1)| = m$ and let $|\mathsf{Dead}(B_1)| = q$. Let x_i, $1 \le i \le m$, be the ith vertex in $\mathsf{Dead}(A_1)$ which gets a non-unique label in $t\langle b\rangle$ (if there is more than one such vertex, choose one of them arbitrarily) and let w_i be the highest node in $\mathsf{tree}(t\langle b\rangle)$ such that x_i has a unique label in $t\langle w_i\rangle$. Let w_i' denote the first node above w_i in $\mathsf{tree}(t\langle b\rangle)$.

Clearly, w_i' must be a ρ-operation which unifies the label that x_i has at $\mathsf{val}(t\langle b\rangle)$ with another label (say, ℓ) where ℓ can be either the dead label (i.e., label 1) or the label of another vertex in $\mathsf{Dead}(A_1)$. Similarly let y_i, $1 \le i \le q$, be the ith vertex in $\mathsf{Dead}(B_1)$ which gets a non-unique label in $t\langle c\rangle$; let z_i be the highest node in $\mathsf{tree}(t\langle c\rangle)$ such that y_i has a unique label in $t\langle z_i\rangle$ and let z_i' denote the first node above z_i in $\mathsf{tree}(t\langle c\rangle)$. Let $X_i = \{x_1, \ldots, x_i\}$, $1 \le i \le m$. Let NX_i, $1 \le i \le m$, denote the set of B-regular vertices which have a neighbor (in G) in the set X_i. Similarly, let $Y_i = \{y_1, \ldots, y_i\}$, $1 \le i \le q$, and let NY_i denote the set of A-regular vertices which have a neighbor (in G) in set Y_i.

Observation 1. *Let v be a vertex which is adjacent to x_i (in G) and does not occur in $t\langle w_i\rangle$. Then the special vertex $s_{x_i,v}$ has the label* $\mathsf{conn}(v)$ *at* $\mathsf{val}(t\langle w_i\rangle)$.

Proof of Observation 1. The vertex $s_{x_i,v}$ must be adjacent to x_i in $\mathsf{val}(t\langle w_i\rangle)$ or else there is no way to connect it to x_i without connecting it to another vertex (other than v) which has the same label as x_i in $\mathsf{val}(t\langle w_i'\rangle)$. Thus $s_{x_i,v}$ has the label $\mathsf{conn}(v)$ at $\mathsf{val}(t\langle w_i\rangle)$. □

Observation 2. *(1) For $1 \le i \le m$, $\mathsf{labels}(t\langle w_i\rangle) \ge |\mathsf{Active}(A_1)| + |A_2| + |NX_i| + m - i = |A| + |NX_i| - i$.*

(2) For $1 \le i \le q$, $\mathsf{labels}(t_z i) \ge |\mathsf{Active}(B_1)| + |B_2| + |NY_i| + q - i = |B| + |NY_i| - i$.

Proof of Observation 2. We prove (1), the proof of (2) is symmetric. Let v be a vertex in $\mathsf{Active}(A_1)$. If v occurs in $t\langle w_i\rangle$, then v has a unique label at $t\langle w_i\rangle$. If v does not occur in $t\langle w_i\rangle$, then by Observation 1 the vertex $s_{x_i,v}$ has the label $\mathsf{conn}(v)$ at $t\langle w_i\rangle$. Thus, so far we have $|\mathsf{Active}(A_1)|$ different labels at $\mathsf{val}(t\langle w_i\rangle)$. Let v be a vertex in $\mathsf{Dead}(A_1) \setminus X_i$. If v occurs in $t\langle w_i\rangle$, then by definition v must have a unique label at $t\langle w_i\rangle$. If v does not occur in $t\langle w_i\rangle$, then by Observation 1 the vertex $s_{x_i,v}$ has the label $\mathsf{conn}(v)$ at $t\langle w_i\rangle$. Thus, we have additional $|\mathsf{Dead}(A_1) \setminus X_i| = m - i$ labels at $\mathsf{val}(t\langle w_i\rangle)$. Let v be a vertex in A_2. By Observation 1, the vertex $s_{x_i,v}$ must have the label $\mathsf{conn}(y)$ at $\mathsf{val}(t\langle w_i\rangle)$. Thus, additional $|A_2|$ labels exists at $\mathsf{val}(t\langle w_i\rangle)$. Let v be a vertex in NX_i. By definition there exists a vertex in X_i (say x_j) such that v is adjacent to x_j in G. By Observation 1, vertex $s_{x_j,v}$ must have the label $\mathsf{conn}(v)$ at $\mathsf{val}(t\langle w_j\rangle)$. Since v is not in $t\langle w_i\rangle$, the vertex $s_{x_j,v}$ also has the label $\mathsf{conn}(v)$ label at $\mathsf{val}(t\langle w_i\rangle)$. Thus, additional $|NX_i|$ labels exists at $\mathsf{val}(t\langle w_i\rangle)$. Summarizing all the labels counted so far gives the formula (1). Note that it is correct to summarize these labels, since by Proposition 1, if $v_1 \neq v_2$, then label $\mathsf{conn}(v_1)$ at $\mathsf{val}(t\langle w_i\rangle)$ is different from the label $\mathsf{conn}(v_2)$ at $\mathsf{val}(t\langle w_i\rangle)$. □

Observation 3. *(1) For $1 \le i \le m$, all the neighbors of x_i in G are included in the set $A \cup NX_i$.*

(2) For $1 \le i \le q$, all the neighbors of y_i in G are included in the set $B \cup NY_i$

Proof of Observation 3. We prove (1), the proof of (2) is symmetric. Let u be a neighbor of x_i in G. If u is not in $A \cup NX_i$, then u must be in B_1. Since x_i has a dead label at a, the special vertex $s_{x_i,u}$ must be in $t\langle b\rangle$. Since $s_{x_i,u}$ is not connected to u at $\mathsf{val}(t\langle a\rangle)$, the vertex u must be active at a. Thus, u must be in $\mathsf{Active}(B_1)$ and has the same label at a as another vertex (say, u') in $\mathsf{Active}(A_1)$. Then the η-operation connecting $s_{x_i,u}$ to u in t connects it also to u', a contradiction. □

Observation 4. $\mathsf{labels}(t\langle a\rangle) \ge |\mathsf{Active}(A_1)| + |A_2| + |B_2|$.

Proof of Observation 4. By definition, each vertex v in $\mathsf{Active}(A_1)$ has a unique label at $\mathsf{val}(t\langle b\rangle)$. Thus there are at least $|\mathsf{Active}(A_1)|$ labels at $\mathsf{val}(t\langle a\rangle)$. Let v be a vertex in A_2 and let u be any vertex in A_1. The special vertex $s_{u,v}$ must have the label $\mathsf{conn}(v)$ at $\mathsf{val}(t\langle a\rangle)$ or else there is no way to connect it to u without connecting it to another vertex (other than u) which has the same label as u at $\mathsf{val}(t\langle a\rangle)$. Thus additional $|A_2|$ labels must exists in $\mathsf{val}(t\langle a\rangle)$. By symmetry, additional $|B_2|$ vertices must exists in $\mathsf{val}(t\langle a\rangle)$.

□

Let e be an expression and let S be a set of special vertices which do not occur in $\mathsf{val}(e)$ such that for each vertex $s = s_{x,y}$ in S both x and y occur in e and have unique labels in $\mathsf{val}(e)$. Then, we denote by $\mathsf{addSpecial}(e, S)$ the expression obtained from e adding each vertex $s = s_{x,y}$ with label $k + 3$, connecting it to x and y, and then renaming its label to 1.

For example: let $e = 2(u_2) \oplus 3(u_3) \oplus 5(u_5)$, $s_1 = s_{u_2,u_3}$, $s_2 = s_{u_3,u_5}$, $S = \{s_1, s_2\}$, and suppose $k = 6$; then $\mathsf{addSpecial}(e, S)$ is the following expression:

$$\rho_{9\to1}(\eta_{9,5}(\eta_{9,3}(9(s_2) \oplus (\rho_{9\to1}(\eta_{9,3}(\eta_{9,2}(9(s_1) \oplus e))))))).$$

We now start the process of constructing a $(k + 4)$-cwd$_1$-expression which defines G''. At each step we show that no more than $k + 4$ labels are used. Moreover, the η-operations added at each step connect special vertices of the form $s_{x,y}$ to x and y, which implies that all edges added in the process belong to G''. Finally, we show in a sequence of observations that for each regular vertex x of G'' the edges which connect x to all its neighbors in G'' exist in the cwd$_1$-expression that we construct. Thus this expression satisfies the conditions of the lemma.

Let e'_1 be the expression obtained by adding all vertices in $A \cup NX_1$ one after the other, each vertex with a unique label. In other words, e'_1 is of the form $\ell_1(v_1) \oplus \ell_2(v_2) \oplus \cdots \oplus \ell_p(v_p)$ where $v_1, \ldots, v_p$ are the vertices in $A \cup NX_1$. Note that by Observation 2, $k \geq p$. Let S_1 denote the set of all special vertices of G'' the form $s_{x,y}$ such that both x and y belong to $A \cup NX_1$. Let $e''_1 = \mathsf{addSpecial}(e'_1, S_1)$. Let ℓ be the label of x_1 in e''_1. Then let $e_1 = \rho_{\ell\to1}(e''_1)$.

Observation 5. $\mathsf{val}(e_1)$ *includes all the edges connecting* x_1 *to all its neighbors in* G''.

Proof of Observation 5. Let s be a neighbor of x_1 in G''. Then s is a special vertex of the form $s_{x_1,u}$ where x_1 and u are adjacent in G. If u is in e_1, then by the construction of e_1, s is connected x_1 (and also to u) in $\mathsf{val}(e_1)$. If u is not in e_1, then u is not in the set $A \cup NX_1$. By Observation 3, u is not a neighbor of x_1, a contradiction. □

For $2 \leq i \leq m$, we define e_i based on e_{i-1} as follows: Let e'_i be the expression obtained by adding all vertices in $NX_i \setminus NX_{i-1}$ to e_{i-1} one after the other, each vertex with a unique label. In other words e'_i is of the form $\ell_1(v_1) \oplus \cdots \oplus \ell_p(v_p) \oplus e_{i-1}$ where $v_1, \ldots, v_p$ are the vertices in $NX_i \setminus NX_{i-1}$. Note that the number of labels used in $\mathsf{val}(e'_i)$ is at most $|NX_i| + |A| + 2 - i$ which is not greater than $k+2$ by Observation 2. Let S_i denote the set of all special vertices of G'' of the form $s_{x,y}$ such that both x and y belong to $A \cup NX_i$ and $s_{x,y}$ is not in e_{i-1}. Let $e''_i = \mathsf{addSpecial}(e'_i, S_i)$. Let ℓ be the label of x_i in e''_i. Then let $e_i = \rho_{\ell\to1}(e''_i)$.

Observation 6. $\mathsf{val}(e_i)$ *includes all the edges connecting* x_i *to all its neighbors in* G''.

Proof of Observation 6. Similar to the proof of Observation 5. □

Observation 7. $\mathsf{labels}(\mathsf{val}(e_i)) \leq |A| + |NX_i| + 2 - i$.

Proof of Observation 7. From the construction of e_1 it is clear that the claim holds for $i = 1$. Assume that the claim holds for $j < i$. In the construction of e_i, $|NX_i \setminus NX_{i-1}|$ new labels are added to e_{i-1} and the label of x_i is released (i.e., is renamed to the dead label). It follows that the claim holds also for $j = i$. □

Let f'_0 be the expression obtained by adding to e_m all vertices in $B_2 \setminus NX_m$ one after the other, each vertex with a unique label. By Observations 7 and 4,

$$\begin{aligned} \mathsf{labels}(f'_0) &= \mathsf{labels}(e_m) + |B_2 \setminus NX_m| \\ &\leq |A| + |NX_m| + 2 - m + |B_2 \setminus NX_m| \\ &= |A| + 2 - m + |B_2| \\ &= |\mathsf{Active}(A_1)| + |A_2| + 2 + |B_2| \leq k + 2. \end{aligned}$$

Hence f'_0 is a $(k+2)$-cwd-expression. Let C denote the set of all special vertices of G'' of the form $s_{x,y}$ such that both x and y belong to $A \cup B_2$ and $s_{x,y}$ is not in e_m. Let $f_0 = \mathsf{addSpecial}(f'_0, C)$. Let $\mathsf{Active}(A_1) = \{u_1, \ldots, u_p\}$ and $\mathsf{Active}(B_1) = \{u'_1, \ldots, u'_p\}$, such that u_i and u'_i, $1 \leq i \leq p$, have the same label in $\mathsf{val}(t\langle a\rangle)$. Let ℓ_{u_i} denote the label of u_i at $\mathsf{val}(f_0)$. Let f_i be the expression obtained by adding vertices u'_i and s_{u_i,u'_i} to f_{i-1} with labels $k+3$ and $k+4$, respectively, then connecting s_{u_i,u'_i} to u_i and u'_i, and finally renaming the labels of u_i and s_{u_i,u'_i} to 1 and renaming the label of u'_i to ℓ_{u_i}. In other words f_i, $1 \leq i \leq p$, is constructed as follows:

$$\begin{aligned} f'_i &= (k+4)(u_i) \oplus (k+3)(s_{u_i,u'_i}) \oplus f_i - 1 \\ f''_i &= \eta_{k+3,k+4}(\eta_{k+3,\ell_{u_i}}(f'_i)). \\ f_i &= \rho_{k+4 \to l_{u_i}}(\rho_{k+3 \to 1}(\rho_{\ell_{u_i} \to 1}(f''_i))). \end{aligned}$$

Clearly, f_i is a $(k+4)$-cwd$_1$-expression.

Observation 8. *For $1 \leq i \leq p$, $\mathsf{val}(f_i)$ includes all the edges connecting u_i to all its neighbors in G''.*

Proof of Observation 8. Let s be a neighbor of u_i in G''. Then s is a special vertex of the form $s_{u_i,v}$ where u_i and v are adjacent in G. If v is in f_0, then by the construction of f_0, s is connected u_i (and also to v) in $\mathsf{val}(f_0)$ and therefore also in $\mathsf{val}(f_i)$. Suppose v is not in f_0. Then v is not in the set $A \cup B_2$ (as all these vertices are in f_0) which implies that v must be in B_1. We assert that in this case v must be equal to u'_i, and thus the edge connecting u_i to $s = s_{u_i,u'_i}$ is established in the construction of f_i. To prove the assertion, suppose v is in B_1 and is not equal to u'_i. If $s_{u_i,v}$ is not in $t\langle a\rangle$ or is on the right side of a, then the η-operation in t which connects $s_{u_i,v}$ to u_i connects it also to u'_i, a contradiction. If $s_{u_i,v}$ is on the left side of a, then v is active at a, which implies that there is another vertex v' on the left side of a which has the same label as v at a. Thus the η-operation in t which connects $s_{u_i,v}$ to v is above a in $\mathsf{tree}(t)$, and therefore connects $s_{u_i,v}$ also to v', a contradiction. □

Let g_1 denote the expression obtained by renaming all labels of vertices of $\mathsf{val}(f_p)$ in $A_2 \setminus NY_q$ to 1.

Observation 9. *For every $u \in A_2 \setminus NY_q$, $\mathsf{val}(g_1)$ includes all the edges connecting u to all its neighbors in G''.*

Proof of Observation 9. Similar arguments as used in to the proof of Observation 8. ☐

Observation 10. $\mathsf{labels}(g_1) \leq |\mathsf{Active}(B_1)| + |NY_q| + 1 + |B_2|$.

Proof of Observation 10. The set of regular vertices in $\mathsf{val}(g_1)$ is $A \cup B_2$. From these vertices, all vertices in $\mathsf{Dead}(A_1)$ are dead (already in e_m), all vertices in $\mathsf{Active}(A_1)$ are dead but instead the vertices in $\mathsf{Active}(B_1)$ have unique labels, and all vertices in $A_2 \setminus NY_q$ are dead in the construction of g_1. Thus, the formula follows. ☐

By Observations 2 and 10, the number of labels in $\mathsf{val}(g_1)$ is at most $k+1$. Let h'_q be the expression obtained by adding y_q to g_1 with a new unique label. Let D_q denote the set of special vertices of the form $s_{x,y}$ such that both x and y are in h'_q and $s_{x,y}$ is not in h'_q. Let $h''_q = \mathsf{addSpecial}(h'_q, D_q)$. Let h_q be the expression obtained from h''_q by renaming to 1 the labels in $\mathsf{val}(h''_q)$ of all vertices in $NY_q \setminus NY_{q-1}$.

Observation 11. *For every $u \in NY_q \setminus NY_{q-1}$, $\mathsf{val}(h_q)$ includes all the edges connecting u to all its neighbors in G''.*

Proof of Observation 11. By similar arguments as used in the proof of Observation 8. ☐

Observation 12. $\mathsf{labels}(h_q) \leq \mathsf{Active}(B_1) + |NY_{q-1}| + 2 + |B_2|$.

Proof of Observation 12. Follows from Observation 10, since in the construction of h_q, one new label is added (for y_q) and all the labels of the vertices in $NY_q \setminus NY_{q-1}$ are released. For $1 \leq i \leq q-1$, we define h_i based on h_{i+1} as follows: Let h'_i be the expression obtained by adding y_i to h_{i+1} with a new unique label. Let D_i denote the set of special vertices of the form $s_{x,y}$ such that both x and y are in h'_i and $s_{x,y}$ is not in h'_i. Let $h''_i = \mathsf{addSpecial}(h'_i, D_i)$. Let h_i be the expression obtained from h''_i by renaming to 1 the labels in $\mathsf{val}(h''_i)$ of all vertices in $NY_i \setminus NY_{i-1}$. ☐

Observation 13. *For every $u \in NY_i \setminus NY_{i-1}$, $\mathsf{val}(h_i)$ includes all the edges connecting u to all its neighbors in G''.*

Proof of Observation 13. By similar arguments as used in the proof of Observation 8. ☐

Observation 14. $\mathsf{labels}(h_i) \leq |\mathsf{Active}(B_1)| + |NY_{i-1}| + |B_2| + 2 + q - i$.

Proof of Observation 14. The proof is by inverse induction on i. For $i = q$ the claim follows from Observation 11. Assume the claim holds for $j > i$. In the construction of h_i, one new label for y_i is added and all the labels of the vertices in $NY_i \setminus NY_{i-1}$ are released. Hence,

$$\begin{aligned}\mathsf{labels}(h_i) &= \mathsf{labels}(h_{i+1}) + 1 - |NY_i \setminus NY_{i-1}| \\ &\leq \mathsf{Active}(B_1) + |NY_i| + |B_2| + 2 + q - (i+1) + 1 - |NY_i - NY_{i-1}| \\ &= \mathsf{Active}(B_1) + |NY_{i-1}| + |B_2| + 2 + q - i.\end{aligned}$$

☐

We claim that h_1 is a $(k+4)$-cwd$_1$-expression which defines G''. By Observations 2, 4, 7, 10, 12, and 14, in the construction of h_1 we never used more than 4 labels than the number of labels used in the construction of t. Thus, h_1 is a $(k+4)$-cwd$_1$-expression. To see that the expression h_1 defines G'', first note that all the edges introduced are correct ones, since at each step the new edges introduced connect a special vertex to its two adjacent regular vertices. By Observations 5, 6, 8, 9, 11, and 13, for each vertex $x \in A$, all edges connecting x to all its neighbors in G'' exists in $\mathsf{val}(h_1)$. Moreover, for each special vertex of the form $s_{x,y}$ such that x is in A, the edges connecting $s_{x,y}$ to both x and y are established (at the same $\mathsf{addSpecial}$ operation in the above construction), even if y belongs to B. It follows that for each vertex $y \in B$ the edges which connect y to all its neighbors in G'' of the form $s_{x,y}$ such that $x \in A$, exist in $\mathsf{val}(h_1)$. Since in h'_1 all the vertices of B have unique labels, it follows that after the operation $\mathsf{addSpecial}(h'_1, D_1)$, $\mathsf{val}(h_1)$ includes all the edges which connect y to all its special vertices of the form $s_{y,z}$ such that $z \in B$. We conclude that for each vertex $y \in B$, all edges connecting y to all its neighbors in G'' exists in $\mathsf{val}(h_1)$. Thus, we have shown that h_1 defines G''. This completes the proof of Lemma 11. ☐

Combining the previous lemmas we now get a proof of Theorem 2.

Proof of Theorem 2. Let t be a k-cwd-expression defining G''. By Lemma 5, there exists a $(k+3)$-cwd-expression t_1 defining G'' such that t_1 has Property 1. By Lemma 6, there exists a $(k+4)$-cwd-expression t_2 defining G'' such that t_2 has Property 2. By Lemma 8 there exists a $(k+6)$-cwd-expression t_3 defining G'' such that t_3 has Property 3. By Lemma 9 there exists a $(k+6)$-cwd-expression t_4 defining G'' such that t_4 has Property 4. By Lemma 10 there exists a $(k+7)$-cwd-expression t_5 defining G'' such that t_5 has Property 5. By Lemma 11, there exists a $(k+11)$-cwd$_1$-expression t' which defines G''. This completes the proof of Theorem 2. ☐

5 From G' to G''

Now to get to the NP-completeness result we shall prove the following (where G' is the graph defined above where edges of G are replaced by three paths of length 3).

Theorem 3. $\mathrm{cwd}_1(G') \leq \mathrm{cwd}(G') + 29$.

For the proof of Theorem 3 we shall use Theorem 2 and the following two lemmas.

Lemma 12. $\mathrm{cwd}_1(G') \leq cwd_1(G'') + 9$.

Lemma 13. $\mathrm{cwd}(G'') \leq \mathrm{cwd}(G') + 9$.

Proof. PENDING ☐

Proof of Theorem 3.

$$\begin{aligned}
\mathrm{cwd}_1(G') &\leq \mathrm{cwd}_1(G'') + 9 \quad \text{(by Lemma 12)} \\
&\leq \mathrm{cwd}_1(G'') + 9 + 11 \quad \text{(by Theorem 2)} \\
&\leq \mathrm{cwd}_1(G') + 20 + 9 \quad \text{(by Lemma 13).}
\end{aligned}$$

☐