Theory of Computing Systems
https://doi.org/10.1007/s00224-021-10058-5

On the structure of solution-sets to regular word equations

Joel D. Day[1] · **Florin Manea**[2]

Accepted: 4 August 2021Published online: 28 October 2021

Abstract
For quadratic word equations, there exists an algorithm based on rewriting rules which generates a directed graph describing all solutions to the equation. For regular word equations – those for which each variable occurs at most once on each side of the equation – we investigate the properties of this graph, such as bounds on its diameter, size, and DAG-width, as well as providing some insights into symmetries in its structure. As a consequence, we obtain a combinatorial proof that the problem of deciding whether a regular word equation has a solution is in NP.

Keywords Quadratic word equations · Regular word equations · String solving · NP

1 Introduction

A *word equation* is a tuple (α, β), which we shall usually write as $\alpha \doteq \beta$, such that α and β are words comprised of letters from a *terminal alphabet* $\Sigma = \{\mathtt{a}, \mathtt{b}, \ldots\}$ and *variables* from a set $X = \{x, y, z, \ldots\}$. Solutions are substitutions of the variables for words in Σ^* making both sides identical. For example, one solution to the word equation $x\mathtt{ab}y \doteq y\mathtt{ba}x$ is given by $x \to \mathtt{b}$ and $y \to \mathtt{bab}$. A system of equations is a set of equations, and a solution to the system is a substitution for the variables which is a solution to all the equations in the system.

This article belongs to the Topical Collection: *Special Issue on International Colloquium on Automata, Languages and Programming (ICALP 2020)*
Guest Editors: Artur Czumaj and Anuj Dawar

✉ Joel D. Day
J.Day@lboro.ac.uk

Florin Manea
florin.manea@informatik.uni-goettingen.de

[1] Loughborough University, Loughborough, UK

[2] Georg-August-Universität Göttingen, Göttingen, DE, Germany

One of the most fundamental questions concerning word equations is the satisfiability problem: determining whether or not a word equation has a solution. The first general algorithm for the satisfiability problem was presented by Makanin [22] in 1977. Since then, several further algorithms have been presented. Most notable among these are the algorithm given by Plandowski [25] which demonstrated that the problem is included in the complexity class PSPACE, the algorithm based on Lempel-Ziv encodings by Plandowksi and Rytter [26], and the method of recompression by Jeż, which has since been shown to require only non-deterministic linear space [15, 16]. On the other hand, it is easily seen that solving word equations is NP-hard due to fact that the subcase when one side of the equation consists only of terminals is exactly the pattern matching problem which is NP-complete [3, 12]. It remains a long-standing open problem whether or not the satisfiability problem for word equations is contained in NP.

Recently, there has been elevated interest in solving more general versions of the satisfiability problem, originating from practical applications in e.g. software verification where several *string solving* tools capable of solving word equations are being developed [1, 2, 4, 6, 18] and database theory [13, 14], where one asks whether a given (system of) word equation(s) has a solution which satisfies some additional constraints. Prominent examples include requiring that the substitution for a variable x belongs to some regular language $\mathcal{L}_x$ (regular constraints), or that the lengths of the substitutions of the variables satisfy a set of given linear diophantine equations. Adding regular constraints makes the problem PSPACE complete (see [10, 25, 27]), while it is another long standing open problem whether the satisfiability problem with length constraints is decidable. There are also many other kinds of constraints, however many lead to undecidable variants of the satisfiability problem [7, 19]. The main difficulty in dealing with additional constraints is that the solution-sets to word equations are often infinite sets with complex structures. For example, they are not parametrisable [24], and the set of lengths of solutions is generally not definable in Presburger arithmetic [20]. Thus, a better understanding of the solution-sets and their structures is a key aspect of improving our ability to solve problems relating to word equations both in theory and practice.

Quadratic word equations (QWEs) are equations in which each variable occurs at most twice. For QWEs, a conceptually simple and easily implemented algorithm exists which produces a representation of the set of all solutions as a graph. Despite this, however, the satisfiability problem for quadratic equations remains NP-hard, even for severely restricted subclasses [8, 11], while inclusion in NP, and whether the satisfiability problem with length constraints is decidable, have remained open for a long time, just as for the general case.

The algorithm solving QWEs is based on iteratively rewriting the equation(s) according to some simple rules called *Nielsen transformations*. If there exists a sequence of transformations from the original equation to the trivial equation $\varepsilon \doteq \varepsilon$, then the equation has a solution. Otherwise, there is no solution. Hence the satisfiability problem becomes a reachability problem for the underlying rewriting transformation relation, which we denote $\Rightarrow_{NT}$. It is natural to represent this relation as a directed graph $\mathcal{G}^{\Rightarrow_{NT}}$ in which the vertices are word equations and the edges are the rewriting transformations. This has the advantage that the set of all solutions to

an equation E corresponds exactly to the set of walks in the graph starting at E and finishing at the trivial equation $\varepsilon \doteq \varepsilon$.[1] Consequently, the properties of the subgraph of $\mathscr{G}^{\Rightarrow_{NT}}$ containing all vertices reachable from E (denoted $\mathscr{G}^{\Rightarrow_{NT}}_{[E]}$) are also informative about the set of solutions to the equation. For example, in [24] a connection is made between the non-parametrisability of the solution set of E and the occurrence of combinations of cycles in the graph. Since equations with a paramtrisable solution set are much easier to work with when dealing with additional constraints, this also establishes a connection between the structure of $\mathscr{G}^{\Rightarrow_{NT}}_{[E]}$ and the potential (un)decidability of variants of the satisfiability problem. Moreover, new insights into the structure and symmetries of these graphs are necessary for better understanding and optimising the practical performance of the algorithm.

Our contribution We consider a subclass of QWEs called regular equations (RWEs) introduced in [23]. A word equation is *regular* if each variable occurs at most once on each side of the equation. Thus, for example, $x\mathtt{a}by \doteq y\mathtt{ba}x$ is regular while $x\mathtt{ab}x \doteq y\mathtt{ba}y$ is not. Understanding RWEs is a vital step towards understanding the quadratic case, not only because they constitute a significant and general subclass, but also because many non-regular quadratic equations can exhibit the same behaviour as regular ones (consider, e.g. $zz \doteq x\mathtt{ab}yy\mathtt{ba}x$ for which all solutions must satisfy $z = x\mathtt{ab}y = y\mathtt{ba}x$). The satisfiability problem was shown in [8] to be NP-hard for RWEs, and shown to be in NP in [9] for some restricted subclasses including the classes of regular-reversed and regular-ordered equations.

For RWEs E, we investigate the structure of the graphs $\mathscr{G}^{\Rightarrow_{NT}}_{[E]}$, and as a consequence, are able to describe some of their most important properties. We achieve this by first noting that $\mathscr{G}^{\Rightarrow_{NT}}_{[E]}$ can be divided into strongly connected components $\mathscr{G}^{\Rightarrow}_{[E']}$ for which all the vertices are equations of the same length ($\Rightarrow$ shall be used to denote the restriction of $\Rightarrow_{NT}$ to length preserving transformations only). The 'full' graph $\mathscr{G}^{\Rightarrow_{NT}}_{[E]}$ is comprised of these individual components $\mathscr{G}^{\Rightarrow}_{[E']}$ arranged in a DAG-like structure of linear depth (see Section 3) and therefore many properties and parameters of the 'full' graph $\mathscr{G}^{\Rightarrow_{NT}}_{[E]}$ are determined by the equivalent properties and parameters of the individual components $\mathscr{G}^{\Rightarrow}_{[E']}$. We then focus on the structure of the subgraphs $\mathscr{G}^{\Rightarrow}_{[E']}$, and as a result are able to give bounds on certain parameters such as diameter, size, and DAG-width.

Our structural results come in two stages, based on whether the equation belongs to a the class of 'jumbled' equations introduced in Section 6. In the first stage, we consider equations which are not jumbled, and we show that for all such equations E, there exists a jumbled equation $\hat{E}$ such that $\mathscr{G}^{\Rightarrow}_{[E]}$ is comprised mainly of several well-

[1] Each choice of edge in a walk can be seen as a decision about the corresponding solution. It is not necessarily true that different walks will result in different solutions. However, all possible decisions are accounted for, so it is guaranteed that for every solution there is a walk from E to $\varepsilon \doteq \varepsilon$ which corresponds to that solution.

connected near-copies of $\mathscr{G}^{\Rightarrow}_{[\hat{E}]}$. For jumbled equations $\hat{E}$, we show in Section 7 that every vertex in $\mathscr{G}^{\Rightarrow}_{[\hat{E}]}$ is close to a vertex in a certain normal form. We show that the vertices in this normal form are determined to a large extent by a property invariant under $\Rightarrow$ introduced in Section 5.

With regards to the diameter of $\mathscr{G}^{\Rightarrow}_{[E']}$, we give upper bounds which are polynomial in the length of the equation. It follows that the diameter of the full graph $\mathscr{G}^{\Rightarrow_{NT}}_{[E]}$ is also polynomial, and consequently, that the satisfiability problem for RWEs is NP-complete. This can be generalised to systems of equations satisfying a natural extension of the regularity property (see Section 11). We also give exact upper and lower bounds on the number of vertices[2] in $\mathscr{G}^{\Rightarrow}_{[E']}$ for a subclass of RWEs called *basic* RWEs (see Section 4), as well as describing exactly for which equations these bounds are achieved. For RWEs which are not basic, we can infer similar bounds, at the cost of a small (linear in the length of the equation) degree of imprecision. Since in the worst case (e.g. for equations without a solution), running the algorithm will perform a full 'search' of the graph, the number of vertices is integral to the running time of the algorithm, and is potentially a better indicator of difficult instances than the complexity class alone. An example of this, comes from comparing two subclasses of RWEs called regular-ordered and regular rotated equations. It follows from our results that while both classes have an NP-complete satisfiability problem, if E' is regular-ordered, then $\mathscr{G}^{\Rightarrow}_{[E']}$ will contain at most n vertices, where n is the length of the equation, while if E' is regular rotated, but not regular-ordered, then $\mathscr{G}^{\Rightarrow}_{[E']}$ will contain $\frac{n!}{2}$ vertices, indicating a vast difference in the number of vertices the algorithm would have to visit.

Motivated by generalisations of the satisfiability problem permitting additional constraints, we also consider the connectivity of the graphs $\mathscr{G}^{\Rightarrow_{NT}}_{[E]}$. To do this, we use DAG-width, a measure for directed graphs which is in several ways analogous to treewidth for undirected graphs. Intuitively, equations for which $\mathscr{G}^{\Rightarrow_{NT}}_{[E]}$ has low DAG-width are likely to be more amenable when considering additional constraints such as length constraints (see Section 3.3). We give an example class of equations for which the DAG-width is unbounded, as well as a class for which the DAG-width is at most two. The latter includes the class of regular-ordered equations which is the most general subclass of QWEs for which it is known that the satisfiability problem with length constraints is decidable [20], and we expect that both cases will be interesting classes to consider in the context of this problem.

2 Preliminaries

For a set S, we denote the cardinality of S by $\mathrm{Card}(S)$. Let Σ be an alphabet. By Σ^*, we denote the set of all words over Σ, and by ε the empty word. By Σ^+,

[2]We consider the number of vertices, rather than edges, because it is the number of vertices which is relevant to the performance of the algorithm, and by definition of $\Rightarrow_{NT}$, the out-degree of the graph is bounded by a constant so the the number of edges is linear in the number of vertices.

we denote the free semigroup $\Sigma^*\backslash\{\varepsilon\}$. A word u is a prefix (resp. suffix) of a word w if there exists v such that $w = uv$ (resp. $w = vu$). Similarly, u is a factor of w if there exist v, v' such that $w = vuv'$. A prefix/suffix/factor is *proper* if is neither the whole word w, nor ε. The length of a word w is denoted $|w|$, while for $\mathtt{a} \in \Sigma$, $|w|_{\mathtt{a}}$ denotes the number of occurrences of $\mathtt{a}$ in w. For a word $w = \mathtt{a}_1\mathtt{a}_2 \ldots \mathtt{a}_n$ with $\mathtt{a}_i \in \Sigma$ for $1 \leq i \leq n$, the notation $w[i]$ refers to the letter $\mathtt{a}_i$ in the i^{th} position. By w^R, we denote the reversal $\mathtt{a}_n\mathtt{a}_{n-1} \ldots \mathtt{a}_1$ of the word w. Two words w_1, w_2 are conjugate (written $w_1 \sim w_2$) if there exist u, v such that $w_1 = uv$ and $w_2 = vu$.

We shall generally distinguish between two types of alphabet: an infinite set $X = \{x_1, x_2, \ldots\}$ of variables, and a set $\Sigma = \{\mathtt{a}, \mathtt{b}, \ldots\}$ of terminal symbols. We shall assume that $\mathrm{Card}(\Sigma) \geq 2$, and that there exists an order on X leading to a lexicographic order on X^*. For a word $\alpha \in (X \cup \Sigma)^*$, we shall denote by $var(\alpha)$ the set $\{x \in X \mid x \text{ is a factor of } \alpha\}$. We shall denote by $qv(\alpha)$ the set $\{x \in var(\alpha) \mid |\alpha|_x = 2\}$. A word equation is a tuple $(\alpha, \beta) \in (X \cup \Sigma)^* \times (X \cup \Sigma)^*$, usually written $\alpha \doteq \beta$. Solutions are morphisms $h : (X \cup \Sigma)^* \to \Sigma^*$ with $h(\mathtt{a}) = \mathtt{a}$ for all $\mathtt{a} \in \Sigma$ such that $h(\alpha) = h(\beta)$. The satisfiability problem is the problem of deciding algorithmically whether a given word equation has a solution. For equations E given by $\alpha \doteq \beta$, we shall often extend notations regarding words in $(X \cup \Sigma)^*$ to E for convenience, so that, e.g. $|E| = |\alpha\beta|$, $var(E) = var(\alpha\beta)$ and $qv(E) = qv(\alpha\beta)$. An equation $\alpha \doteq \beta$ is quadratic if $|\alpha\beta|_x \leq 2$ for all $x \in X$. It is regular if $|\alpha|_x \leq 1$ and $|\beta|_x \leq 1$ hold for all $x \in X$. Thus all regular equations are quadratic, but not all quadratic equations are regular. We shall usually abbreviate regular (resp. quadratic) word equation to RWE (resp. QWE). For $Y \subseteq X$, let $\pi_Y : (X \cup \Sigma^*) \to Y^*$ be the morphism such that $\pi_Y(x) = x$ if $x \in Y$ and $\pi_Y(x) = \varepsilon$ otherwise; i.e. π_Y is a projection from $(X \cup \Sigma)^*$ onto Y^*. A regular equation E given by $\alpha \doteq \beta$ is regular-ordered if $\pi_{qv(E)}(\alpha) = \pi_{qv(E)}(\beta)$, it is regular rotated if $\pi_{qv(E)}(\alpha) \sim \pi_{qv(E)}(\beta)$ and it is regular reversed if $\pi_{qv(E)}(\alpha) = \pi_{qv(E)}(\beta)^R$.

Given a set S and binary relation $\mathscr{R} \subseteq S \times S$, we denote the reflexive-transitive closure of $\mathscr{R}$ as $\mathscr{R}^*$. For each $s \in S$, we denote by $[s]_{\mathscr{R}}$ the set $\{s' \mid (s, s') \in \mathscr{R}^*\}$. The relation $\mathscr{R}$ may be represented as a directed graph, which we denote $\mathscr{G}^{\mathscr{R}}$, with vertices from S and edges from $\mathscr{R}$. Usually, we will be interested in the subgraph of $\mathscr{G}^{\mathscr{R}}$ containing vertices belonging to $[s]_{\mathscr{R}}$ for some $s \in S$. Thus, for a subset T of S we shall denote by $\mathscr{G}^{\mathscr{R}}_T$ the subgraph of $\mathscr{G}^{\mathscr{R}}$ containing vertices from T. Given a (directed) graph $\mathscr{G}$, with vertices $V(\mathscr{G})$ and edges $E(\mathscr{G})$, a root vertex is some $v \in V(\mathscr{G})$ such that there does not exist $(u, v) \in E(\mathscr{G})$. We denote by $diam(\mathscr{G})$ the diameter of the graph $\mathscr{G}$, by which we mean the maximum length of a shortest (directed) path between two vertices. For our purposes, we are really interested in the maximum length of shortest paths only when they exist, meaning that we shall not adopt the convention that $diam(\mathscr{G}) = \infty$ when $\mathscr{G}$ is a directed graph which is not strongly connected.

For $W, V' \subseteq V(\mathscr{G})$, we say that W *guards* V' if for all $(u, v) \in E(\mathscr{G})$ with $u \in V'$, we have $v \in V' \cup W$. If $\mathscr{G}$ is acyclic, we write $v_1 \leq_{\mathscr{G}} v_2$ if there is a directed path from v_1 to v_2 in $\mathscr{G}$ or $v_1 = v_2$. Following [5], A DAG-decomposition of $\mathscr{G}$ is a pair (D, χ) such that D is a directed acyclic graph (DAG) with vertices $V(D)$, and $\chi = \{X_d \mid d \in V(D)\}$ is a family of subsets of $V(\mathscr{G})$ satisfying:

(D1) $V(\mathscr{G}) = \bigcup_{d \in V(D)} X_d$,
(D2) if $d, d', d'' \in V(D)$ such that $d \leq_D d' \leq_D d''$, then $X_d \cap X_{d''} \subseteq X_{d'}$,
(D3) For all edges (d, d') of D, $X_d \cap X_{d'}$ guards $X_{\geq d'} \setminus X_d$, where $X_{\geq d'} = \bigcup_{d'' \geq_D d'} X_{d''}$, and for all root vertices d, $X_{\geq d}$ is guarded by $\emptyset$.

The width of the DAG-decomposition is $\max\{\text{Card}(X_d) \mid d \in V(D)\}$. The DAG-width of $\mathscr{G}$ is the minimum width of any possible DAG-decomposition of $\mathscr{G}$ and is denoted $dgw(\mathscr{G})$.

3 An Algorithm for Solving Regular Word Equations

In this section we present the algorithm for solving QWEs as a rewriting system defined by a relation $\Rightarrow_{NT}$. The rewriting relation is derived from morphisms called Nielsen transformations, and we shall abuse this terminology slightly and generally also refer to the rewriting transformations themselves as Nielsen transformations. The Nielsen transformations never introduce new variables or terminal symbols, and never increase the length of the equation. They also preserve the properties of being quadratic (resp. regular). Thus, given a quadratic (resp. regular) word equation E, the set $\{E' \mid E \Rightarrow_{NT}^* E'\}$ of equations reachable via Nielsen transformations is finite. Moreover, given an equation which has a solution h, there is always a Nielsen transformation which produces an equation which has a solution, such that at least one of the new equation or the new solution is strictly shorter than the previous one. It follows that, given an equation which possesses a solution, it is possible to reach the equation $\varepsilon \doteq \varepsilon$ after finitely many rewriting steps. For a more detailed description of the algorithm, we refer the reader to e.g. Chapter 12 of [21].

3.1 Nielsen Transformations

The Nielsen transformations (morphisms) are defined as follows: for $x \in X \cup \Sigma$ and $y \in X$, let $\psi_{x<y} : (X \cup \Sigma)^* \to (X \cup \Sigma)^*$ be the morphism given by $\psi_{x<y}(y) = xy$ and $\psi_{x<y}(z) = z$ whenever $z \neq y$. We define the rewriting transformations via the relations $\Rightarrow_L, \Rightarrow_R, \Rightarrow_>$ as follows. Suppose we have a QWE E of the form $x\alpha \doteq y\beta$ where $x, y \in X \cup \Sigma$ and $\alpha, \beta \in (X \cup \Sigma)^*$. Then:

1. if $x \in qv(E)$ and $x \neq y$, then $x\alpha \doteq y\beta \Rightarrow_L x\psi_{y<x}(\alpha) \doteq \psi_{y<x}(\beta)$, and
2. if $y \in qv(E)$ and $x \neq y$, then $x\alpha \doteq y\beta \Rightarrow_R \psi_{x<y}(\alpha) \doteq y\psi_{x<y}(\beta)$, and
3. if $x \in X \setminus qv(E)$, then $x\alpha \doteq y\beta \Rightarrow_> x\alpha \doteq \beta$, and
4. if $y \in X \setminus qv(E)$, then $x\alpha \doteq y\beta \Rightarrow_> \alpha \doteq y\beta$, and
5. if $x = y$, then $x\alpha \doteq y\beta \Rightarrow_> \alpha \doteq \beta$.

Moreover, for a QWE E of the form $\alpha \doteq \beta$ with $\alpha, \beta \in (X \cup \Sigma)^*$, and for each $Y \subseteq var(E)$, we have the additional transformations $\alpha \doteq \beta \Rightarrow_> \pi_{X \setminus \{Y\}}(\alpha) \doteq \pi_{X \setminus \{Y\}}(\beta)$.

Now, our full rewriting relation, $\Rightarrow_{NT}$, is given by $\Rightarrow_L \cup \Rightarrow_R \cup \Rightarrow_>$.[3] For convenience, we shall define $\Rightarrow$ to be $\Rightarrow_L \cup \Rightarrow_R$. We shall call the rewriting transformations from $\Rightarrow$ *length-preserving*, since they are exactly those for which the resulting equation has the same length as the original. The following observation follows directly from the definition of $\Rightarrow_{NT}$.

Remark 3.1 Let E, E' be QWEs such that $E \Rightarrow_{NT} E'$. If E is regular, then E' is regular. Moreover, if $E \Rightarrow E'$, then $var(E) = var(E')$, $qv(E) = qv(E')$, and $|E| = |E'|$. Similarly, if $E \Rightarrow_> E'$, then $var(E') \subseteq var(E)$, $qv(E') \subseteq qv(E)$, and $|E'| < |E|$. Hence the set $\{E'' \mid E \Rightarrow_{NT}^* E''\}$ is finite.

If E_1, E_2 are RWEs such that $E_1 \Rightarrow_L E_2$, then it follows from the definitions that there exist $x, y \in X$ and $\alpha_1, \alpha_2, \beta_1, \beta_2 \in (X \backslash \{x, y\})^*$ such that E_1 is given by $x\alpha_1 y\alpha_2 \doteq y\beta_1 x\beta_2$ and E_2 is given by $x\alpha_1 y\alpha_2 \doteq \beta_1 yx\beta_2$. Extending this observation to multiple applications of $\Rightarrow_L$, we may conclude that the set $\{E_2 \mid E_1 \Rightarrow_L^* E_2'\}$ is exactly the set $\{x\alpha_1 y\alpha_2 \doteq \beta_3 x\beta_2 \mid \beta_3 \sim y\beta_1\}$. A similar statement can be made for $\Rightarrow_R^*$. Consequently, the reflexive transitive closures $\Rightarrow_L^*$ and $\Rightarrow_R^*$ are symmetric. Hence, we may also observe the following.

Remark 3.2 Let E be a RWE and $Z \in \{L, R\}$. Then $\mathrm{Card}(\{E' \mid E \Rightarrow_Z^* E'\}) < |E|$ and $\Rightarrow_Z^*$ is an equivalence relation. It follows that $\Rightarrow^*$ is also an equivalence relation.

The following well-known result forms the basis for the algorithm for solving QWEs.

Theorem 3.3 [21] *Let E be a QWE. Then E has a solution if and only if $E \Rightarrow_{NT}^* \varepsilon \doteq \varepsilon$.*

3.2 Representing the Set of Solutions as a Graph

Theorem 3.3 provides the basis for treating the satisfiability of QWEs as a reachability problem for the rewriting relation $\Rightarrow_{NT}$. Since any relation R is naturally represented as a (directed) graph $\mathcal{G}^R$, it is also natural to interpret the resulting algorithm as a search in the graph $\mathcal{G}_{[E]}^{\Rightarrow_{NT}}$: it suffices to to determine whether there exists a path in the graph from the original equation E to the trivial equation $\varepsilon \doteq \varepsilon$. In fact, the graph $\mathcal{G}_{[E]}^{\Rightarrow_{NT}}$ can tell us significantly more than simply whether a solution to E exists: every walk from E to $\varepsilon \doteq \varepsilon$ in $\mathcal{G}_{[E]}^{\Rightarrow_{NT}}$ corresponds to a solution to E and likewise, every solution to E is represented by a walk in $\mathcal{G}_{[E]}^{\Rightarrow_{NT}}$ from E to $\varepsilon \doteq \varepsilon$. Thus the graphs $\mathcal{G}_{[E]}^{\Rightarrow}$ contain a full description of all solutions to E, and as such, their

[3] There are several possible variations on the definition of the length-reducing rewriting transformations $\Rightarrow_>$ for which the algorithm remains correct and is guaranteed to terminate. However, for our results, the exact choice is not important as we concentrate our investigations on the length preserving part $\Rightarrow$ of the rewriting relation for reasons described in Section 3.2.

properties and structure are of inherent interest to the study of QWEs and their solutions. An immediate example of this is the diameter, which is strongly related to the complexity of the satisfiability problem, as demonstrated in the following proposition.

Proposition 3.4 *Let $\mathscr{C}$ be a class of QWEs. Suppose there exists a constant $k \in \mathbb{N}$ such that for each $E \in \mathscr{C}$, we have $diam(\mathscr{G}^{\Rightarrow_{NT}}_{[E]}) \in O(|E|^k)$. Then the satisfiability problem for $\mathscr{C}$ is in NP.*

Proof Let $\mathscr{C}$ be a class of quadratic word equations and let $k \in \mathbb{N}$ such that for each $E \in \mathscr{C}$, $diam(\mathscr{G}^{\Rightarrow_{NT}}_{[E]}) \in O(|E|^k)$. By Theorem 3.3, to check whether an equation $E \in \mathscr{C}$ has a solution, we have to check whether there is a path from E to $\varepsilon \doteq \varepsilon$ in $\mathscr{G}^{\Rightarrow_{NT}}_{[E]}$. If such a path exists, then due to our assumptions about the diameter, one exists of length at most $O(|E|^k)$. Moreover, for each edge $E_1 \Rightarrow_{NT} E_2$ in the path, we have that $|E_2| \leq |E_1| \leq |E|$, so verifying that $E_1 \Rightarrow_{NT} E_2$ can be achieved in linear time. Hence, subject to appropriate non-deterministic choices, we may find such a path whenever it exists in $O(|E|^{k+1})$ time and the satisfiability problem for $\mathscr{C}$ is in NP. □

Many properties will be determined mostly (i.e. up to some small imprecision) on the subgraphs obtained by restricting our rewriting relation to length-preserving transformations only (i.e. to $\Rightarrow$). Since the rewriting relation $\Rightarrow_{NT}$ allows us to preserve or decrease the length, but never increase it again, any walk in the graph will visit a subgraph containing equations of each length only once, and in order of decreasing length. The following proposition confirms how we may infer a global property of $\mathscr{G}^{\Rightarrow_{NT}}_{[E]}$ from its 'local' values in the individual subgraphs $\mathscr{G}^{\Rightarrow}_{[E']}$ in the case of two properties we are particularly interested in: diameter and DAG-width.

Proposition 3.5 *Let E be a QWE. Then*

1. $diam(\mathscr{G}^{\Rightarrow_{NT}}_{[E]}) \leq (|E|+1)(1+\max\{diam(\mathscr{G}^{\Rightarrow}_{[E']}) \mid E \Rightarrow^*_{NT} E'\}) - 1$, *and*
2. $dgw(\mathscr{G}^{\Rightarrow_{NT}}_{[E]}) = \max\{dgw(\mathscr{G}^{\Rightarrow}_{[E']}) \mid E \Rightarrow^*_{NT} E'\}$.

Proof The second statement is a direct consequence of Theorem 6 in [5]. We shall consider the first statement. Let E be a quadratic word equation. Let

$$m = \max\{diam(\mathscr{G}^{\Rightarrow}_{[E']} \mid E \Rightarrow^*_{NT} E')\}.$$

Let $E_1, E_2, \ldots E_n$ be the shortest path in $\mathscr{G}^{\Rightarrow_{NT}}_{[E]}$ between E_1 and E_n. Then $E_i \Rightarrow_{NT} E_{i+1}$ for $1 \leq i < n$. Consequently, for each i, $1 \leq i < n$ either $|E_i| = |E_{i+1}|$ or $|E_i| > |E_{i+1}|$. Let $j_1, j_2, \ldots, j_k$ be all the indices i for which the latter holds. Then, since the length of an equation cannot be negative, we necessarily have that $k \leq |E|$. Moreover, we have that $E_1 \Rightarrow^* E_{j_1}$, $E_{j_k+1} \Rightarrow^* E_n$, and for each i, $1 \leq i < k$, $E_{j_i+1} \Rightarrow^* E_{j_{i+1}}$. Since, for each E_i, $\mathscr{G}^{\Rightarrow}_{[E_i]}$ is a subgraph of $\mathscr{G}^{\Rightarrow_{NT}}_{[E]}$, and by our assumption that the path $E_1, E_2, \ldots E_n$ is minimal in $\mathscr{G}^{\Rightarrow_{NT}}_{[E]}$, it follows that the path $E_1, E_2, \ldots, E_{j_1}$ is minimal in $\mathscr{G}^{\Rightarrow}_{[E_1]}$, and thus $j_1 - 1 \leq m$. By the same argument, the path $E_{j_k+1}, E_{j_k+2}, \ldots, E_n$ is minimal in $\mathscr{G}^{\Rightarrow}_{[E_{j_k+1}]}$ so we get that $n - j_k - 1 \leq m$

and similarly, for each i, $1 \leq i < k$, we may conclude that $j_{i+1} - j_i - 1 \leq m$. It follows that

$$n = (n - j_k) + (j_k - j_{k-1}) + \ldots + (j_2 - j_1) + j_1 \leq (k+1)(m+1)$$

meaning the length of the path $E_1, E_2, \ldots E_n$ is at most $(|E|+1)(m+1)$. Since this holds for all choices of E_1, E_n, we have that $diam(\mathscr{G}_{[E]}^{\Rightarrow NT}) \leq (|E|+1)(m+1) - 1$ as claimed. □

In what follows, we shall focus predominantly on the structure of the (sub)graphs $\mathscr{G}_{[E']}^{\Rightarrow}$ corresponding to the length-preserving transformations belonging to $\Rightarrow$ (see Fig. 1). This has the advantage of allowing us to apply further restrictions, in particular a reduction to the case of basic equations introduced in Section 4, without significantly altering the structure of the graph. It is worth pointing out that due to Remark 3.2, the graph $\mathscr{G}_{[E]}^{\Rightarrow}$ is strongly connected whenever E is a RWE. The same is generally not true in the case of arbitrary QWEs E, or for the full graph $\mathscr{G}_{[E]}^{\Rightarrow NT}$.

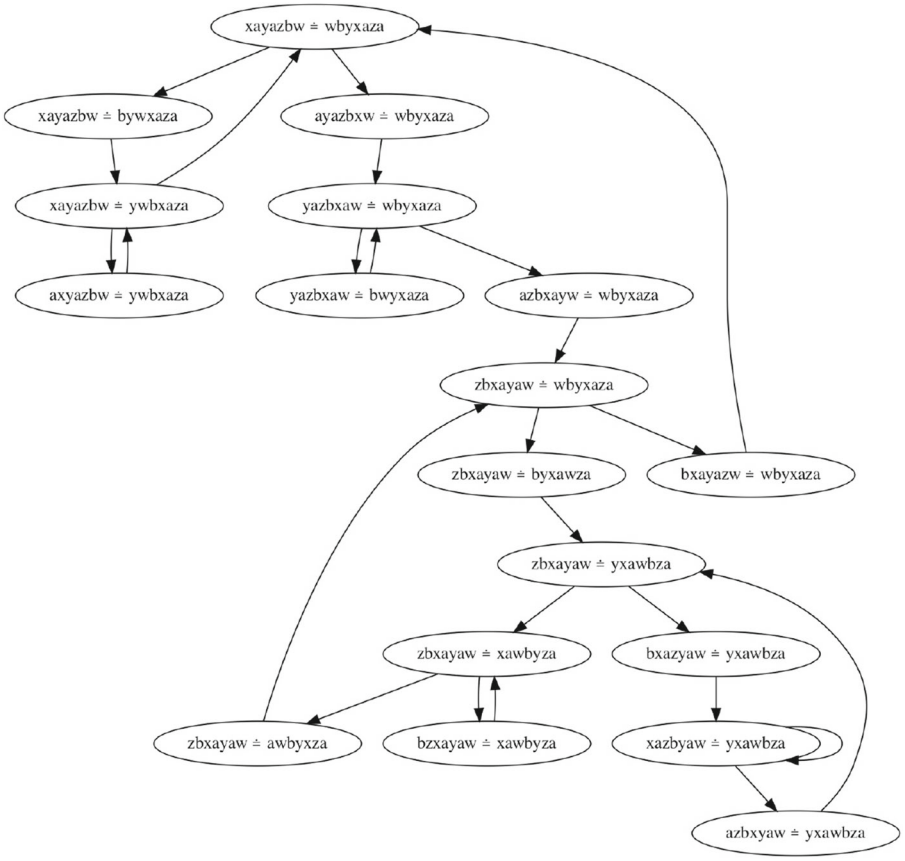

Fig. 1 The graph $\mathscr{G}_{[E]}^{\Rightarrow}$ in the case that E is the equation $x\mathtt{a}y\mathtt{a}z\mathtt{b}w \doteq w\mathtt{b}y x\mathtt{a}z\mathtt{a}$ with variables x, y, z, w and terminal symbols $\mathtt{a}, \mathtt{b}$. Generated in python using the PyDot graph drawing package

3.3 Solving Equations Modulo Constraints

Often, it is important to determine whether a given equation has a solution which satisfies some additional constraints. For some types of constraints, it is possible to adapt the algorithm by finding, for each Nielsen transformation, an appropriate corresponding transformation of the constraints. For example, if $x, y, z \in X$ and we have the length constraint $|x| = |z|$, when we apply the Nielsen transformation associated with $\psi_{y<x}$ to our equation, we replace each occurrence of x with yx. Thus, the updated constraint would be $|x| + |y| = |z|$. Unfortunately, as is the case for length constraints, the resulting set of possible equation/constraint combinations can become infinite, meaning that the modified version of the algorithm is not guaranteed to terminate.

A possible solution to this is to find finite descriptions of the potentially infinite sets of constraints which may occur alongside each equation. The task of finding such descriptions, and consequently the potential decidability of the corresponding extended satisfiability problems, is dependent on the structural properties of the graph, as can be seen e.g. in [20, 24].

One case in which computing finite descriptions is straightforward is when the graph $\mathscr{G}_{[E]}^{\Rightarrow_{NT}}$ is acyclic (i.e. a DAG). Unfortunately, inspection of the definition of $\Rightarrow_{NT}$ reveals that this is not true for the majority of RWEs (or QWEs). Hence, when considering the existence of algorithms for solving word equations with length constraints (or constraints of other types), it is natural to specifically consider classes of equations E where the graphs $\mathscr{G}_{[E]}^{\Rightarrow_{NT}}$ have particularly DAG-like (or un-DAG-like) structures, which we can measure using parameters such as DAG-width.

3.4 Properties of the Graphs $\mathscr{G}_{[E]}^{\Rightarrow_{NT}}$ for Regular Equations E

In order to understand the full graphs $\mathscr{G}_{[E]}^{\Rightarrow_{NT}}$, we mostly need to understand the (strongly connected) components corresponding to the length-preserving transformations, as we can easily see that these components will be connected in a DAG-like structure whose depth is at most $|E|$. Hence, our main goal is to describe the structure of the graphs $\mathscr{G}_{[E]}^{\Rightarrow}$ for RWEs E. This is done in several steps, with each one accounting for a particular structural feature or aspect as follows.

(1) In the first step (Section 4), we describe the effect of terminal symbols, single occurrence variables, and 'decomposability' on the structure of $\mathscr{G}_{[E]}^{\Rightarrow}$, essentially reducing the structure of $\mathscr{G}_{[E]}^{\Rightarrow}$ to $\mathscr{G}_{[E']}^{\Rightarrow}$ for a 'basic' equation E' which does not contain any of these features.

(2) Building on an important technical tool developed in Section 5, the second step (Section 6) introduces the class of jumbled equations. For equations E' which are not jumbled, but which have nevertheless been simplified as per the first step, there exists a specific repetitive structure allowing us to express $\mathscr{G}_{[E']}^{\Rightarrow}$ as a combination of (near) copies of some smaller graph $\mathscr{G}_{[E'']}^{\Rightarrow}$ where E'' is a jumbled equation obtained by deleting the appropriate variables from E'.

(3) In the third step (Section 7), we show that for jumbled equations E'', all vertices in $\mathscr{G}^{\Rightarrow}_{[E'']}$ are 'close' to a vertex from a small subset conforming to a very particular structure called Lex Normal Form.
(4) Finally, in Sections 8, 9 and 10, we exploit our structural results to investigate the diameter, number of vertices and connectivity (DAG-width) of $\mathscr{G}^{\Rightarrow}_{[E]}$ respectively. In Section 11 we note a generalisation of our results to systems of equations.

4 Basic Equations: A Convenient Abstraction

The current section is devoted to reducing the study of the graphs $\mathscr{G}^{\Rightarrow}_{[E]}$ to the case of basic equations. This has several advantages, including a significant reduction in the size of the graphs which is useful for working with examples, as well as allowing for the simpler formulation of precise results, e.g. regarding the size of the graphs in Section 9, as well as avoiding unnecessary repetition in the formal statements and their proofs.

Definition 4.1 (Basic Equations) Let E be a QWE given by $\alpha \doteq \beta$. Then E is *decomposable* if there exist proper prefixes α', β' of α and β such that $var(\alpha') \cap qv(E) = var(\beta') \cap qv(E)$. Otherwise, E is *indecomposable*. E is *basic* if it is indecomposable and $\alpha, \beta \in qv(E)^*$.

For a basic RWE, both sides of the equation are permutations of the same set of variables, for example $x_1x_2x_3 \doteq x_3x_1x_2$ and $xywz \doteq wzxy$ are both basic RWEs. On the other hand, $xyzw \doteq yxzw$, $\mathtt{a}x\mathtt{b}y \doteq y\mathtt{ba}x$ and $xy \doteq yz$ are not – the first being decomposable and the latter two containing terminal symbols and variables occurring on one side only.

We firstly consider decomposable equations E, showing that in this case the graph $\mathscr{G}^{\Rightarrow}_{[E]}$ is isomorphic to $\mathscr{G}^{\Rightarrow}_{[E']}$ for some shorter equation E'. The main step in this respect is the following observation.

Lemma 4.2 *Let E be a RWE given by $\alpha_1\alpha_2 \doteq \beta_1\beta_2$ where $\alpha_1, \alpha_2, \beta_1, \beta_2 \in (X \cup \Sigma)^*$ such that $\alpha_1, \beta_1 \neq \varepsilon$ and $var(\alpha_1) \cap qv(E) = var(\beta_1) \cap qv(E)$. Let E' be a RWE. Then $E \Rightarrow E'$ if and only if there exist $\alpha_3, \beta_3 \in (X \cup \Sigma)^*$ such that E' is given by $\alpha_3\alpha_2 \doteq \beta_3\beta_2$ and $\alpha_1 \doteq \beta_1 \Rightarrow \alpha_3 \doteq \beta_3$.*

Proof Suppose E is a RWE given by $\alpha_1\alpha_2 \doteq \beta_1\beta_2$ where $\alpha_1, \alpha_2, \beta_1, \beta_2 \in (X \cup \Sigma)^*$ with $\alpha_1, \beta_1 \neq \varepsilon$ such that $var(\alpha_1) \cap qv(E) = var(\beta_1) \cap qv(E)$. Let E' be a RWE. Suppose firstly that $\alpha_3, \beta_3 \in (X \cup \Sigma)^*$ such that $\alpha_1 \doteq \beta_1 \Rightarrow_L \alpha_3 \doteq \beta_3$ (the case that $\alpha_1 \doteq \beta_1 \Rightarrow_R \alpha_3 \doteq \beta_3$ is symmetric). Then it follows from the definition of $\Rightarrow_L$ that α_1 has a prefix $y \in qv(E)$. Hence, there exist $x \in X \cup \Sigma$ and $\gamma, \delta_1, \delta_2 \in (X \cup \Sigma)^*$ such that $\alpha_1 = y\gamma$, $\beta_1 = x\delta_1y\delta_2$, $\alpha_3 = \alpha_1$ and $\beta_3 = \delta_1xy\delta_2$. By the definition of $\Rightarrow_L$, it follows that $\alpha_1\alpha_2 \doteq \beta_1\beta_2 \Rightarrow_L \alpha_3\alpha_2 \doteq \beta_3\beta_2$ and thus $E \Rightarrow E'$.

Now suppose instead that $E \Rightarrow_L E'$ (again, the case that $E \Rightarrow_R E'$ is symmetric). Then by definition of $\Rightarrow_L$, there exists a variable $y \in qv(E)$ in the leftmost position

of α_1 which also occurs in $\beta_1\beta_2$. Moreover, it follows from the definition of $\Rightarrow_L$ and the fact that $E \Rightarrow_L E'$ that $y \neq \beta_1[1]$. Furthermore, since $var(\alpha_1) \cap qv(E) = var(\beta_1) \cap qv(E)$, y must in fact occur somewhere in β_1, so there exist $x \in X \cup \Sigma$ and $\gamma, \delta_1, \delta_2 \in (X \cup \Sigma)^*$ such that $\alpha_1 = y\gamma$ and $\beta_1 = x\delta_1 y\delta_2$, and such that E' is given by $\alpha_3\alpha_2 \doteq \beta_3\beta_2$ where $\alpha_3 = \alpha_1$ and $\beta_3 = \delta_1 x y \delta_2$. It follows from the definition of $\Rightarrow_L$ that $\alpha_1 \doteq \beta_1 \Rightarrow_L \alpha_3 \doteq \beta_3$ and thus the statement holds. □

It follows immediately from Lemma 4.2 that the relation $\Rightarrow$ preserves the properties of being (in)decomposable and basic.

Corollary 4.3 *Let E_1, E_2 be RWEs such that $E_1 \Rightarrow E_2$. Then E_1 is indecomposable if and only if E_2 is indecomposable. Consequently E_1 is basic if and only if E_2 is basic.*

Moreover, a straightforward induction yields the following description of the graphs $\mathscr{G}^{\Rightarrow}_{[E]}$ in the case that E is decomposable.

Corollary 4.4 *Let E be a decomposable RWE given by $\alpha_1\alpha_2 \doteq \beta_1\beta_2$ where $\alpha_1, \alpha_2, \beta_1, \beta_2 \in (X \cup \Sigma)^*$ such that $\alpha_1, \beta_1 \neq \varepsilon$ and $var(\alpha_1) \cap qv(E) = var(\beta_1) \cap qv(E)$. Then $\mathscr{G}^{\Rightarrow}_{[E]}$ is isomorphic to $\mathscr{G}^{\Rightarrow}_{[\alpha_1 \doteq \beta_1]}$ and can be obtained from $\mathscr{G}^{\Rightarrow}_{[\alpha_1 \doteq \beta_1]}$ by replacing each vertex $\alpha_3 \doteq \beta_3 \in [\alpha_1 \doteq \beta_1]_{\Rightarrow}$ with $\alpha_3\alpha_2 \doteq \beta_3\beta_2$.*

Corollary 4.4 accounts for decomposable equations. It remains to consider the case of equations containing terminal symbols and variables occurring on only one side (and therefore once overall). For this case, we need the following notion for relating the structure of two graphs.

Definition 4.5 (Isolated path compression) Let G_1, G_2 be (directed) graphs. We say that G_1 is an *isolated path compression* of order n of G_2 if G_2 may be obtained from G_1 by replacing each edge (e, e') in G_1 by a path $(e, e_1), (e_1, e_2), \ldots (e_{k-1}, e_k), (e_k, e')$ such that $k \leq n$ and $e_1, e_2, e_3, \ldots, e_k$ are new vertices unique to the edge (e, e').

Informally, an isolated path compression of a graph is obtained simply by replacing 'isolated paths' (paths whose internal vertices are not adjacent to to any vertices outside the path) of a bounded length with single edges. Therefore, the overall structure is generally preserved, and most properties will be preserved, or change proportionally to the order n (Fig. 2).

Remark 4.6 Consider graphs G_1, G_2 such that G_1 is an isolated path compression of order n of G_2. If $dgw(G_1) = 1$, then $dgw(G_2) \in \{1, 2\}$.[4]

[4]The case that $dgw(G_1) = 1$ and $dgw(G_2) = 2$ is a special case arising from the possibility of 'isolated cycles' being compressed into singleton self-loops.

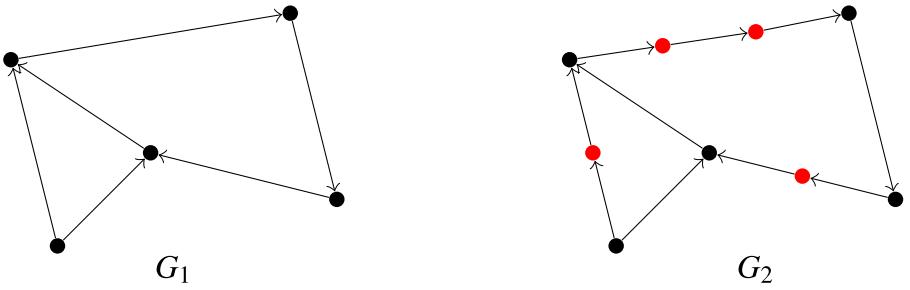

Fig. 2 The graph G_1 is an isolated path compression of order two of the graph G_2

If $dgw(G_1) \geq 2$, then the $dgw(G_1) = dgw(G_2)$. Moreover, $diam(G_2) \leq (n+1)diam(G_1)$, and the number of vertices (resp. edges) in G_2 is at most the number of vertices in G_1 plus n times the number of edges of G_1.

Using isolated path compressions, it is possible to describe the structure of the graph $\mathscr{G}^{\Rightarrow}_{[E]}$ for any RWE E in terms of the graph $\mathscr{G}^{\Rightarrow}_{[E']}$ for the RWE E' obtained from E by erasing all terminal symbols and single-occurrence variables from E (i.e. projecting onto $qv(E)$).

Lemma 4.7 *Let E be an indecomposable RWE given by $\alpha \doteq \beta$. Then the graph $\mathscr{G}^{\Rightarrow}_{[\pi_{qv(E)}(\alpha) \doteq \pi_{qv(E)}(\beta)]}$ is isomorphic to an isolated path compression of order $|E|$ of $\mathscr{G}^{\Rightarrow}_{[E]}$.*

Proof Let E be an indecomposable RWE given by $\alpha \doteq \beta$. Note that by Corollary 4.4, it follows that E' is indecomposable for every $E' \in [E]_{\Rightarrow}$. We begin by considering the simple cases arising when $\mathrm{Card}(qv(E)) < 2$. If $\mathrm{Card}(qv(E)) = 0$, then $\mathscr{G}^{\Rightarrow}_{[E]}$ is a single vertex with no edges. Moreover, $\pi_{qv(E)}(\alpha) \doteq \pi_{qv(E)}(\beta)$ is the trivial equation $\varepsilon \doteq \varepsilon$, so $\mathscr{G}^{\Rightarrow}_{[\pi_{qv(E)}(\alpha) \doteq \pi_{qv(E)}(\beta)]}$ is also a single vertex with no edges. The two graphs are clearly isomorphic, so the lemma holds trivially.

Now suppose that $\mathrm{Card}(qv(E)) = 1$. Then E has the form $w_1 x w_2 \doteq w_3 x w_4$ where $qv(E) = \{x\}$ and $w_1, w_2, w_3, w_4 \in ((X \cup \Sigma)\backslash\{x\})^*$. It necessarily follows that the equation $\pi_{qv(E)}(\alpha) \doteq \pi_{qv(E)}(\beta)$ has the form $x \doteq x$, meaning that $\mathscr{G}^{\Rightarrow}_{[\pi_{qv(E)}(\alpha) \doteq \pi_{qv(E)}(\beta)]}$ is again a single vertex with no edges. If $w_1, w_2 \neq \varepsilon$, then E is decomposable, a contradiction. Otherwise, $\mathscr{G}^{\Rightarrow}_{[E]}$ is a cycle of length $\max\{|w_1|, |w_2|\} < |E|$, so again the statement of the lemma follows directly. Thus, for the remainder of the proof, we shall suppose that $\mathrm{Card}(qv(E)) \geq 2$.

Before proceeding, we remark that for any equation E' given by $\alpha' \doteq \beta'$, if $\alpha'[1], \beta'[1] \notin qv(E')$, then either E' is decomposable, or $|\alpha'|, |\beta'| \in \{0, 1\}$. Both are contradictions to previous assumptions (the former to the fact that E is indecomposable, and hence E' is indecomposable for all $E' \in [E]_{\Rightarrow}$, and the latter to the assumption that $\mathrm{Card}(qv(E)) \geq 2$ which is only possible if $|\alpha|, |\beta| \geq 2$). Consequently, we may partition $[E]_{\Rightarrow}$ into two sets S_1 and S_2 where S_1 contains all equations E' given by $\alpha' \doteq \beta'$ such that $\alpha'[1]$ and $\beta'[1]$ are both in $qv(E')$, and S_2

contains all equations E' given by $\alpha' \doteq \beta'$ such that exactly one of $\alpha'[1], \beta'[1]$ is in $qv(E')$. Intuitively, S_1 will be the set of 'surviving' vertices in the isolated path compression while S_2 consists of those vertices which belong only to the 'isolated paths' which are contracted/compressed. Supporting this, we show the following two claims regarding elements of S_2.

Claim 4.7.1 Suppose that $E' \in S_2$. Then the in-degree and out-degree of E' in $\mathscr{G}^{\Rightarrow}_{[E]}$ are both exactly one.

Proof W.l.o.g. suppose that E' is given by $x\alpha'_1 y\alpha'_2 \doteq y\beta'$ with $y \in qv(E')$ and $x \notin qv(E')$. It follows from the definitions of $\Rightarrow_L$ and $\Rightarrow_R$ that there is no E'' such that $E' \Rightarrow_R E''$, and exactly one E'' such that $E' \Rightarrow_L E''$. Thus the out-degree is one as claimed. Now consider the in-degree and let $E'' \in [E]_{\Rightarrow}$ such that $E'' \Rightarrow E'$. Note that by the definition of $\Rightarrow_R$, we cannot have that $E'' \Rightarrow_R E'$, so we must have that $E'' \Rightarrow_L E'$. It follows from the fact that the Nielsen transformation morphisms $\psi_{y<x}$ are injective that there is exactly one such E'', and thus we also have that the in-degree of E' is one as claimed. □

Claim 4.7.2 Let $E' \in S_2$. Then there exists $k \leq |E| - 2$ and $E_0, E_1, \ldots, E_{k+1} \in [E]_{\Rightarrow}$ and $Z \in \{L, R\}$ such that all the following statements hold:

1. $E_0, E_{k+1} \in S_1$,
2. $E_i \in S_2$ for $1 \leq i \leq k$,
3. $E_i \Rightarrow_Z E_{i+1}$ for $0 \leq i \leq k$,
4. there exists i, $1 \leq i \leq k$ such that $E' = E_i$.

Proof W.l.o.g. suppose that the RHS of E' has a prefix contained in $qv(E')$. Then since $\mathrm{Card}(qv(E')) \geq 2$ and since E' is regular, the LHS also contains at least one variable in $qv(E')$ and we may either write E' as

(1) $a_i a_{i+1} \ldots a_k x\alpha'_1 x' a_1 a_2 \ldots a_{i-1} y\alpha'_2 \doteq y\beta'$, or
(2) $a_i a_{i+1} \ldots a_k x a_1 a_2 \ldots a_{i-1} y\alpha'_2 \doteq y\beta'$

where $k \leq |E| - 2$, $a_j \in (X \backslash qv(E)) \cup \Sigma$ for $1 \leq j \leq k$, $x, x', y \in qv(E)$ with $x, x' \neq y$, and $\alpha'_1, \alpha'_2, \beta' \in (X \cup \Sigma)^*$. Consider the first case. Let E_0 be the equation given by

$$x' a_1 a_2 \ldots a_k x\alpha'_1 y\alpha'_2 \doteq y\beta',$$

let E_{k+1} be the equation given by

$$x\alpha'_1 x' a_1 a_2 \ldots a_k y\alpha'_2 \doteq y\beta',$$

and for $1 \leq j \leq k$, let E_j be the equation given by

$$a_j a_{j+1} \ldots a_k x\alpha'_1 x' a_1 a_2 \ldots a_{j-1} y\alpha'_2 \doteq y\beta'.$$

Then clearly, $E_i = E'$, $E_0, E_{k+1} \in S_1$, $E_j \in S_2$ for $1 \leq j \leq k$, and $E_j \Rightarrow_R E_{j+1}$ for $0 \leq j \leq k$ as claimed.

Now consider the second case. Let $E_0 = E_{k+1}$ be the equation given by

$$x a_1 a_2 \ldots a_k y\alpha'_2 \doteq y\beta'$$

and for $1 \leq j \leq k$, let E_j be the equation given by

$$a_j a_{j+1} \dots a_k x a_1 a_2 \dots a_{j-1} y \alpha_2' \doteq y \beta'.$$

Then clearly, $E_i = E'$, $E_0, E_{k+1} \in S_1$, $E_j \in S_2$ for $1 \leq j \leq k$, and $E_j \Rightarrow_R E_{j+1}$ for $0 \leq j \leq k$ as claimed. □

Claims 4.7.1 and 4.7.2 are sufficient to show that the equations/vertices in S_1 are exactly those which survive in an isolated path compression of order $|E|$ of $\mathscr{G}^{\rightrightarrows}_{[E]}$. To state this more formally, we define a relation $\diamond$ on the equations in S_1 such that $E' \diamond E''$ if $E', E'' \in S_1$ and either $E' \Rightarrow E''$, or there exist $E_1, E_2, \dots E_k \in S_2$ and $Z \in \{L, R\}$ such that $E' \Rightarrow_Z E_1 \Rightarrow_Z E_2 \Rightarrow_Z \dots \Rightarrow_Z E_k \Rightarrow_Z E''$. Then we get the following.

Claim 4.7.3 The graph $\mathscr{G}^{\diamond}_{S_1}$ is an isolated path compression of order $|E|$ of $\mathscr{G}^{\rightrightarrows}_{[E]}$.

Proof Directly from Claims 4.7.1 and 4.7.2. □

It remains to show that $\mathscr{G}^{\diamond}_{S_1}$ is isomorphic to $\mathscr{G}^{\rightrightarrows}_{[\hat{E}]}$ where $\hat{E}$ is given by $\pi_{qv(E)}(\alpha) \doteq \pi_{qv(E)}(\beta)$. In other words, we must show that there is an isomorphism $f : S_1 \to [\hat{E}]_{\Rightarrow}$ such that for any $E', E'' \in S_1$, $f(E_1) \Rightarrow F(E_2)$ if and only if $E_1 \diamond E_2$. Before we can define f, we must firstly show that there exists $\tilde{E} \in S_1$ given by $\tilde{\alpha} \doteq \tilde{\beta}$ such that $\pi_{qv(\tilde{E})}(\tilde{\alpha}) = \pi_{qv(E)}(\alpha)$ and $\pi_{qv(\tilde{E})}(\tilde{\beta}) = \pi_{qv(E)}(\beta)$. If $E \in S_1$ then we may simply take $\tilde{E} = E$. Otherwise, $E \in S_2$, meaning exactly one of $\alpha[1], \beta[1]$ is in $qv(E)$. W.l.o.g. suppose that $\alpha[1] \notin qv(E)$. Then we may write $\alpha = \gamma x \alpha_1 y \alpha_2$ and $\beta = y \beta_1$ where $\gamma \in ((X \backslash qv(E)) \cup \Sigma)^+$, $x, y \in qv(E)$, and $\alpha_1, \alpha_2, \beta_1 \in (X \cup \Sigma)^*$. Furthermore, we have $E \Rightarrow_R^* \tilde{E}$ where $\tilde{E} \in S_1$ is given by $x \alpha_1 \gamma_2 y \alpha_2 \doteq y \beta_1$, in which case we have that $\pi_{qv(E)}(\alpha) = \pi_{qv(\tilde{E})}(x \alpha_1 \gamma_2 y \alpha_2)$ and $\pi_{qv(E)}(\beta) = \pi_{qv(\tilde{E})}(y \beta_1)$ (note that we have that $qv(E) = qv(\tilde{E})$ since $\tilde{E} \in [E]_{\Rightarrow}$).

Since $\tilde{E} \in S_1$, we may write $\tilde{E}$ as

$$y_1 \gamma_1 y_2 \gamma_2 \dots y_n \gamma_n \doteq y_1' \delta_1 y_2' \delta_2 \dots y_n' \delta_n$$

where $y_i, y_i' \in qv(\tilde{E})$ and $\gamma_i, \delta_i \in ((X \backslash qv(\tilde{E})) \cup \Sigma)^*$ for $1 \leq i \leq n$. Consequently, by our assumptions about $\tilde{E}$, it follows that $\hat{E}$ may be written as $y_1 y_2, \dots y_n \doteq y_1' y_2' \dots y_n'$. With this information, we are now ready to define our isomorphism $f : S_1 \to [\hat{E}]_{\Rightarrow}$ via two morphisms σ_{LHS} and σ_{RHS}. In particular, let $\sigma_{LHS} : qv(\tilde{E})^* \to (X \cup \Sigma)^*$ be the morphism such that $\sigma_{LHS}(y_i) = y_i \gamma_i$ for $1 \leq i \leq n$ and $\sigma_{RHS} : qv(\tilde{E})^* \to (X \cup \Sigma)^*$ be the morphism such that $\sigma_{RHS}(y_j') = y_j' \delta_i$ for $1 \leq j \leq n$. Then we define f such that $f(\alpha' \doteq \beta')$ is $\sigma_{LHS}(\alpha') \doteq \sigma_{RHS}(\beta')$ for all $\alpha' \doteq \beta' \in S_1$. In order to show that f is indeed an isomorphism with the desired property that $f(E_1) \Rightarrow F(E_2)$ if and only if $E_1 \diamond E_2$, we need the following claim.

Claim 4.7.4 Let $\hat{\alpha_1}, \hat{\alpha_2}, \hat{\beta_1}, \hat{\beta_2} \in qv(E)^*$ such that $\hat{\alpha_1} \doteq \hat{\beta_1} \in [\hat{E}]_{\Rightarrow}$, and $\sigma_{LHS}(\hat{\alpha_1}) \doteq \sigma_{RHS}(\hat{\beta_1}) \in [E]_{\Rightarrow}$. Then $\hat{\alpha_1} \doteq \hat{\beta_1} \Rightarrow \hat{\alpha_2} \doteq \hat{\beta_2}$ if and only if $\sigma_{LHS}(\hat{\alpha_1}) \doteq \sigma_{RHS}(\hat{\beta_1}) \diamond \sigma_{LHS}(\hat{\alpha_2}) \doteq \sigma_{RHS}(\hat{\beta_2})$.

Proof Suppose firstly that $\hat{\alpha_1} \doteq \hat{\beta_1} \Rightarrow \hat{\alpha_2} \doteq \hat{\beta_2}$, and w.l.o.g. suppose that $\hat{\alpha_1} \doteq \hat{\beta_1} \Rightarrow_L \hat{\alpha_2} \doteq \hat{\beta_2}$. Then there exist $z_1, z_2, \ldots, z_n \in qv(E)$, $\mu \in qv(E)^*$ such that $\hat{\alpha_1} = z_i\mu$ for some i, $1 \leq i \leq n$, $\hat{\beta_1} = z_1 z_2 \ldots z_n$, $\hat{\alpha_2} = \hat{\alpha_1}$ and $\hat{\beta_2} = z_2 \ldots z_{i-1} z_1 z_i \ldots z_n$. Let $a_1, a_2, \ldots a_k \in (X \backslash qv(E)) \cup \Sigma$ such that $\sigma_{RHS}(z_1) = z_1 a_1 a_2 \ldots a_k$. Let E_0 be given by $\sigma_{LHS}(z_i\mu) \doteq \sigma_{RHS}(z_1 z_2 \ldots z_n)$, and for $1 \leq j \leq k$, let E_j be given by $\sigma_{LHS}(z_i\mu) \doteq a_j a_{j+1} \ldots a_k \sigma_{RHS}(z_2 \ldots z_{i-1}) z_1 a_1 a_2 \ldots a_{j-1} \sigma_{RHS}(z_i \ldots z_n)$, and let E_{k+1} be given by $\sigma_{LHS}(z_i\mu) \doteq \sigma_{RHS}(z_2 \ldots z_{i-1}) z_1 a_1 a_2 \ldots a_k \sigma_{RHS}(z_i \ldots z_n)$. Then we have $E_0 \Rightarrow_L E_1 \Rightarrow_L \ldots \Rightarrow_L E_{k+1}$. Moreover, we have that $E_0 \in S_1$ is given by $\sigma_{LHS}(\hat{\alpha_1}) \doteq \sigma_{RHS}(\hat{\beta_1})$, $E_{k+1} \in S_1$ is given by $\sigma_{LHS}(\hat{\alpha_2}) \doteq \sigma_{RHS}(\hat{\beta_2})$, and $E_j \in S_2$ for $1 \leq j \leq k$ so $E_0 \diamond E_{k+1}$ as required.

Now suppose that $\sigma_{LHS}(\hat{\alpha_1}) \doteq \sigma_{RHS}(\hat{\beta_1}) \diamond \sigma_{LHS}(\hat{\alpha_2}) \doteq \sigma_{RHS}(\hat{\beta_2})$. Then by the definition of $\diamond$, there exist $E_0, E_1, \ldots, E_{k+1} \in [E]_{\Rightarrow}$ such that $E_0 \in S_1$ is given by $\sigma_{LHS}(\hat{\alpha_1}) \doteq \sigma_{RHS}(\hat{\beta_1})$, $E_{k+1} \in S_1$ is given by $\sigma_{LHS}(\hat{\alpha_2}) \doteq \sigma_{RHS}(\hat{\beta_2})$, $E_0 \Rightarrow_Z E_1 \Rightarrow_Z \ldots \Rightarrow_Z E_{k+1}$ for some $Z \in \{L, R\}$, and $E_j \in S_2$ for $1 \leq j \leq k$.

W.l.o.g. suppose that $Z = L$. Then there exist $z_1, z_2, \ldots, z_n \in qv(E)$, $\mu \in qv(E)^*$, and $a_1, a_2, \ldots, a_\ell \in (X \backslash qv(E)) \cup \Sigma$ such that $\hat{\alpha_1} = z_i\mu$ for some i, $1 \leq i \leq n$, $\hat{\beta} = z_1 z_2 \ldots z_n$, and $\sigma_{RHS}(z_1) = z_1 a_1 a_2 \ldots a_\ell$. Hence E_0 can be written as

$$\sigma_{LHS}(z_i\mu) \doteq z_1 a_1 a_2 \ldots a_\ell \sigma_{RHS}(z_2 z_3 \ldots z_n).$$

Moreover, we have that $E_0 \Rightarrow_L E'_1 \Rightarrow_L E'_2 \Rightarrow_L \ldots \Rightarrow E'_\ell \Rightarrow_L E'_{\ell+1}$ where E'_j is given by

$$\sigma_{LHS}(z_i\mu) \doteq a_j a_{j+1} \ldots a_\ell \sigma_{RHS}(z_2 \ldots z_{i-1}) z_1 a_1 \ldots a_{j-1} \sigma_{RHS}(z_i z_{i+1} \ldots z_n)$$

for $1 \leq j \leq k$, and $E'_{\ell+1}$ is given by

$$\sigma_{LHS}(z_i\mu) \doteq \sigma_{RHS}(z_2 \ldots z_{i-1}) z_1 a_1 \ldots a_\ell \sigma_{RHS}(z_i z_{i+1} \ldots z_n).$$

Note that $E_{\ell+1}$ may also be written

$$\sigma_{LHS}(z_i\mu) \doteq \sigma_{RHS}(z_2 z_3 \ldots z_{i-1} z_1 z_i z_{i+1} \ldots z_n).$$

Now, since $\Rightarrow_L$ is deterministic, and since $E_{\ell+1}, E_{k+1} \in S_1$ while $E'_{j_1}, E_{j_2} \in S_2$ for each j_1, $1 \leq j_1 \leq \ell$ and $j_2, 1 \leq j_2 \leq k$, we must necessarily have that $k = \ell$. Since σ_{LHS} and σ_{RHS} are injective, we must have $\hat{\alpha_2} = \hat{\alpha_1}$ and $\hat{\beta_2} = z_2 z_3 \ldots z_{i-1} z_1 z_i z_{i+1} \ldots z_n$. It follows from the definitions that $\hat{\alpha_1} \doteq \hat{\beta_1} \Rightarrow_L \hat{\alpha_2} \doteq \hat{\beta_2}$. □

It follows from Claim 4.7.4 by a simple induction with $\tilde{E}$ as the base case that $S_1 = \{\sigma_{LHS}(\hat{\alpha'}) \doteq \sigma_{RHS}(\hat{\beta'}) \mid \hat{\alpha'} \doteq \hat{\beta'} \in [\hat{E}]_{\Rightarrow}\}$, or equivalently that $f(S_1) = [\hat{E}]_{\Rightarrow}$. The claim also states explicitly that $\sigma_{LHS}(\hat{\alpha'}) \doteq \sigma_{RHS}(\hat{\beta'}) \diamond \sigma_{LHS}(\hat{\alpha''}) \doteq \sigma_{RHS}(\hat{\beta''})$ if and only if $\hat{\alpha'} \doteq \hat{\beta'} \Rightarrow \hat{\alpha''} \doteq \hat{\beta''}$ and thus f is an isomorphism such that $f(E_1) \Rightarrow f(E_2)$ if and only if $E_1 \diamond E_2$ for all $E_1, E_2 \in S_1$. We may therefore conclude that $\mathscr{G}^{\diamond}_{S_1}$ is indeed isomorphic to $\mathscr{G}^{\Rightarrow}_{[\hat{E}]}$ as required. □

Combining Corollary 4.4 and Lemma 4.7, it is now possible to formulate the main result of this section, describing the graphs $\mathscr{G}^{\rightrightarrows}_{[E]}$ for arbitrary RWEs E in terms of graphs $\mathscr{G}^{\rightrightarrows}_{[E']}$ for basic RWEs E'. An example of the theorem is given in Fig. 3.

Theorem 4.8 *Let E be a RWE given by $\alpha \doteq \beta$. Let α', β' be the shortest non-empty prefixes of α, β respectively such that $var(\alpha') \cap qv(E) = var(\beta') \cap qv(E)$. Let E' be the equation given by $\pi_{qv(E)}(\alpha') \doteq \pi_{qv(E)}(\beta')$. Then E' is basic, and $\mathscr{G}^{\rightrightarrows}_{[E']}$ is isomorphic to an isolated path compression of order $|E|$ of $\mathscr{G}^{\rightrightarrows}_{[E]}$.*

Proof Let $S = qv(\alpha' \doteq \beta')$. Firstly, we shall show that $\alpha' \doteq \beta'$ is indecomposable. Suppose for contradiction that $\alpha' \doteq \beta'$ is decomposable. Then there exist proper prefixes α'', β'' of α' and β' respectively such that $var(\alpha'') \cap S = var(\beta'') \cap S$. Then α'' and β'' are proper prefixes of α and β, and since they are shorter than α' and β', by our assumptions about α' and β', we cannot have that $var(\alpha'') \cap qv(E) = var(\beta'') \cap qv(E)$. Consequently, either there exists $x \in var(\alpha'') \cap qv(E)$ such that $x \notin var(\beta'') \cap qv(E)$ or there exists $x \in var(\beta'') \cap qv(E)$ such that $x \notin var(\alpha'') \cap qv(E)$. W.l.o.g. suppose the former is true. Then $x \notin var(\beta'')$, but since $x \in qv(E)$, it follows from $var(\alpha') \cap qv(E) = var(\beta') \cap qv(E)$ that $x \in var(\beta')$. However, this implies that $x \in S$, and since $x \in var(\alpha'')$ but $x \notin var(\beta'')$, we arrive at a contradiction to our assumption that $var(\alpha'') \cap S = var(\beta'') \cap S$.

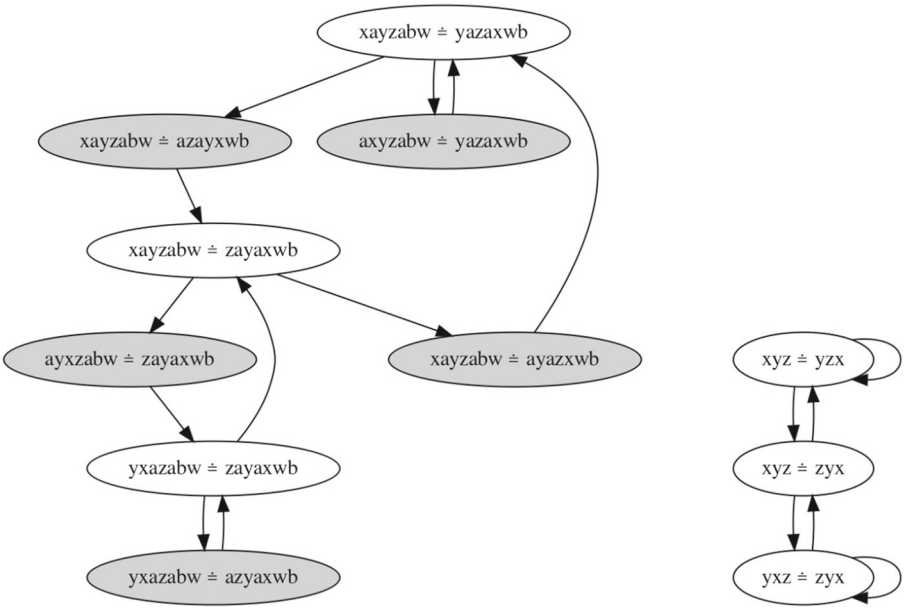

Fig. 3 An example of Theorem 4.8. On the left is the graph $\mathscr{G}^{\rightrightarrows}_{[E]}$ in the case that E is given by $x\mathtt{a}yz\mathtt{ab}w \doteq y\mathtt{a}z\mathtt{a}xw$ with variables x, y, z, w and terminal symbols $\mathtt{a}, \mathtt{b}$. On the right is $\mathscr{G}^{\rightrightarrows}_{[E']}$ for the corresponding basic equation E', which in this case is given by $xyz \doteq yzx$. The graph on the right is isomorphic to an isolated path compression of order 2 of the graph on the right. Vertices internal to the isolated paths (i.e. those which are removed by the compression are shown in grey

Now, let E'' be the equation given by $\pi_S(\alpha') \doteq \pi_S(\beta')$. By the assumption that $var(\alpha') \cap qv(E) = var(\beta') \cap qv(E)$, there is no variable $x \in qv(E) \backslash S$ occurring in α' or β'. Consequently, $E'' = E'$, and by Lemma 4.7, we have that $\mathscr{G}^{\Rightarrow}_{[E']}$ is isomorphic to an isolated path compression of order $|E|$ of $\mathscr{G}^{\Rightarrow}_{[\alpha' \doteq \beta']}$, which by Corollary 4.4 is isomorphic to $\mathscr{G}^{\Rightarrow}_{[E]}$. □

5 A Useful Invariant

When reasoning about the graphs $\mathscr{G}^{\Rightarrow}_{[E]}$, we need a way to help determine whether or not, for two equations E_1, E_2, we have $E_1 \Rightarrow^* E_2$. Showing the positive case that $E_1 \Rightarrow^* E_2$ can be achieved by simply finding an appropriate sequence of length-preserving Nielsen transformations from E_1 to E_2. However, showing that $E_1 \not\Rightarrow^* E_2$ presents more of a challenge: the naive way would be to enumerate all vertices in $\mathscr{G}^{\Rightarrow}_{[E_1]}$ and show that E_2 is not among them. However, this is not suitable for abstract reasoning, and, even in concrete cases, is inelegant and time-consuming.

The contribution of this section is a property of basic RWEs, defined as Υ_E below, which is preserved under the relation $\Rightarrow$ and thus provides a concise and more general means for showing that $E_1 \not\Rightarrow^* E_2$. It is an indispensable component of the proofs of our main results.

Definition 5.1 (The invariant Υ_E) Let E be a basic RWE such that $\text{Card}(var(E)) > 1$. Let # be a new symbol not in X. Then we may write E as $x\alpha_1 y\alpha_2 \doteq y\beta_1 x\beta_2$ with $x, y \in X$ and $\alpha_1, \alpha_2, \beta_1, \beta_2 \in (X \backslash \{x, y\})^*$. Let $\mathscr{Z}_E = var(\alpha_1\alpha_2\beta_1\beta_2) \cup \{\#\}$. Let the function $Q_E : \mathscr{Z}_E \to X^2$ be defined as follows: for each $z \in \mathscr{Z}_E \backslash \{\#\}$, let $Q_E(z) = (u, v)$ where uz is a factor of $x\alpha_1 y\alpha_2$ and vz is a factor of $y\beta_1 x\beta_2$. Let $Q_E(\#) = (u, v)$ where uy is a factor of $x\alpha_1 y\alpha_2$ and vx is a factor of $y\beta_1 x\beta_2$. Let $\Upsilon_E = \{Q_E(z) \mid z \in \mathscr{Z}_E\}$. If $\text{Card}(var(E)) \leq 1$, then $\Upsilon_E = \emptyset$.

Intuitively, given a basic RWE E of the form $\alpha \doteq \beta$, we construct Υ_E by taking, for each variable $x \in var(E)$, the pair (u, v) of predecessors of x in E, i.e. such that ux is a factor of α and vx is a factor of β. It follows directly from the definition of basic RWEs that this pair is unique, and it exists whenever x is not the leftmost variable in either α or β. The special case that x is the leftmost variable of α or β is handled by the special symbol #. The following observations follow directly from the definitions, but are central to the use of Υ_E in later proofs.

Remark 5.2 Let E be a basic regular word equation given by $\alpha y \doteq \beta x$ with $x, y \in X$ and $\alpha, \beta \in X^*$. Then for each $z \in var(\alpha)$, there is exactly one element $(u, v) \in \Upsilon_E$ such that $u = z$. For each $z \notin var(\alpha)$, there is no element $(u, v) \in \Upsilon_E$ such that $u = z$. Similarly, for each $w \in var(\beta)$, there is exactly one element $(u, v) \in \Upsilon_E$ such that $v = w$ and for each $w \notin var(\beta)$, there is no element $(u, v) \in \Upsilon_E$ such that $v = w$.

The usefulness of Υ_E as a property of basic RWEs arises from the fact that it is invariant under the length-preserving Nielsen transformations. Consequently for a given basic RWE E, we can use the set $\{E' \mid \Upsilon_{E'} = \Upsilon_E\}$ as an over-approximation of the set $[E]_{\Rightarrow}$.

Theorem 5.3 *Let E_1, E_2 be basic RWEs such that $E_1 \Rightarrow^* E_2$. Then $\Upsilon_{E_1} = \Upsilon_{E_2}$.*

Proof It is sufficient to prove the same statement for the case that $E_1 \Rightarrow E_2$. W.l.o.g. we may assume that $E_1 \Rightarrow_L E_2$. The case that $E_1 \Rightarrow_R E_2$ is symmetric. Moreover, if $E_1 = E_2$, then the statement holds trivially, thus we may assume that $E_1 \neq E_2$. The statement trivially holds for equations of the form $xy \doteq yx$, since $[xy \doteq yx]_{\Rightarrow} = \{xy \doteq yx\}$. Otherwise, taking into account the fact that E_1 and E_2 are basic and therefore indecomposable, we have two cases: we may write E_1 and E_2 as either

1. $x\alpha_1 w\alpha_2 y\alpha_3 \doteq yw\beta_1 x\beta_2$ and $x\alpha_1 w\alpha_2 y\alpha_3 \doteq w\beta_1 yx\beta_2$, or
2. $x\alpha_1 y\alpha_2 w\alpha_3 \doteq yw\beta_1 x\beta_2$ and $x\alpha_1 y\alpha_2 w\alpha_3 \doteq w\beta_1 yx\beta_2$

respectively, where $w, x, y \in X$ with $x \neq y$ and $\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2 \in (X \backslash \{x, y, w\})^*$ such that $var(\alpha_1\alpha_2\alpha_3) = var(\beta_1\beta_2)$.

Suppose that we have the first case, then $\mathcal{Z}_{E_1} = var(\alpha_1\alpha_2\alpha_3) \cup \{\#, w\}$ and $\mathcal{Z}_{E_2} = var(\alpha_1\alpha_2\alpha_3) \cup \{\#, y\}$. Moreover, for each $z \in var(\alpha_1\alpha_2\alpha_3)$, there exist $u, v \in X$ such that uz (resp. vz) is a factor of the LHS (resp. RHS) of both E_1 and E_2, so $Q_{E_1}(z) = Q_{E_2}(z)$. Now, let a, b, c be the rightmost variables in $x\alpha_1$, $w\alpha_2$ and $w\beta_1$ respectively (i.e. their length-1 suffixes). Then we have that $Q_{E_1}(w) = (a, y)$, $Q_{E_1}(\#) = (b, c)$, $Q_{E_2}(y) = (b, c)$, and $Q_{E_2}(\#) = (a, y)$. Thus $\Upsilon_{E_1} = \Upsilon_{E_2}$.

Now suppose instead that we have the second case. Similarly to the first case, we have that $\mathcal{Z}_{E_1} = var(\alpha_1\alpha_2\alpha_3) \cup \{\#, w\}$, $\mathcal{Z}_{E_2} = var(\alpha_1\alpha_2\alpha_3) \cup \{\#, y\}$ and for each $z \in var(\alpha_1\alpha_2\alpha_3)$, $Q_{E_1}(z) = Q_{E_2}(z)$. Now, let a, b, c be the rightmost variables in $x\alpha_1$, $w\beta_1$ and $y\alpha_2$ respectively. Then we have that $Q_{E_1}(w) = (c, y)$, $Q_{E_1}(\#) = (a, b)$, $Q_{E_2}(y) = (c, y)$, and $Q_{E_2}(\#) = (a, b)$. Thus $\Upsilon_{E_1} = \Upsilon_{E_2}$ in both cases as required. □

As an example, let E_1 be the basic RWE given by $xuzwy \doteq ywuxz$. Then $\mathcal{Z}_{E_1} = \{u, z, w, \#\}$ and Q_{E_1} is the function with $Q_{E_1}(u) = (x, w)$, $Q_{E_1}(z) = (u, x)$, $Q_{E_1}(w) = (z, y)$ and $Q_{E_1}(\#) = (w, u)$. Thus, $\Upsilon_{E_1} = \{(w, u), (x, w), (u, x), (z, y)\}$. Similarly, if E_2 is the basic RWE given by $xuwzy \doteq yuxwz$, then $\Upsilon_{E_2} = \{(x, y), (u, x), (w, w), (z, u)\}$. Consequently, we may conclude that $E_1 \not\Rightarrow^* E_2$ (and symmetrically that $E_2 \not\Rightarrow^* E_1$).

Since the invariant Υ_E provides a necessary condition on when two basic RWEs belong to the same equivalence class under $\Rightarrow^*$, we might also ask whether it is also sufficient, and hence characteristic. However, this is not the case. For instance, if E_3 is given by $xuvwy \doteq ywvux$ and E_4 is given by $xwvuy \doteq yuvwx$, then $\Upsilon_{E_3} = \Upsilon_{E_4} = \{(x, v), (u, w), (v, y), (w, u)\}$ but it can be verified (e.g. by enumerating $[E_3]_{\Rightarrow}$ and $[E_4]_{\Rightarrow}$) that $E_3 \not\Rightarrow^* E_4$.

6 Jumbled Equations and a Special Case of Symmetry

The invariant property Υ_E introduced in the Section 5 consists of pairs of variables. The case that $(x, x) \in \Upsilon_E$ for some $x \in var(E)$ is special in the sense that it leads to a particular repetitive structure in the graph $\mathcal{G}^{\Rightarrow}_{[E]}$, described in the current section. We shall call basic RWEs E for which no pair of the form (x, x) occurs in Υ_E jumbled.

Definition 6.1 (Jumbled Equations and $\Delta(E)$) Let E be a basic RWE and let $\Delta(E) = \{x \in var(E) \mid (x, x) \in \Upsilon_E\}$. If $\mathrm{Card}(\Delta(E)) = 0$, then E is *jumbled*.

For example, if we consider the equation E given by $xyzw \doteq wyzx$, then $\Upsilon_E = \{(x, w), (y, y), (z, z)\}$ so $\Delta(E) = \{y, z\}$ and E is not jumbled. On the other hand, for E' given by $xyzw \doteq wzyx$, we have $\Upsilon_{E'} = \{(x, z), (y, w), (z, y)\}$, so $\Delta(E') = \emptyset$ and E' is jumbled.

Note that since Υ_E is invariant under $\Rightarrow^*$, so is the property of being jumbled. Furthermore, it follows from the definitions that $(x, x) \in \Upsilon_E$ for some basic RWE E and $x \in X$ if and only if there exists $y \in X$ such that one of the following holds:[5]

1. xy occurs as a factor of both the LHS and RHS of E, or
2. there exists E' with $E \Rightarrow E'$ such that xy occurs as a factor of both the LHS and RHS of E'.

The cardinality of $\Delta(E)$ can be interpreted as a measure of the similarity of the two sides of the equation. If $\mathrm{Card}(\Delta(E))$ is large in comparison to $\mathrm{Card}(E)$, then the orders in which the variables occur on the LHS and RHS of E will be similar. On the other hand, when $\Delta(E) = \emptyset$, there will be no common order in the variables on each side, and hence the equation is 'jumbled'. In general, we may observe the following bounds on $\mathrm{Card}(\Delta(E))$ as follows.

Remark 6.2 Let E be a basic RWE. It follows directly from Definition 5.1 that if $\mathrm{Card}(var(E)) < 2$, then $\mathrm{Card}(\Delta(E)) = 0$. Otherwise, E can be written as $\alpha x \doteq \beta y$ for some $x, y \in X$, $\alpha \in (X\backslash\{x\})^*$ and $\beta \in (X\backslash\{y\})^*$. Since E is basic, it is indecomposable, so we may additionally conclude that $x \neq y$. By Remark 5.2, neither (x, x) nor (y, y) can be contained in Υ_E, so we must have $\mathrm{Card}(\Delta(E)) \leq \mathrm{Card}(var(E)) - 2$.

The rest of this section is devoted to describing the structure of the graphs $\mathscr{G}^{\Rightarrow}_{[E]}$ in the general case in terms of the graphs $\mathscr{G}^{\Rightarrow}_{[E']}$ where E' is jumbled. The first step is to notice that we can easily transform any basic RWE E into one which is jumbled by simply removing all variables x such that $(x, x) \in \Delta(E)$.

Lemma 6.3 *Let E be a basic RWE given by $\alpha \doteq \beta$ and let $Y = var(E)\backslash\Delta(E)$. Then the equation E_Y given by $\pi_Y(\alpha) \doteq \pi_Y(\beta)$ is a jumbled basic RWE.*

Proof If $\Delta(E) = \emptyset$, then the lemma holds trivially. Assume that $\Delta(E) \neq \emptyset$. We shall prove the following statement, from which the lemma follows by a simple induction.

Claim 6.3.1 Suppose that E is a basic RWE given by $\alpha \doteq \beta$, and that $x \in \Delta(E)$. Let E' be the equation $\pi_{var(E)\backslash\{x\}}(\alpha) \doteq \pi_{var(E)\backslash\{x\}}(\beta)$. Then E' is a basic RWE and $\Upsilon_{E'} = \Upsilon_E\backslash\{(x, x)\}$.

[5]The first case corresponds to the possibility that $Q_E(y) = (x, x)$ for some variable y. The second case corresponds to the possibility that $Q_E(\#) = (x, x)$, meaning that E has the form $y\alpha_1 xz\alpha_2 \doteq z\beta_1 xy\beta_2$, with $x, y, z \in X$ and $\alpha_1, \alpha_2, \beta_1, \beta_2 \in X^*$, in which case $E \Rightarrow \alpha_1 xyz\alpha_2 \doteq z\beta_1 xy\beta_2$.

Proof Let $Q_E, \mathcal{Z}_E$ be defined as per Definition 5.1. We shall consider two cases depending on whether $Q_E(\#) = (x, x)$. Suppose firstly that $Q_E(\#) \neq (x, x)$. Then there exist $\alpha_1, \alpha_2, \beta_1, \beta_2$ such that $\alpha = \alpha_1 xy\alpha_2$, $\beta = \beta_1 xy\beta_2$, $\pi_{var(E)\backslash\{x\}}(\alpha) = \alpha_1 y\alpha_2$ and $\pi_{var(E)\backslash\{x\}}(\beta) = \beta_1 y\beta_2$. Suppose for contradiction that E' is not basic. Clearly both sides of E' belong to $qv(E')$, so we may infer that E' is decomposable, and thus that there exist proper prefixes α' and β' of $\alpha_1 y\alpha_2$ and $\beta_1 y\beta_2$ respectively such that $var(\alpha') \cap qv(E') = var(\beta') \cap qv(E')$. Clearly, either y occurs in both α' and β', or in neither. Let $\tau : var(E')^* \to var(E)^*$ be the morphism such that $\tau(y) = xy$ and $\tau(z) = z$ for $z \in var(E')\backslash\{y\}$. Then $\alpha'' = \tau(\alpha')$ and $\beta'' = \tau(\beta')$ are proper prefixes of α and β respectively which satisfy $var(\alpha'') \cap qv(E) = var(\beta'') \cap qv(E)$. Thus E is decomposable and therefore not basic, a contradiction.

To see that $\Upsilon_{E'} = \Upsilon_E\backslash\{(x, x)\}$, suppose firstly that x is not a prefix of α or β, and thus that $\alpha_1 \neq \varepsilon$ and $\beta_1 \neq \varepsilon$. Then $\mathcal{Z}_E = (var(E)\backslash\{\alpha_1[1], \beta_1[1]\}) \cup \{\#\}$, and $\mathcal{Z}_{E'} = (var(E)\backslash\{\alpha_1[1], \beta_1[1], x\}) \cup \{\#\}$. It follows from the definitions that $Q_{E'}(y) = Q_E(x) = (\alpha_1[|\alpha_1|], \beta_1[|\beta_1|])$. Since $\alpha_1, \beta_1 \neq \varepsilon$, $\alpha_1[1] \notin \{x, y\}$ and $\beta_1[1] \notin \{x, y\}$. Consequently there exist $u_\#, v_\# \in var(E)\backslash\{x\}$ such that $u_\#\alpha_1[1]$ is a factor of both α and $\pi_{var(E)\backslash\{x\}}(\alpha)$ and such that $v_\#\beta_1[1]$ is a factor of both β and $\pi_{var(E)\backslash\{x\}}(\beta)$. It follows that $Q_E(\#) = Q_{E'}(\#) = (u_\#, v_\#)$. Likewise, for any $z \notin \{x, y, \alpha_1[1], \beta_1[1]\}$, there exist $u, v \in var(E)\backslash\{x\}$ such that uz is a factor of both α and $\pi_{var(E)\backslash\{x\}}(\alpha)$ and such that vz is a factor of both β and $\pi_{var(E)\backslash\{x\}}(\beta)$. It follows that $Q_E(z) = Q_{E'}(z) = (u, v)$. Thus we may conclude that $\Upsilon_{E'} = \Upsilon_E\backslash\{(x, x)\}$.

Next, suppose that $\alpha_1 = \varepsilon$ and $\beta_1 \neq \varepsilon$ (the case that $\beta_1 = \varepsilon$ and $\alpha_1 \neq \varepsilon$ is symmetric). Then $\mathcal{Z}_E = (var(E)\backslash\{x, \beta_1[1]\}) \cup \{\#\}$ and $\mathcal{Z}_{E'} = (var(E)\backslash\{y, x, \beta_1[1]\}) \cup \{\#\}$. Then $Q_E(\#) = (u_\#, \beta_1[|\beta_1|])$ where $u_\#\beta_1[1]$ is a factor of $xy\alpha_2$. Since E is regular, each variable occurs once per side, so we may infer that $\beta_1[1] \neq y$, and hence that $u_\# \neq x$. It follows that $u_\#\beta_1[1]$ is also a factor of $y\alpha_2$, so we may further conclude that $Q_{E'}(\#) = (u_\#, \beta_1[|\beta_1|]) = Q_E(\#)$. Note that $Q_E(y) = (x, x)$. Let $z \in var(E)\backslash\{x, y, \beta_1[1]\}$. Then there exist $u, v \in var(E)\backslash\{x\}$ such that uz is a factor of both $xy\alpha_2$ and $y\alpha_2$, and such that vz is a factor of both $\beta_1 xy\beta_2$ and $\beta_1 y\beta_2$. It follows that $Q_E(z) = Q_{E'}(z) = (u, v)$. Again we have $\Upsilon_{E'} = \Upsilon_E\backslash\{(x, x)\}$. Finally, note that if $\alpha_1 = \beta_1 = \varepsilon$, then E is decomposable, which is a contradiction to the assumption that E is basic.

It remains to consider the case that $Q_E(\#) = (x, x)$. This implies that there exist $u, v \in var(E)\backslash\{x\}$ and $\alpha_1, \alpha_2, \beta_1, \beta_2 \in var(E)^*$ such that $\alpha = u\alpha_1 xv\alpha_2$, $\beta_2 = v\beta_1 xu\beta_2$, meaning E' is given by $u\alpha_1 v\alpha_2 \doteq v\beta_1 u\beta_2$. Suppose for contradiction that E' is not basic. Then as in the previous case, it must be decomposable, and there exist proper prefixes α', β' of $u\alpha_1 v\alpha_2$ and $v\beta_1 u\beta_2$ respectively which satisfy $var(\alpha') \cap qv(E') = var(\beta') \cap qv(E')$. Then we must have that $\alpha' = u\alpha_1 v\alpha_3$ and $\beta' = v\beta_1 u\beta_3$ for some $\alpha_3, \beta_3 \in X^*$. However, it follows that $\alpha'' = u\alpha_1 xv\alpha_3$ and $\beta'' = v\beta_1 xu\beta_3$ are proper prefixes of α and β satisfying $var(\alpha'') \cap qv(E) = var(\beta'') \cap qv(E)$, so E is decomposable which is a contradiction to the assumption that E is basic.

To see that $\Upsilon_{E'} = \Upsilon_E\backslash\{(x, x)\}$, note that in this case $\mathcal{Z}_E = (var(E)\backslash\{u, v\}) \cup \{\#\}$ and $\mathcal{Z}_{E'} = (var(E)\backslash\{u, v, x\}) \cup \#$. It follows from the definitions that $Q_{E'}(\#) = Q_E(x) = (w_1, w_2)$, where w_1 is the leftmost variable in $u\alpha_1$ and w_2 is the leftmost variable in $v\beta_1$. Moreover, for any $z \in var(E)\backslash\{u, v, x\}$, there exist $w_1', w_2' \in$

$var(E)\backslash\{x\}$ such that $w_1'z$ is a factor of both $u\alpha_1 x v\alpha_2$ and $u\alpha_1 v\alpha_2$, and such that $w_2'z$ is a factor of both $v\beta_1 x u\beta_2$ and $v\beta_1 u\beta_2$, meaning that $Q_{E'}(z) = Q_E(z) = (w_1', w_2')$. It follows that $\Upsilon_{E'} = \Upsilon_E\backslash\{(x, x)\}$ as required. □

We conclude the proof by noting that if $\Delta(E) = \{x_1, x_2, \ldots, x_k\}$, then there exist equations E_i for $0 \leq i \leq k$ given by $\alpha_i \doteq \beta_i$ such that

1. $E_0 = E$ and $E_k = E_Y$, and
2. for $1 \leq i \leq k$, $\alpha_i = \pi_{var(E_{i-1})\backslash\{x_i\}}(\alpha_{i-1})$ and $\beta_i = \pi_{var(E_{i-1})\backslash\{x_i\}}(\beta_{i-1})$.

Since E is basic, it follows by Claim 6.3.1 that E_i is basic for $1 \leq i \leq k$, and moreover by the same claim that $\Upsilon_{E_Y} = \Upsilon_E\backslash\{(x_i, x_i) \mid 1 \leq i \leq k\}$ meaning that E_Y is both basic and jumbled. □

There is a strong relation between the graph $\mathscr{G}^{\Rightarrow}_{[E]}$ for a basic RWE E and $\mathscr{G}^{\Rightarrow}_{[E_Y]}$ where E_Y is the jumbled basic RWE obtained from E by deleting the variables in $\Delta(E)$. The relation is described formally in Theorem 6.8. Before presenting the theorem, it is useful to first introduce some additional notions. Essentially, $\mathscr{G}^{\Rightarrow}_{[E]}$ is made up of approximate copies of $\mathscr{G}^{\Rightarrow}_{[E_Y]}$. Each copy is a subgraph $\mathscr{H}^E_\varphi$ of $\mathscr{G}^{\Rightarrow}_{[E]}$ which is associated with a certain morphism $\varphi : Y^* \to var(E)^*$ from a set Φ_E defined below. Intuitively, φ can be seen as a way of assigning variables in $\Delta(E)$ to variables in $Y = var(E)\backslash\Delta(E)$.

Definition 6.4 (The set Φ_E) Let E be a basic RWE. Let $Y = var(E)\backslash\Delta(E)$. Let Φ_E be the set of morphisms $\varphi : Y^* \to var(E)^*$ satisfying $\varphi(y) \in \Delta(E)^* y$ for all $y \in Y$, and $\sum_{y\in Y} |\varphi(y)|_x = 1$ for all $x \in \Delta(E)$.

The subgraphs $\mathscr{H}^E_\varphi$ are obtained by restricting $\mathscr{G}^{\Rightarrow}_{[E]}$ to subsets H^E_φ defined below. More precisely, $\mathscr{H}^E_\varphi$ consists of vertices H^E_φ and edges (E_1, E_2) whenever $E_1, E_2 \in H^E_\varphi$ and $E_1 \Rightarrow E_2$ (i.e. whenever (E_1, E_2) is an edge of $\mathscr{G}^{\Rightarrow}_{[E]}$). We shall say that $\mathscr{H}^E_\varphi$ is the subgraph of $\mathscr{G}^{\Rightarrow}_{[E]}$ *induced* by H^E_φ.

Definition 6.5 (V^E_φ, U^E_φ and H^E_φ) Let E be a basic RWE given by $\alpha \doteq \beta$ and let $Y = var(E)\backslash\Delta(E)$. Let E_Y be the equation $\pi_Y(\alpha) \doteq \pi_Y(\beta)$. Let $\varphi \in \Phi_E$. Then we define the sets V^E_φ, U^E_φ and H^E_φ as follows:

1. $V^E_\varphi = \{\varphi(\hat{\alpha}) \doteq \varphi(\hat{\beta}) \mid \hat{\alpha} \doteq \hat{\beta} \in [E_Y]_{\Rightarrow}\}$,
2. $H^E_\varphi = \{E' \mid \exists E'' \in V^E_\varphi, Z \in \{L, R\}.\ E'' \Rightarrow^*_Z E'\}$,
3. $U^E_\varphi = H^E_\varphi\backslash V^E_\varphi$.

For each $\varphi \in \Phi_E$, the subgraph $\mathscr{H}^E_\varphi$ is an approximate copy of $\mathscr{G}^{\Rightarrow}_{[E_Y]}$ in the sense that $\mathscr{G}^{\Rightarrow}_{[E_Y]}$ is isomorphic to an isolated path contraction of $\mathscr{H}^E_\varphi$. The intuition behind the sets V^E_φ and U^E_φ is that they provide a decomposition of the set H^E_φ of vertices of $\mathscr{H}^E_\varphi$ into those which survive after the isolated path compression (V^E_φ) and those

which are compressed/removed (U_φ^E). The underlying isomorphism is the function which maps equations $\hat{\alpha} \doteq \hat{\beta} \in [E_Y]_{\Rightarrow}$ to $\varphi(\hat{\alpha}) \doteq \varphi(\hat{\beta})$.

The structure of each subgraph $\mathscr{H}_\varphi^E$ is therefore essentially the same as the structure of $\mathscr{G}_{[E_Y]}^{\Rightarrow}$. In order to fully understand the structure of $\mathscr{G}_{[E]}^{\Rightarrow}$ however, we also need to know how the individual subgraphs are connected, or in other words, when two of subgraphs $\mathscr{H}_{\varphi_1}^E, \mathscr{H}_{\varphi_2}^E$ share a common vertex. We shall later see (Lemma 6.14) that $\mathscr{H}_{\varphi_1}^E$ and $\mathscr{H}_{\varphi_2}^E$ share a vertex if and only if the corresponding morphisms φ_1, φ_2 satisfy a 'closeness' condition defined as follows. See Fig. 4 for a complete example of the resulting relation.

Definition 6.6 (Close morphisms $\varphi_1, \varphi_2 \in \Phi_E$) Let E be a basic RWE and let $Y = var(E) \backslash \Delta(E)$. Let $\varphi_1, \varphi_2 \in \Phi_E$. Then φ_1, φ_2 are *close* if there exist $y_1, y_2 \in Y$ with $y_1 \neq y_2$ and $\gamma_1, \gamma_2 \in \Delta(E)^*$ such that:

1. For all $y \in Y \backslash \{y_1, y_2\}$, $\varphi_1(y) = \varphi_2(y)$, and
2. $\varphi_1(y_1) = \gamma_1\gamma_2 y_1$, $\varphi_2(y_1) = \gamma_2 y_1$, and $\varphi_2(y_2) = \gamma_1\varphi_1(y_2)$.

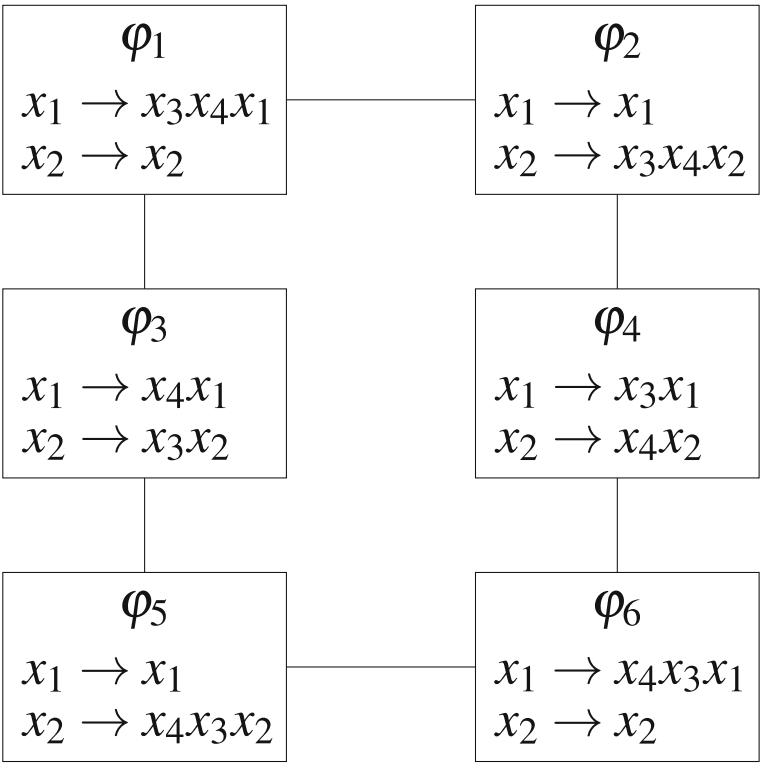

Fig. 4 A graph representing the closeness relation for morphisms in Φ_E for a basic RWE E with $var(E) = \{x_1, x_2, x_3, x_4\}$ and $\Delta(E) = \{x_3, x_4\}$, meaning that $Y = \{x_1, x_2\}$. In this case, Φ_E contains six morphisms, φ_i, $1 \leq i \leq 6$, which make up the vertices of the graph. Vertices connected by an edge are close in the sense of Definition 6.6

Informally, two morphisms $\varphi_1, \varphi_2 \in \Phi_E$ are close if we can obtain one from the other by removing some prefix of the image of a variable y_1 and appending it to the left of the image of another variable y_2. For example, suppose that $var(E) = \{x_1, x_2, x_3, x_4, x_5, x_6\}$ and $\Delta(E) = \{x_3, x_4, x_5, x_6\}$, and consider the two morphisms $\varphi_1, \varphi_2 : \{x_1, x_2\}^* \to \{x_1, x_2, x_3, x_4, x_5, x_6\}^*$ given by $\varphi_1(x_1) = x_4x_3x_5x_1$, $\varphi_1(x_2) = x_6x_2$, $\varphi_2(x_1) = x_5x_1$ and $\varphi_2(x_2) = x_4x_3x_6x_2$. Then φ_1, φ_2 both belong to Φ_E and are close, since we can get one from the other simply by moving the the prefix x_4x_3 from the image of x_1 to the image of x_2.

The following lemma shows that even when φ_1 and φ_2 are not close, we can find a sequence of intermediate morphisms in Φ_E starting with φ_1 and ending with φ_2, such that each morphism in the sequence and its successor are close, and such that this sequence is 'short'. This will form the basis of our claim that the subgraphs $\mathscr{H}_\varphi^E$ which make up the graph $\mathscr{G}_{[E]}^{\rightrightarrows}$ are well-connected, and in particular means that there is a (short) path in $\mathscr{G}_{[E]}^{\rightrightarrows}$ between any two of the subgraphs.

Lemma 6.7 *Let E be a basic RWE and suppose that $\varphi', \varphi'' \in \Phi_E$ with $\varphi' \neq \varphi''$. Then there exist $k \leq 4Card(\Delta(E)) + 1$ and $\varphi_1, \varphi_2, \varphi_3, \ldots, \varphi_k \in \Phi_E$ such that $\varphi' = \varphi_1$, $\varphi'' = \varphi_k$, and φ_i, φ_{i+1} are close for all i, $1 \leq i < k$.*

Proof Let $Y = var(E)\backslash\Delta(E)$. If $\Delta(E) = \emptyset$, then Φ_E contains only the identity morphism. Thus we may assume that $\Delta(E) \neq \emptyset$ and consequently by Remark 6.2 that $\text{Card}(Y) \geq 2$. Note the following claim.

Claim 6.7.1 Let $\varphi_1, \varphi_2 \in \Phi_E$, $y_1, y_2 \in Y$, $z \in \Delta(E)$ and $\gamma_1, \gamma_2 \in \Delta(E)^*$ such that $y_1 \neq y_2$ and

1. $\varphi_1(y_1) = \gamma_1 z \gamma_2 y_1$, $\varphi_2(y_1) = \gamma_1\gamma_2 y_1$ and $\varphi_2(y_2) = z\varphi_1(y_2)$, and
2. $\varphi_1(y) = \varphi_2(y)$ for all $y \in Y\backslash\{y_1, y_2\}$.

Then there exists $\varphi_3 \in \Phi_E$ such that φ_1, φ_3 are close, and φ_3, φ_2 are close.

Proof Let φ_3 be the morphism such that $\varphi_3(y_1) = \gamma_2 y_1$, $\varphi_3(y_2) = \gamma_1 z \varphi_1(y_2)$, and $\varphi_3(y) = \varphi_1(y)$ for all $y \in Y\backslash\{y_1, y_2\}$. Then it follows directly from the definitions that φ_1, φ_3 are close. Moreover, since $\varphi_2(y) = \varphi_1(y)$ for all $y \in Y\backslash\{y_1, y_2\}$, it also follows from the definitions that φ_2, φ_3 are also close. □

Claim 6.7.1 shows us that with two successors in a sequence, we can 'move' any variable $z \in \Delta(E)$ from $\varphi(y_1)$ to the prefix of $\varphi(y_2)$ where $y_1, y_2 \in Y$ with $y_1 \neq y_2$ (leaving the rest of the morphism unchanged). Given any $\varphi' \in \Phi_E$ we can reach any other morphism $\varphi'' \in \Phi_E$ by moving each variable $z \in \Delta(E)$ twice in this manner according to the following strategy: firstly, we move each variable $z \in \Delta(E)$ to the prefix of the image of a variable $y \in Y$ such that $z \notin var(\varphi''(y))$. Note that this is possible due to the assumption that $\text{Card}(Y) \geq 2$ and requires moving each variable in $\Delta(E)$ at most once. Then, we move the variables $z \in \Delta(E)$ back to the images of the 'correct' $y \in Y$ in the appropriate order. For example, if $\varphi''(y) = z_1 z_2 \ldots z_n y$, then we would first move z_n to the prefix of the image of y, then z_{n-1}, and so on. Again this requires moving each variable at most once, and once we have done this for all variables, then we will be left with exactly the morphism φ''. Overall we have

moved each variable at most twice. Since each move requires two successors in the underlying sequence, we need at most $4\Delta(E)$ successors in total and the statement of the lemma follows. □

We are now ready to give the full statement relating $\mathscr{G}^{\Rightarrow}_{[E]}$ and $\mathscr{G}^{\Rightarrow}_{[E_Y]}$ formally as follows. An example demonstrating the theorem is given in Fig. 5.

Theorem 6.8 *Let E be a basic RWE given by $\alpha \doteq \beta$. Let $Y = var(E)\backslash\Delta(E)$. Let E_Y be the equation $\pi_Y(\alpha) \doteq \pi_Y(\beta)$. Let $d = \max\{1, diam(\mathscr{G}^{\Rightarrow}_{[E_Y]})\}$. Then:*

1. *for each $\varphi \in \Phi_E$, $H^E_\varphi \subseteq [E]_\Rightarrow$ and $\mathscr{G}^{\Rightarrow}_{[E_Y]}$ is isomorphic to an isolated path contraction of order $Card(\Delta(E))$ of the subgraph $\mathscr{H}^E_\varphi$ of $\mathscr{G}^{\Rightarrow}_{[E]}$ induced by H^E_φ.*
2. $\mathscr{G}^{\Rightarrow}_{[E]} = \bigcup_{\varphi\in\Phi_E} \mathscr{H}^E_\varphi$.
3. $diam(\mathscr{G}^{\Rightarrow}_{[E]}) \in O(d|E|^2)$.

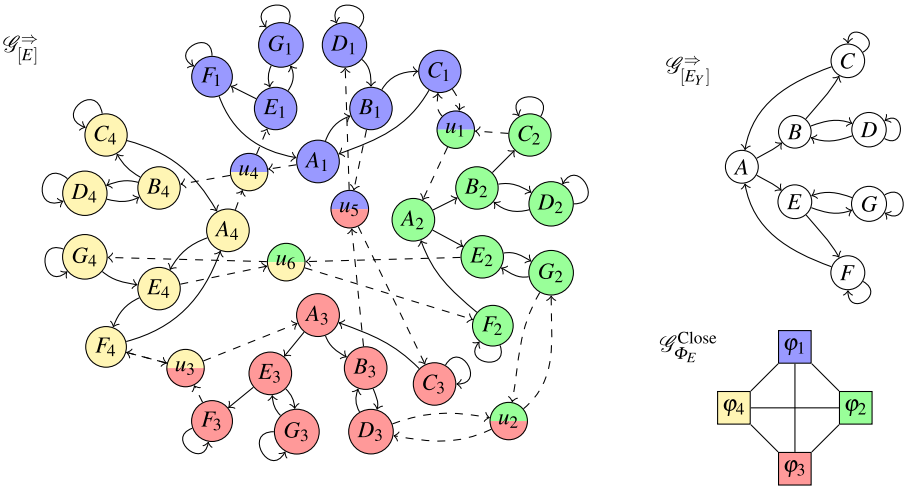

Fig. 5 Example illustrating Theorem 6.8. On the left is $\mathscr{G}^{\Rightarrow}_{[E]}$ for the equation E given by $y_1xy_2y_3y_4 \doteq y_4y_3xy_2y_1$. Note that $\Delta(E) = \{x\}$, so $Y = \{y_1, y_2, y_3, y_4\}$ and E_Y is given by $y_1y_2y_3y_4 \doteq y_4y_3y_2y_1$. The graph $\mathscr{G}^{\Rightarrow}_{[E_Y]}$ is shown on the top-right, where the equations in $[E_Y]_\Rightarrow$ have been labelled A, B, C, D, E, F, G. The set Φ_E contains four morphisms φ_i, $1 \le i \le 4$, such that $\varphi_i(y_i) = xy_i$ and $\varphi_i(y_j) = y_j$ for $j \neq i$. In this case, all morphisms in Φ_E are close to each other so the closeness relation (depicted as the graph $\mathscr{G}^{\text{Close}}_{\Phi_E}$ on the bottom-right) is a complete graph. The graph $\mathscr{G}^{\Rightarrow}_{[E]}$ is comprised of four subgraphs $\mathscr{H}^E_{\varphi_i}$, $1 \le i \le 4$. Each subgraph and morphism from Φ_E is depicted with a distinct colour in the figure. For each $Z \in \{A, B, C, D, E, F, G\}$ given by $\alpha_Z \doteq \beta_Z$, Z_i denotes the equation $\varphi_i(\alpha_Z) \doteq \varphi_i(\beta_Z)$. Thus the set of vertices unique to the subgraph $\mathscr{H}^E_{\varphi_i}$ is given by $V^E_{\varphi_i} = \{A_i, B_i, C_i, D_i, E_i, F_i, G_i\}$. The vertices shared between two subgraphs (i.e. those belonging to $U^E_{\varphi_i}$) are labelled $u_1, u_2, \ldots, u_6$. Since any two morphisms from Φ_E are close, each pair of subgraphs have at least one vertex in common. Each subgraph can be made isomorphic to $\mathscr{G}^{\Rightarrow}_{[E]}$ by contracting the paths (dashed) passing through the shared vertices $i_1, i_2, \ldots, i_6$. For example, the subgraph $\mathscr{H}^E_{\varphi_1}$ containing the vertices $A_1, B_1, C_1, D_1, E_1, F_1, G_1, u_1, u_4, u_5$ can be made isomorphic to $\mathscr{G}^{\Rightarrow}_{[E_Y]}$ by contracting the paths (A_1, u_4, E_1), (B_1, u_5, D_1), and (C_1, u_1, C_1) into single edges (A_1, E_1), (B_1, D_1) and (C_1, C_1)

Before we proceed with proving Theorem 6.8, it deserves a few further comments. Firstly, we note that since each morphism $\varphi \in \Phi_E$ is clearly injective, the subsets V_φ^E of vertices of each subgraph $\mathscr{H}_\varphi^E$ are pairwise disjoint. Consequently, while the subgraphs $\mathscr{H}_\varphi^E$ do overlap (and it is precisely these overlaps which mean they are all connected), each one contains a unique copy of the vertices of $\mathscr{G}_{[E_Y]}^{\Rightarrow}$.

Secondly, note that the number of morphisms in the set Φ_E will grow exponentially with respect to $\mathrm{Card}(\Delta(E))$. More precisely, we may assume some order $Y = \{y_1, y_2, \ldots, y_n\}$ on the variables in Y and represent each morphism $\varphi \in \Phi_E$ as a word $\varphi(y_1)\varphi(y_2)\ldots\varphi(y_n)$. This representation is clearly unique to φ. Furthermore, a word over $var(E)^*$ is a representation of this form for some $\varphi \in \Phi_E$ if and only if each variable occurs exactly once, the variables y_i occur in order from left to right, and y_n occurs as a suffix. Thus, the number of morphisms in total is given by

$$\mathrm{Card}(\Phi_E) = \frac{(\mathrm{Card}(var(E)) - 1)!}{(\mathrm{Card}(var(E)) - \mathrm{Card}(\Delta(E)))!}.$$

Since each subgraph contains a subset of vertices not shared with any other, it follows that the number of vertices in $\mathscr{G}_{[E]}^{\Rightarrow}$ will also be (at least) exponential in $\mathrm{Card}(\Delta(E))$. We shall see later in Section 9 that this is essentially the worst case for the size of $\mathscr{G}_{[E]}^{\Rightarrow}$ for RWEs E, with the largest graphs corresponding exactly to the case that $\mathrm{Card}(\Delta(E))$ is maximal. Nevertheless, it is worth pointing out that in the same case, the graph $\mathscr{G}_{[E_Y]}^{\Rightarrow}$ will be consist of a single vertex and two self-loops and thus the $diam(\mathscr{G}_{[E]}^{\Rightarrow})$ will be (at most) quadratic in $|E|$. This is significantly better than our upper bound in the general case.

Proof of Theorem 6.8 The rest of the section focuses on the proof of Theorem 6.8. The main technical content is presented in the following series of lemmas. Statement 1 is given by Lemmas 6.15 and 6.16, while Statements 2 and 3 are given by Lemmas 6.17 and 6.18 respectively. Throughout the remainder of this section, for a basic RWE E given by $\alpha \doteq \beta$ and a morphism φ, we shall use the notation $\varphi(E)$ as shorthand for $\varphi(\alpha) \doteq \varphi(\beta)$. We begin by noting some properties of equations belonging to the sets H_φ^E. The first deals with equations belonging to V_φ^E and follows directly from the definitions.

Fact 6.9 Let E be a basic RWE. Let $Y = var(E)\backslash\Delta(E)$, $n = \mathrm{Card}(\Delta(E))$ and let $E_Y = \pi_Y(E)$. Suppose that $\varphi \in \Phi_E$. Then $E' \in V_\varphi^E$ if and only if there exists a permutation $\sigma : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ and $y_1, y_2, \ldots, y_n$ with $Y = \{y_1, y_2, \ldots, y_n\}$ such that $y_1 y_2 \ldots y_n \doteq y_{\sigma(1)} y_{\sigma(2)} \ldots y_{\sigma(n)} \in [E_Y]_\Rightarrow$ and such that E' can be written as

$$\varphi(y_1)\varphi(y_2)\ldots\varphi(y_n) \doteq \varphi(y_{\sigma(1)})\varphi(y_{\sigma(2)})\ldots\varphi(y_{\sigma(n)}).$$

With a little additional reasoning, we can give a similar characterisation of equations contained in U_φ^E.

Lemma 6.10 *Let E be a basic RWE. Let $Y = var(E)\backslash\Delta(E)$, $n = Card(\Delta(E))$ and let $E_Y = \pi_Y(E)$. Suppose that $\varphi \in \Phi_E$. Then $E' \in U_\varphi^E$ if and only if there*

exist a permutation $\sigma : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ *and* $y_1, y_2, \ldots, y_n$ *with* $Y = \{y_1, y_2, \ldots, y_n\}$ *such that one of the following holds:*

1. $y_1 y_2 \ldots y_n \doteq y_{\sigma(1)} y_{\sigma(2)} \ldots y_{\sigma(n)} \in [E_Y]_\Rightarrow$ *and* E' *may be written as:*

$$\varphi(y_1)\varphi(y_2)\ldots\varphi(y_n) \doteq \delta_2\varphi(y_{\sigma(2)})\ldots\varphi(y_{\sigma(\iota-1)})\delta_1\varphi(y_{\sigma(\iota)})\ldots\varphi(y_{\sigma(n)})$$

2. $y_{\sigma(1)} y_{\sigma(2)} \ldots y_{\sigma(n)} \doteq y_1 y_2 \ldots y_n \in [E_Y]_\Rightarrow$ *and* E' *may be written as:*

$$\delta_2\varphi(y_{\sigma(2)})\ldots\varphi(y_{\sigma(\iota-1)})\delta_1\varphi(y_{\sigma(\iota)})\ldots\varphi(y_{\sigma(n)}) \doteq \varphi(y_1)\varphi(y_2)\ldots\varphi(y_n)$$

where $\sigma(\iota) = 1$, $\delta_1\delta_2 = \varphi(y_{\sigma(1)})$, *and* $\delta_1, \delta_2 \neq \varepsilon$.

Proof Suppose that E' satisfies the conditions of the lemma. We shall consider the case that Statement 1 holds. The case that Statement 2 holds is symmetric. Then $E'' \Rightarrow_L^* E'$ where E'' is the equation given by

$$\varphi(y_1)\varphi(y_2)\ldots\varphi(y_n) \doteq \overbrace{\delta_1\delta_2}^{\varphi(y_{\sigma(1)})}\ \varphi(y_{\sigma(2)})\ldots\varphi(y_{\sigma(\iota-1)})\varphi(y_{\sigma(\iota)})\ldots\varphi(y_{\sigma(n)}).$$

Consequently, $E'' = \varphi(\hat{E})$ for some $\hat{E} \in [E_Y]_\Rightarrow$, so $E'' \in V_\varphi^E$ and thus $E' \in H_\varphi^E$. Note however, that since E' is a basic RWE, each variable occurs exactly once on each side of the equation. We may therefore conclude that $\delta_1\delta_2 = \varphi(y_{\sigma(1)})$ is not a factor of the RHS of E', and consequently, by Fact 6.9, $E' \notin V_\varphi^E$. Thus $E' \in U_\varphi^E$.

Now suppose instead that $E' \in U_\varphi^E$. Then there exists some $E'' \in V_\varphi^E$, $k \in \mathbb{N}$ and $Z \in \{L, R\}$ such that $E'' \Rightarrow_Z^k E'$. Suppose we choose E'', Z and k such that k is minimal. Suppose additionally that $Z = L$. We shall show that Statement 1 of the lemma is satisfied. The case that $Z = R$ is symmetric and results in Statement 2 being satisfied.

Since we have $E'' \in V_\varphi^E$, it follows from Fact 6.9 that there exists a permutation $\sigma : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ and $y_1, y_2, \ldots, y_n$ with $Y = \{y_1, y_2, \ldots, y_n\}$ such that $y_1 y_2 \ldots y_n \doteq y_{\sigma(1)} y_{\sigma(2)} \ldots y_{\sigma(n)} \in [E_Y]_\Rightarrow$ and such that E'' can be written as

$$\varphi(y_1)\varphi(y_2)\ldots\varphi(y_n) \doteq \varphi(y_{\sigma(1)})\varphi(y_{\sigma(2)})\ldots\varphi(y_{\sigma(n)}).$$

Let $\ell = |\varphi(y_{\sigma(1)})|$ and let E''' be the equation given by

$$\varphi(y_1)\varphi(y_2)\ldots\varphi(y_n) \doteq \varphi(y_{\sigma(2)})\ldots\varphi(y_{\sigma(\iota-1)})\varphi(y_{\sigma(1)})\varphi(y_{\sigma(\iota)})\ldots\varphi(y_{\sigma(n)})$$

where $\sigma(\iota) = 1$. Then $E'' \Rightarrow_Z^\ell E'''$. However, since $y_1 y_2 \ldots y_n \doteq y_{\sigma(1)} y_{\sigma(2)} \ldots y_{\sigma(n)} \in [E_Y]_\Rightarrow$ and

$$y_1 y_2 \ldots y_n \doteq y_{\sigma(1)} y_{\sigma(2)} \ldots y_{\sigma(n)} \Rightarrow y_1 y_2 \ldots y_n \doteq y_{\sigma(2)} \ldots y_{\sigma(\iota-1)} y_{\sigma(1)} y_{\sigma(\iota)} \ldots y_{\sigma(n)},$$

we may conclude that $y_1 y_2 \ldots y_n \doteq y_{\sigma(2)} \ldots y_{\sigma(\iota-1)} y_{\sigma(1)} y_{\sigma(\iota)} \ldots y_{\sigma(n)} \in [E_Y]_\Rightarrow$. Thus, by Fact 6.9, $E''' \in V_\varphi^E$. Consequently, since V_φ^E and U_φ^E are by definition disjoint, we must have that $k \notin \{0, \ell\}$. Moreover, by our assumption that k is as minimal, we must have that $k < \ell$ (otherwise we could choose E''' in place of E' and get a smaller value of k). This directly implies that there exist $\delta_1, \delta_1 \neq \varepsilon$ with $\delta_1\delta_2 = \varphi(y_{\sigma(1)})$ such that E' may be written as

$$\varphi(y_1)\varphi(y_2)\ldots\varphi(y_n) \doteq \delta_2\varphi(y_{\sigma(2)})\ldots\varphi(y_{\sigma(\iota-1)})\delta_1\varphi(y_{\sigma(\iota)})\ldots\varphi(y_{\sigma(n)})$$

and thus Statement 1 of the lemma is satisfied. The case that $Z = R$ is symmetrical, leading instead to the satisfaction of Statement 2. □

We shall now focus on the claim that $H_\varphi^E \subseteq [E]_\Rightarrow$ for each $\varphi \in \Phi$. The first step is to show that for at least one $\varphi \in \Phi_E$, the equation $\varphi(E_Y)$ is contained in $[E]_\Rightarrow$.

Lemma 6.11 *Let E be a basic RWE. Let $Y = var(E)\backslash\Delta(E)$ and let $E_Y = \pi_Y(E)$. Then there exists $\varphi \in \Phi_E$ such that $\varphi(E_Y) \in [E]_\Rightarrow$.*

Proof Note that if $\Delta(E) = \emptyset$, then $E_Y = E$ and Φ_E contains only the identity morphism, so the lemma holds trivially. Suppose that $\Delta(E) \neq \emptyset$. By Remark 6.2, we may therefore assume that E is a basic RWE with at least two variables, so may write it as $x\alpha_1 u_1 u_2 \dots u_n y\alpha_2 \doteq y\beta_1 u_1 u_2 \dots u_n x\beta_2$ where $x, y, u_1, u_2, \dots, u_n \in X$ are pairwise distinct variables and $\alpha_1, \alpha_2, \beta_1, \beta_2 \in (var(E)\backslash\{x, y, u_1, u_2, \dots, u_n\})^*$, and such that α_1 and β_1 do not share a common non-empty suffix. Then $E \Rightarrow_R^* E'$ where E' is given by $u_1 u_2 \dots u_n x\alpha_1 y\alpha_2 \doteq y\beta_1 u_1 u_2 \dots u_n x\beta_2$.

Now, consider the function $Q_{E'}$ as defined in Definition 5.1. Note in particular that $Q_{E'}(\#) = (v, w)$ where $v, w \in X$ are the length-1 suffixes of $x\alpha_1$ and $y\beta_1$, and hence $v \neq w$. By Theorem 5.3, $\Upsilon_E = \Upsilon_{E'}$ (and hence $\Delta(E) = \Delta(E')$). Thus, for every $z \in \Delta(E)$, there exists $z' \in var(E)$ such that $Q_{E'}(z') = (z, z)$, meaning that z occurs directly to the left of z' on both the LHS and RHS of E'. It follows that each $z \in \Delta(E)$ has a unique 'successor' variable z' occurring to the right of z on both sides of the equation, and therefore that there exists some morphism $\varphi \in \Phi_E$ such that $E' = \varphi(\pi_Y(E'))$. Finally, notice that $u_i \in \Delta(E') = \Delta(E)$ for $1 \leq i \leq n$, and consequently, $\pi_Y(E') = \pi_Y(E) = E_Y$. □

The following lemma shows a correspondence between edges in $\mathscr{G}^{\Rightarrow}_{[E_Y]}$ and paths in the subgraphs $\mathscr{H}_\varphi^E$ of $\mathscr{G}^{\Rightarrow}_{[E]}$ which start and end with vertices from V_φ^E and whose internal vertices (if there are any) belong to U_φ^E.

Lemma 6.12 *Let E be a basic RWE. Let $Y = var(E)\backslash\Delta(E)$ and let $E_Y = \pi_Y(E)$. Let $Z \in \{L, R\}$ and suppose that $E', E'' \in [E_Y]_\Rightarrow$ such that $E' \Rightarrow_Z E''$. Let $\varphi \in \Phi_E$. Then there exist $k \leq Card(\Delta(E))$ and $E_0, E_1, E_2, \dots, E_{k+1}$ such that*

1. *$\varphi(E') = E_0$ and $\varphi(E'') = E_{k+1}$, and*
2. *$E_i \in U_\varphi^E$ for $1 \leq i \leq k$, and*
3. *$E_0 \Rightarrow_Z E_1 \Rightarrow_Z E_2 \Rightarrow_Z \dots \Rightarrow_Z E_k \Rightarrow_Z E_{k+1}$.*

Proof Note that if $\text{Card}(Y) < 2$, then $[E_Y]_\Rightarrow$ is a singleton and the lemma holds trivially. We may therefore assume that $\text{Card}(Y) \geq 2$. Suppose that $Z = R$. The case that $Z = L$ is symmetric. Then there exist $x, y \in Y$ and $\alpha_1, \alpha_2, \beta_1, \beta_2 \in Y^*$ such that E' may be written as $x\alpha_1 y\alpha_2 \doteq y\beta_1 x\beta_2$ and E'' may be written as $\alpha_1 x y\alpha_2 \doteq y\beta_1 x\beta_2$. Let $E_0 = \varphi(E')$ and $E_{k+1} = \varphi(E'')$. If $\varphi(x) = x$, then $E_0 \Rightarrow_R E_{k+1}$ so the lemma holds for $k = 0$. Suppose that $\varphi(x) \neq x$.

Then there exists k, $1 \leq k \leq \text{Card}(\Delta(E))$ and $z_1, z_2, \dots, z_k \in \Delta(E)$ such that $\varphi(x) = z_1 z_2 \dots z_k x$. For each i, $1 \leq i \leq k$, let E_i be the equation given by:

$$z_{i+1} \dots z_k x\varphi(\alpha_1) z_1 z_2 \dots z_i \varphi(y)\varphi(\alpha_2) \doteq \varphi(y)\varphi(\beta_1)\varphi(x)\varphi(\beta_2).$$

Then it follows directly from Lemma 6.10 that $E_i \in U_\varphi^E$ for $1 \leq i \leq k$. Moreover,

$$E_0 \Rightarrow_R E_1 \Rightarrow_R E_2 \Rightarrow_R \ldots \Rightarrow_R E_k \Rightarrow_R E_{k+1}$$

as required. □

A straightforward induction on Lemma 6.12 allows us to conclude that if, for some $\varphi \in \Phi_E$, $\varphi(E_Y) \in [E]_\Rightarrow$, then $H_\varphi^E \subseteq [E]_\Rightarrow$. We have already shown (Lemma 6.11 that this is true for at least one choice of φ. The next step is to show that $\varphi(E_Y) \in [E]_\Rightarrow$ for all $\varphi \in \Phi_E$, which we obtain as a consequence of Lemmas 6.7 and 6.14 below. Before proving Lemma 6.14, we need the following result, which we shall reuse later and is therefore stated separately.

Lemma 6.13 *Let E be a basic RWE. Then there exist $n_1, n_2 < |E|^2$ and $\hat{E}$ such that $E \Rightarrow^{n_1} \hat{E}$ and $\hat{E} \Rightarrow^{n_2} E$ where $\hat{E}$ can be written as $x\alpha y \doteq y\beta x$ where $x, y \in var(E)$ and $\alpha, \beta \in (var(E)\backslash\{x, y\})^*$*

Proof We shall prove the case that $E \Rightarrow^{n_1} \hat{E}$. The case that $\hat{E} \Rightarrow^{n_2} E$ is easily adapted. Recall that we may write any basic RWE as $x\alpha_1 y\alpha_2 \doteq y\beta_1 x\beta_2$ where $x, y \in X$ and $\alpha_1, \alpha_2, \beta_1, \beta_2 \in (X\backslash\{x, y\})^*$. We have the following claim:

Claim 6.13.1 For every basic, regular equation E given by $x\alpha_1 y\alpha_2 \doteq y\beta_1 x\beta_2$, either $\alpha_2 = \beta_2 = \varepsilon$, or there exists $n < |E|$ and E' such that $E \Rightarrow^n E'$ and E' may be written as $x'\alpha_1' y'\alpha_2' \doteq y'\beta_1' x'\beta_2'$ where $x', y' \in X$, $\alpha_1', \alpha_2', \beta_1', \beta_2' \in (X\backslash\{x', y'\})^*$, and such that $|\alpha_1'| + |\beta_1'| > |\alpha_1| + |\beta_1|$.

Proof Let E be given by $x\alpha_1 y\alpha_2 \doteq y\beta_1 x\beta_2$ where $x, y \in var(E)$ and $\alpha_1, \alpha_2, \beta_1, \beta_2 \in (var(E)\backslash\{x, y\})^*$. We have two cases, either $var(\alpha_1) = var(\beta_1)$, in which case, due to the fact that E is basic and therefore indecomposable, we must have that $\alpha_2 = \beta_2 = \varepsilon$, so the claim holds. Otherwise, there exists $z \in (var(\alpha_1)\backslash var(\beta_1)) \cup (var(\beta_1)\backslash var(\alpha_1))$. W.l.o.g. suppose that $z \in var(\alpha_1)\backslash var(\beta_1)$. Then since E is regular, $z \in var(\beta_2)$ and we can write E as $x\gamma_1 z\gamma_2 y\alpha_2 \doteq y\beta_1 x\delta_1 z\delta_2$ where $\gamma_1, \gamma_2, \delta_1, \delta_2 \in (var(E)\backslash\{x, y, z\})^*$. Consequently, we have that $E \Rightarrow_R^* E'$ where E' is given by $z\gamma_2 x\gamma_1 y\alpha_2 \doteq y\beta_1 x\delta_1 z\delta_2$. By Remark 3.2, we have that $E \Rightarrow^n E'$ where $n < |E|$. Moreover, E' clearly has the form described in the claim as witnessed by $x' = z$, $y' = y$, $\alpha_1' = \gamma_2 x\gamma_1$, $\alpha_2' = \alpha_2$, $\beta_1' = \beta_1 x\delta_1$ and $\beta_2' = \delta_2$. □

Since for any equation E of the form $x\alpha_1 y\alpha_2 \doteq y\beta_1 x\beta_2$, we must clearly have that $|\alpha_1| + |\beta_1| < |E|$, it follows from a simple induction on Claim 6.13.1 that $E \Rightarrow^n \hat{E}$ for some $n < |E|^2$ and $\hat{E}$ of the form $x'\alpha y' \doteq y'\beta x'$ as claimed. □

Lemma 6.14 *Let E be a basic RWE. Let $Y = var(E)\backslash\Delta(E)$ and let $E_Y = \pi_Y(E)$. Let $\varphi_1, \varphi_2 \in \Phi_E$. Then $H_{\varphi_1}^E \cap H_{\varphi_2}^E \neq \emptyset$ if and only if φ_1, φ_2 are close.*

Proof If $\mathrm{Card}(Y) < 2$, then Φ_E consists of the identity morphism only so the statement holds trivially. Suppose that $\mathrm{Card}(Y) \geq 2$. Suppose firstly that φ_1, φ_2 are close. Then there exist $y_1, y_2 \in Y$ with $y_1 \neq y_2$ and $\gamma_1, \gamma_2 \in \Delta(E)^*$ such that $\varphi_1(y_1) = \gamma_1\gamma_2 y_1$, $\varphi_2(y_1) = \gamma_2 y_1$, $\varphi_2(y_2) = \gamma_1\varphi_1(y_2)$, and for $y \in Y\backslash\{y_1, y_2\}$, $\varphi_1(y) = \varphi_2(y)$. In order to show that $H_{\varphi_1}^E \cap H_{\varphi_2}^E \neq \emptyset$, we need the following claim.

Claim 6.14.1 There exist $\hat{E} \in [E_Y]_{\Rightarrow}$ and $\hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta}_1, \hat{\beta}_2 \in (Y \backslash \{y_1, y_2\})^*$ such that $\hat{E}$ can be written either as:

1. $y_1\hat{\alpha}_1 y_2\hat{\alpha}_2 \doteq y_2\hat{\beta}_1 y_1\hat{\beta}_2$, or
2. $y_2\hat{\alpha}_1 y_1\hat{\alpha}_2 \doteq y_1\hat{\beta}_1 y_2\hat{\beta}_2$

Proof By Lemma 6.13, there exists $\hat{E}' \in [E_Y]_{\Rightarrow}$ such that $\hat{E}'$ may be written as $x\hat{\alpha}z \doteq z\hat{\beta}x$ where $x, z \in Y$, $x \neq z$ and $\hat{\alpha}, \hat{\beta} \in (Y \backslash \{x, z\})^*$. By Lemma 6.3, E_Y is basic, meaning that each variable in $Y = var(E_Y)$ occurs exactly once on each side of E_Y. It follows by properties of $\Rightarrow$ that each variable in Y also occurs exactly once in each of $x\hat{\alpha}z$ and $z\hat{\beta}x$. Hence there exist $\hat{\alpha}', \hat{\alpha}'' \in (Y \backslash \{y_2\})^*$ such that $x\hat{\alpha}z = \hat{\alpha}' y_2 \hat{\alpha}''$ (and such that y_1 occurs in either $\hat{\alpha}'$ or $\hat{\alpha}''$).

Suppose w.l.o.g. that y_1 occurs to the left of y_2 in the RHS. We shall show that Statement 1 of the lemma is satisfied. The case that y_2 occurs to the right of y_1 is symmetric and leads to Statement 2 being satisfied. Then there exist $\hat{\beta}', \hat{\beta}'', \hat{\beta}''' \in (Y \backslash \{y_1, y_2\})^*$ such that $z\hat{\beta}x = \hat{\beta}' y_1 \hat{\beta}'' y_2 \hat{\beta}'''$. Then we may write $\hat{E}'$ as

$$\hat{\alpha}' y_2 \hat{\alpha}'' \doteq \hat{\beta}' y_1 \hat{\beta}'' y_2 \hat{\beta}'''.$$

Note that z is a suffix of $y_2\hat{\alpha}''$ and a prefix of $\hat{\beta}' y_1$. Since y_2 does not occur in $\hat{\beta}' y_1$, we have $y_2 \neq z$. Consequently, we may write $\hat{\alpha}'' = \hat{\alpha}''' z$ for some $\hat{\alpha}'''$. Then

$$\begin{array}{rl} & \overbrace{\hat{\alpha}' y_2 \hat{\alpha}''' z \doteq \hat{\beta}' y_1 \hat{\beta}'' y_2 \hat{\beta}'''}^{\hat{E}'} \\ \Rightarrow_R^* & y_2 \hat{\alpha}''' \hat{\alpha}' z \doteq \hat{\beta}' y_1 \hat{\beta}'' y_2 \hat{\beta}''' \\ \Rightarrow_L^* & y_2 \underbrace{\hat{\alpha}''' \hat{\alpha}' z}_{\hat{\alpha}_1 y_1 \hat{\alpha}_2} \doteq y_1 \underbrace{\hat{\beta}'' \hat{\beta}'}_{\hat{\beta}_1} y_2 \underbrace{\hat{\beta}'''}_{\hat{\beta}_2} \end{array}$$

so $y_2\hat{\alpha}'''\hat{\alpha}' z \doteq y_1\hat{\beta}''\hat{\beta}' y_2\hat{\beta}''' \in [E_Y]_{\Rightarrow}$. Since y_1 occurs either in $\hat{\alpha}'$ or in $\hat{\alpha}'' = \hat{\alpha}''' z$, we may write $\hat{\alpha}'''\hat{\alpha}' z$ as $\hat{\alpha}_1 y_1 \hat{\alpha}_2$ for some $\hat{\alpha}_1, \hat{\alpha}_2 \in (Y \backslash \{y_1, y_2\})^*$. Thus the first statement of the lemma holds with $\hat{\beta}_1 = \hat{\beta}''\hat{\beta}'$ and $\hat{\beta}_2 = \hat{\beta}'''$. □

Assume that the first statement of Claim 6.14.1 holds. The case that the second statement holds is symmetric. Then there exists $\hat{E} \in [E_Y]_{\Rightarrow}$ such that $\hat{E}$ has the form $y_1\hat{\alpha}_1 y_2\hat{\alpha}_2 \doteq y_2\hat{\beta}_1 y_1\hat{\beta}_2$, for some $\hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta}_1, \hat{\beta}_2 \in (Y \backslash \{y_1, y_2\})^*$. Let E_{INT} be the equation given by

$$\gamma_2 y_1 \varphi_1(\hat{\alpha}_1)\gamma_1\varphi_1(y_2)\varphi_1(\hat{\alpha}_2) \doteq \varphi_1(y_2)\varphi_1(\hat{\beta}_1)\gamma_1\gamma_2 y_1 \varphi_1(\hat{\beta}_2)$$

and notice that

$$\begin{array}{rl} & \overbrace{\varphi_1(y_1)\varphi_1(\hat{\alpha}_1)\varphi_1(y_2)\varphi_1(\hat{\alpha}_2) \doteq \varphi_1(y_2)\varphi_1(\hat{\beta}_1)\varphi_1(y_1)\varphi_1(\hat{\beta}_2)}^{\varphi_1(\hat{E})} \\ \Rightarrow_R^* & \underbrace{\gamma_2 y_1 \varphi_1(\hat{\alpha}_1)\gamma_1\varphi_1(y_2)\varphi_1(\hat{\alpha}_2) \doteq \varphi_1(y_2)\varphi_1(\hat{\beta}_1)\gamma_1\gamma_2 y_1\varphi_1(\hat{\beta}_2)}_{E_{INT}}. \end{array}$$

Moreover, recall that $\varphi_2(y_1) = \gamma_2 y_1$, $\varphi_2(y_2) = \gamma_1\varphi_1(y_2)$. Since $\hat{\alpha}_1, \hat{\alpha}_2 \in (Y\backslash\{y_1, y_2\})^*$, we also have $\varphi_2(\hat{\alpha}_1) = \varphi_1(\hat{\alpha}_1)$ and $\varphi_2(\hat{\alpha}_2) = \varphi_1(\hat{\alpha}_2)$. Consequently

$$\begin{aligned} & \overbrace{\gamma_2 y_1\varphi_1(\hat{\alpha}_1)\gamma_1\varphi_1(y_2)\varphi_1(\hat{\alpha}_2) \doteq \gamma_1\varphi_1(y_2)\varphi_1(\hat{\beta}_1)\gamma_2 y_1\varphi_1(\hat{\beta}_2)}^{\varphi_2(\hat{E})} \\ \Rightarrow_L^* \quad & \underbrace{\gamma_2 y_1\varphi_1(\hat{\alpha}_1)\gamma_1\varphi_1(y_2)\varphi_1(\hat{\alpha}_2) \doteq \varphi_1(y_2)\varphi_1(\hat{\beta}_1)\gamma_1\gamma_2 y_1\varphi_1(\hat{\beta}_2)}_{E_{INT}}. \end{aligned}$$

Since $\hat{E} \in [E_Y]_{\Rightarrow}$, by definition $\varphi_1(\hat{E}) \in V_{\varphi_1}^E$ and $\varphi_2(\hat{E}) \in V_{\varphi_2}^E$. Thus it follows that $E_{INT} \in U_{\varphi_1}^E \cap U_{\varphi_2}^E$ and consequently $H_{\varphi_1}^E \cap H_{\varphi_2}^E \neq \emptyset$.

Now suppose instead that $H_{\varphi_1}^E \cap H_{\varphi_2}^E \neq \emptyset$. Let $E_{INT} \in H_{\varphi_1}^E \cap H_{\varphi_2}^E$. If $\varphi_1 = \varphi_2$ then the statement holds trivially. Thus we assume that $\varphi_1 \neq \varphi_2$. Before we proceed, we need the following claim.

Claim 6.14.2 Let $\varphi', \varphi'' \in \Phi_E$ and $\mu', \mu'' \in Y^*$ such that $|\mu'|_y = |\mu''|_y = 1$ for all $y \in Y$. If $\varphi'(\mu') = \varphi''(\mu'')$, then $\varphi' = \varphi''$ and $\mu' = \mu''$.

Proof Suppose that $\varphi'(\mu') = \varphi''(\mu'')$. It follows from the definition of Φ_E that for any $\varphi \in \Phi$, the morphism $\pi_Y \circ \varphi$ is the identity over Y. Thus $\mu' = \pi_Y(\varphi'(\mu')) = \pi_Y(\varphi''(\mu'')) = \mu''$. Furthermore, for each $y \in Y$, we may uniquely reconstruct $\varphi'(y)$ and $\varphi''(y)$ as the longest factors of the form $\Delta(E)^* y$ in $\varphi'(\mu')$ and $\varphi''(\mu'')$ respectively. It follows from the definition of Φ_E and the fact that $|\mu'|_y, |\mu''|_y = 1$ that these factors will exist and be unique. Thus, under the assumption that $\varphi'(\mu') = \varphi''(\mu'')$, it follows that $\varphi'(y) = \varphi''(y)$ for all $y \in Y$ and hence $\varphi' = \varphi''$. □

It follows from Fact 6.9 and Lemma 6.10 that for each $i \in \{1, 2\}$, there exists $\mu_i \in Y^*$ with $|\mu_i|_y = 1$ for all $y \in Y$ such that at least one of the LHS or RHS of E_{INT} has the form $\varphi_i(\mu_i)$. By Claim 6.14.2, and since $\varphi_1 \neq \varphi_2$, a single side of E_{INT} cannot have the form $\varphi_i(\mu_i)$ for both $i = 1$ and $i = 2$. By Fact 6.9, this means that $E_{INT} \notin V_{\varphi_1}^E, V_{\varphi_2}^E$ and consequently that $E_{INT} \in U_{\varphi_1}^E \cap U_{\varphi_2}^E$. Thus, either Statement 1 or Statement 2 of Lemma 6.10 holds with $\varphi = \varphi_1$ and $E' = E_{INT}$. W.l.o.g. suppose that the LHS of E_{INT} has the form $\varphi_1(\mu_1)$ and the RHS of E_{INT} has the form $\varphi_2(\mu_2)$. This corresponds to the case that Statement 1 of Lemma 6.10 holds, so there exist $y_1, y_2, \ldots, y_n$ with $Y = \{y_1, y_2, \ldots, y_n\}$ and a permutation $\sigma : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ such that E_{INT} may be written as

$$\varphi_1(y_1)\varphi_1(y_2)\ldots\varphi_1(y_n) \doteq \delta_2\varphi_1(y_{\sigma(2)})\ldots\varphi_1(y_{\sigma(\iota-1)})\delta_1\varphi_1(y_{\sigma(\iota)})\ldots\varphi_1(y_{\sigma(n)})$$

where $\delta_1\delta_2 = \varphi_1(y_{\sigma(1)})$ with $\delta_1, \delta_2 \neq \varepsilon$ and $\sigma(\iota) = 1$. Note that by the definition of Φ_E, the fact that $\delta_2 \neq \varepsilon$ implies that $\delta_1 \in \Delta(E)^*$ and $\delta_2 = \delta_3 y_{\sigma(1)}$ for some $\delta_3 \in \Delta(E)^*$.

Recalling that the RHS of E_{INT} has the form $\varphi_2(\mu_2)$, we may directly infer that $\mu_2 = y_{\sigma(1)}y_\sigma(2)\ldots y_{\sigma(n)}$ and subsequently $\varphi_2(y_{\sigma(2)}) = \delta_2$, $\varphi_2(y_{\sigma(\iota)}) = \delta_1\varphi_1(y_{\sigma(\iota)})$, and $\varphi_2(y) = \varphi_1(y)$ for all $y \notin \{y_{\sigma(2)}, y_{\sigma(\iota)}\}$. Thus φ_1 and φ_2 are close as required. □

We are now able to prove that each set H_φ^E is in fact a subset of the vertices of $\mathscr{G}_{[E]}^{\Rightarrow}$, and thus that the subgraphs $\mathscr{H}_\varphi^E$ of $\mathscr{G}_{[E]}^{\Rightarrow}$ are well-defined.

Lemma 6.15 *Let E be a basic RWE. Then $H_\varphi^E \subseteq [E]_\Rightarrow$ for each $\varphi \in \Phi_E$.*

Proof Let $Y = var(E)\backslash\Delta(E)$ and let $E_Y = \pi_Y(E)$. By Lemma 6.11, there exists $\varphi \in \Phi_E$ such that $\varphi(E_Y) \in [E]_\Rightarrow$. Let $\tilde{E} \in H_{\varphi'}^E$ for some arbitrary $\varphi' \in \Phi_E$. By Lemma 6.7, there exist $k \leq 4\text{Card}(\Delta(E)) + 1$ and $\varphi_1, \varphi_2, \ldots, \varphi_k \in \Phi_E$ such that $\varphi = \varphi_1, \varphi' = \varphi_k$, and for $1 \leq i < k$, φ_i and φ_{i+1} are close. Thus, by Lemma 6.14, there exist $E_1, E_2, \ldots, E_k$ such that $E_i \in H_{\varphi_i}^E \cap H_{\varphi_{i+1}}^E$ for $1 \leq i < k$.

It follows from Lemma 6.12 that if $E', E'' \in H_{\varphi_i}^E$ for some i, $1 \leq i \leq k$, then $E' \Rightarrow^* E''$. Thus, $\varphi(E_Y) \Rightarrow^* \varphi(E_1)$, $E_k \Rightarrow^* \tilde{E}$, and for $1 \leq i \leq k$, $E_i \Rightarrow^* E_{i+1}$. Consequently, $\tilde{E} \in [E]_\Rightarrow$. Since this holds for all $\tilde{E} \in \mathscr{H}_{\varphi'}^E$ for all $\varphi' \in \Phi_E$, the lemma follows. □

The following lemma completes the proof of Statement 1 of Theorem 6.8.

Lemma 6.16 *Let E be a basic RWE. Let $Y = var(E)\backslash\Delta(E)$, let $E_Y = \pi_Y(E)$, and let $\varphi \in \Phi_E$. Then $\mathscr{G}_{[E_Y]}^{\Rightarrow}$ is isomorphic to an isolated path contraction of order Card$(\Delta(E))$ of $\mathscr{H}_\varphi^E$.*

Proof For $k \geq 0$, we shall say that a sequence of equations $E_0, E_1, \ldots, E_{k+1}$ as a *U-path* if $E_0, E_{k+1} \in V_\varphi^E$, $E_i \in U_\varphi^E$ for $1 \leq i \leq k$, and there exists $Z \in \{L, R\}$ such that $E_0 \Rightarrow_Z E_1 \Rightarrow_Z E_2 \Rightarrow_Z \ldots \Rightarrow_Z E_k \Rightarrow_Z E_{k+1}$. Let $\diamond$ be the relation on V_φ^E such that $E' \diamond E''$ if and only if $E', E'' \in V_\varphi^E$ and there exists a U-path starting with E' and ending with E''. We shall show firstly that the graph $\mathscr{G}_{V_\varphi^E}^{\diamond}$ is an isolated path compression of order Card$(\Delta(E))$ of $\mathscr{H}_\varphi^E$, and secondly that $\mathscr{G}_{[E_Y]}^{\Rightarrow}$ is isomorphic to $\mathscr{G}_{V_\varphi^E}^{\diamond}$.

Clearly, every U-path is a path in $\mathscr{H}_\varphi^E$. Moreover, it follows from the definition of H_φ^E, along with the fact that $\Rightarrow_Z^*$ is an equivalence relation for $Z \in \{L, R\}$, that for every vertex $E' \in U_\varphi^E$, there exist $E'', E''' \in V_\varphi^E$ and $Z \in \{L, R\}$ such that $E'' \Rightarrow_Z^* E'$ and $E' \Rightarrow_Z^* E'''$. Consequently, every vertex in $\mathscr{H}_\varphi^E$ either belongs to V_φ^E or is the internal vertex of some U-path. It follows as a direct consequence of the following claim that U-path containing a given vertex in U_φ^E is unique, and therefore that no two distinct U-paths share an internal vertex. Thus $\mathscr{G}_{V_\varphi^E}^{\diamond}$ is an isolated path compression of order k of $\mathscr{H}_\varphi^E$ where k is the number of internal vertices in the longest U-path in $\mathscr{H}_\varphi^E$.

Claim 6.16.1 Let $E' \in U_\varphi^E$. Then the in- and out-degrees of E' in $\mathscr{H}_\varphi^E$ are exactly one.

Proof Since $E' \in U_\varphi^E$, there exist a permutation $\sigma : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ and $y_1, y_2, \ldots, y_n$ with $Y = \{y_1, y_2, \ldots, y_n\}$ such either Statement 1 or Statement 2

of Lemma 6.10 holds. Suppose that Statement 1 holds. The case that Statement 2 holds is symmetric. Then we may write E' as follows:

$$\varphi(y_1)\varphi(y_2)\ldots\varphi(y_n) \doteq \delta_2\varphi(y_{\sigma(2)})\ldots\varphi(y_{\sigma(\iota-1)})\delta_1\varphi(y_{\sigma(\iota)})\ldots\varphi(y_{\sigma(n)})$$

where $\sigma(\iota) = 1$, $\delta_1\delta_2 = \varphi(y_{\sigma(1)})$ and $\delta_1, \delta_2 \neq \varepsilon$. Moreover, $\hat{E} \in [E_Y]_{\Rightarrow}$ where $\hat{E}$ is given by $y_1y_2\ldots y_n \doteq y_{\sigma(1)}y_{\sigma(2)}\ldots y_{\sigma(n)}$. Note that $\varphi(\hat{E}) \Rightarrow_L^* E'$.

Let $E'_{\mathrm{pre}_L}, E'_{\mathrm{suc}_L}$ be the equations such that $E'_{\mathrm{pre}_L} \Rightarrow_L E'$ and $E' \Rightarrow_L E'_{\mathrm{suc}_L}$. It follows from the definitions that $\varphi(\hat{E}) \Rightarrow_L^* E'_{\mathrm{pre}_L}$ and $\varphi(\hat{E}) \Rightarrow_L^* E'_{\mathrm{suc}_L}$, so both belong to H_φ^E and the in- and out-degree of E' in $\mathscr{H}_\varphi^E$ are both at least one. To see that they are exactly one, we must show that for the equations E'_{pre_R} and E'_{suc_R} such that $E'_{\mathrm{pre}_R} \Rightarrow_R E'$ and $E' \Rightarrow_R E'_{\mathrm{suc}_R}$, neither E'_{pre_R} nor E'_{suc_R} is contained in the set H_φ^E. We may write E'_{pre_R} as

$$z\varphi(y_1)\varphi(y_2)\ldots\delta_3\delta_2\ldots\varphi(y_n) \doteq \delta_2\varphi(y_{\sigma(2)})\ldots\varphi(y_{\sigma(i-1)})\delta_1\varphi(y_{\sigma(i)})\ldots\varphi(y_{\sigma(n)})$$

where $z \in X$ and $\delta_3 \in X^*$ such that $\delta_3 z = \delta_1$, and we may write E'_{suc_R} as

$$\gamma\varphi(y_2)\ldots\delta_1 z'\delta_2\ldots\varphi(y_n) \doteq \delta_2\varphi(y_{\sigma(2)})\ldots\varphi(y_{\sigma(i-1)})\delta_1\varphi(y_{\sigma(i)})\ldots\varphi(y_{\sigma(n)})$$

where $z' \in X$ and $\gamma \in X^*$ such that $z'\gamma = \varphi(y_1)$. It follows by Fact 6.9 and Lemma 6.10 that any equation in $V_\varphi^E \cup U_\varphi^E = H_\varphi^E$ must have $\varphi(y_{\sigma(1)}) = \delta_1\delta_2$ occurring as a factor of at least one side. However, since each variable occurs exactly once on each side of the equations $E'_{\mathrm{pre}_R}, E'_{\mathrm{suc}_R}$, we may immediately observe that $\varphi(y_{\sigma(1)})$ does not occur as a factor of the LHS or of the RHS of either equation. Thus $E'_{\mathrm{pre}_R}, E'_{\mathrm{suc}_R} \notin H_\varphi^E$, and the in- and out-degrees of E' in $\mathscr{H}_\varphi^E$ are exactly one as claimed. □

The following claim asserts that each vertex in $E' \in U_\varphi^E$ occurs on a U-path with at most $\mathrm{Card}(\Delta(E))$ internal vertices. Since we have already shown that E' occurs on exactly one U-path, it follows that all U-paths have at most $\mathrm{Card}(\Delta(E))$ internal vertices and thus that the order of the isolated path compression is at most $\mathrm{Card}(\Delta(E))$.

Claim 6.16.2 Let $E' \in U_\varphi^E$. Then there exist $k \leq \mathrm{Card}(\Delta(E))$, $E_0, E_1, \ldots, E_{k+1}$ and $Z \in \{L, R\}$ such that:

1. $E_0, E_{k+1} \in V_\varphi^E$, and
2. $E_i \in U_\varphi^E$ for $1 \leq i \leq k$, and
3. $E_i \Rightarrow_Z E_{i+1}$ for $0 \leq i \leq k$, and
4. there exists i, $1 \leq i \leq k$ such that $E' = E_i$.

Proof Since $E' \in U_\varphi^E$, there exist a permutation $\sigma : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ and $y_1, y_2, \ldots, y_n$ with $Y = \{y_1, y_2, \ldots, y_n\}$ such either Statement 1 or Statement 2 of Lemma 6.10 holds. Suppose that Statement 1 holds. The case that Statement 2 holds is symmetric. Then the equation $\hat{E}$ given by $y_1y_2\ldots y_n \doteq y_{\sigma(1)}y_{\sigma(2)}\ldots y_{\sigma(n)}$ is contained in $[E_Y]_{\Rightarrow}$ and we may write E' as follows

$$\varphi(y_1)\varphi(y_2)\ldots\varphi(y_n) \doteq z_{j+1}\ldots z_k\varphi(y_{\sigma(2)})\ldots\varphi(y_{\sigma(\iota-1)})z_1\ldots z_j\varphi(y_{\sigma(\iota)})\ldots\varphi(y_{\sigma(n)})$$

where $\sigma(\iota) = 1$, $z_1 z_2 \dots z_k = \varphi(y_{\sigma(1)})$ and $1 \leq j < k \leq \mathrm{Card}(\Delta(E)) + 1$.

Now, let E_0 be the equation given by

$$\varphi(y_1)\varphi(y_2)\dots\varphi(y_n) \doteq \overbrace{z_1 z_2 \dots z_k}^{\varphi(y_1)} \varphi(y_{\sigma(2)}) \dots \varphi(y_{\sigma(\iota-1)})\varphi(y_{\sigma(\iota)}) \dots \varphi(y_{\sigma(n)}),$$

let E_{k+1} be the equation

$$\varphi(y_1)\varphi(y_2)\dots\varphi(y_n) \doteq \varphi(y_{\sigma(2)}) \dots \varphi(y_{\sigma(\iota-1)}) \overbrace{z_1 z_2 \dots z_k}^{\varphi(y_1)} \varphi(y_{\sigma(\iota)}) \dots \varphi(y_{\sigma(n)}),$$

and for $1 \leq i < k$, let E_i be the equation given by

$$\varphi(y_1)\varphi(y_2)...\varphi(y_n) \doteq z_{i+1}... z_k y_{\sigma(1)} \varphi(y_{\sigma(2)})...\varphi(y_{\sigma(\iota-1)}) z_1 \dots z_i \varphi(y_{\sigma(\iota)})...\varphi(y_{\sigma(n)}).$$

Then clearly we have $E_0 = \varphi(\hat{E}) \in V_\varphi^E$. Let $\hat{E}'$ be the equation given by $y_1 y_2 \dots y_n \doteq y_{\sigma(2)} \dots y_{\sigma(\iota-1)} y_{\sigma(1)} y_{\sigma(\iota)} \dots y_{\sigma(n)}$. Then $\hat{E} \Rightarrow \hat{E}'$ so $\hat{E}' \in [E_Y]_\Rightarrow$, and moreover $E_{k+1} = \varphi(\hat{E}')$ so $E_{k+1} \in V_\varphi^E$. Thus Statement 1 is satisfied. Note also that $E_i \Rightarrow_L E_{i+1}$ for $0 \leq i \leq k$, so Statement 3 is satisfied, and furthermore we have that $E_i \in H_\varphi^E$ for $1 \leq i \leq k$. For each i, $1 \leq i \leq k$, since each variable $y \in Y$ occurs exactly once on each side E_i, we may conclude that $\varphi(y_{\sigma(1)}) = z_1 z_2 \dots z_k$ is not a factor of the RHS of E_i. Thus, by Fact 6.9, $E_i \notin V_\varphi^E$ so $E_i \in U_\varphi^E$ and Statement 2 is satisfied. Finally note that $E' = E_j$, so Statement 4 is also satisfied. □

It remains to show that $\mathscr{G}^{\Rightarrow}_{[E_Y]}$ is isomorphic to $\mathscr{G}^{\diamond}_{V_\varphi^E}$. Recall that by definition $V_\varphi^E = \{\varphi(E') \mid E' \in [E_Y]_\Rightarrow\}$ and note that the function mapping equations $\hat{E} \in [E_Y]_\Rightarrow$ to their counterparts $\varphi(\hat{E}) \in V_\varphi^E$ is a bijection. Consequently, the fact that $\mathscr{G}^{\diamond}_{V_\varphi^E}$ is isomorphic to $\mathscr{G}^{\Rightarrow}_{[E_Y]}$ follows directly from the following claim.

Claim 6.16.3 Let $\hat{E}_1, \hat{E}_2 \in [E_Y]_\Rightarrow$. Then $\hat{E}_1 \Rightarrow \hat{E}_2$ if and only if $\varphi(\hat{E}_1) \diamond \varphi(\hat{E}_2)$.

Proof Suppose that $\hat{E}_1 \Rightarrow \hat{E}_2$. Then it follows from Lemma 6.12 that $\varphi(\hat{E}_1) \diamond \varphi(\hat{E}_2)$. Suppose instead that $\varphi(\hat{E}_1) \diamond \varphi(\hat{E}_2)$. Since $\hat{E}_1 \in [E_Y]_\Rightarrow$, it may be written as

$$y_1 y_2 \dots y_n \doteq y_{\sigma(1)} y_{\sigma(2)} \dots y_{\sigma(n)}$$

where $Y = \{y_1, y_2, \dots, y_n\}$ and $\sigma : \{1, 2, \dots, n\} \to \{1, 2, \dots, n\}$ is a permutation.

By definition of $\diamond$, there exists $Z \in \{L, R\}$ and $\ell \in \mathbb{N}$ such that $\varphi(\hat{E}_1) \Rightarrow_Z^\ell \varphi(\hat{E}_2)$. Suppose that $Z = L$. The case that $Z = R$ is symmetric. For $i > 1$, let E_i be the equation such that $\varphi(\hat{E}_1) \Rightarrow_L^i \varphi(E_i)$. Let $k = |\varphi(y_{\sigma(1)})| - 1$. Then we may write E_{k+1} as

$$\varphi(y_1)\varphi(y_2)\dots\varphi(y_n) \doteq \varphi(y_{\sigma(2)}) \dots \varphi(y_{\sigma(\iota-1)})\varphi(y_{\sigma(1)})\varphi(y_{\sigma(\iota)}) \dots \varphi(y_{\sigma(n)}).$$

Let $\hat{E}_3$ be the equation given by $y_1 y_2 \dots y_n \doteq y_{\sigma(2)} \dots y_{\sigma(\iota-1)} y_{\sigma(1)} y_{\sigma(\iota)} \dots y_{\sigma(n)}$. Then $\hat{E}_1 \Rightarrow \hat{E}_3$ so $\hat{E}_3 \in [E_Y]_\Rightarrow$ and it follows from Fact 6.9 that $E_{k+1} \in V_\varphi^E$. Hence we must have $\ell \leq k + 1$. Moreover, for $1 \leq i \leq k$, there exist δ_1, δ_2 such that $\delta_1 \delta_2 = \varphi(y_{\sigma(1)})$ and $\delta_1, \delta_2 \neq \varepsilon$ and such that we may write E_i as

$$\varphi(y_1)\varphi(y_2)\dots\varphi(y_n) \doteq \delta_2 \varphi(y_{\sigma(2)}) \dots \varphi(y_{\sigma(\iota-1)})\delta_1 \varphi(y_{\sigma(\iota)}) \dots \varphi(y_{\sigma(n)}).$$

Consequently, by Lemma 6.10, $E_i \in U_\varphi^E$ for $1 \leq i \leq k$. By definition, $\varphi(\hat{E}_2) \in V_\varphi^E$, so it follows that $\ell > k$ and thus $\ell = k+1$, and thus that in fact $\hat{E}_2 = \hat{E}_3$, meaning that $\hat{E}_1 \Rightarrow \hat{E}_2$ as required. □

Claims 6.16.1 and 6.16.2 show that the graph $\mathscr{G}^{\diamond}_{V_\varphi^E}$ is an isolated path compression of order $\mathrm{Card}(\Delta(E))$ of $\mathscr{H}_\varphi^E$. Claim 6.16.3 shows that $\mathscr{G}^{\Rightarrow}_{[E]}$ is isomorphic to $\mathscr{G}^{\diamond}_{V_\varphi^E}$, so the statement of the lemma holds. □

The following lemma deals with the second statement of Theorem 6.8. It asserts that the subgraphs $\mathscr{H}_\varphi^E$ completely cover the graph $\mathscr{G}^{\Rightarrow}_{[E]}$: each edge and each vertex of $\mathscr{G}^{\Rightarrow}_{[E]}$ also belong to at least one subgraph $\mathscr{H}_\varphi^E$.

Lemma 6.17 *Let E be a basic RWE. Then* $\mathscr{G}^{\Rightarrow}_{[E]} = \bigcup_{\varphi \in \Phi_E} \mathscr{H}_\varphi^E$.

Proof We have already shown in Lemma 6.15 that each vertex of $\bigcup_{\varphi \in \Phi_E} \mathscr{H}_\varphi^E$ is a vertex of $\mathscr{G}^{\Rightarrow}_{[E]}$. Moreover, it follows directly from the definition of $\mathscr{H}_\varphi^E$ that each edge in $\bigcup_{\varphi \in \Phi_E} \mathscr{H}_\varphi^E$ is also an edge of $\mathscr{G}^{\Rightarrow}_{[E]}$. It remains to show that each vertex/edge of $\mathscr{G}^{\Rightarrow}_{[E]}$ is a vertex/edge of $\mathscr{H}_\varphi^E$ for some $\varphi \in \Phi_E$. The main step is Claim 6.17.1 as follows.

Claim 6.17.1 For every $E' \in [E]_\Rightarrow$, and $Z \in \{L, R\}$, there exists $\varphi \in \Phi_E$ and $E'' \in V_\varphi^E$ such that $E'' \Rightarrow_Z^* E'$.

Proof Note that by Lemma 6.11, there exists $E_0 \in [E]_\Rightarrow$ and $\varphi_0 \in \Phi_E$ such that $E_0 \in V_{\varphi_0}^E$ and thus the claim holds for $E_0 = E'$. Note also that for every $E' \in [E]_\Rightarrow$, since $\Rightarrow^*$ is an equivalence relation, we have $E_0 \Rightarrow^* E'$. Thus it is sufficient to show that if the claim holds for E_i and $E_i \Rightarrow E_{i+1}$, then it also holds for E_{i+1}.

Suppose that the claim holds for $E_i \in [E]_\Rightarrow$ and that $E_i \Rightarrow_{Z_i} E_{i+1}$. Then there exist $\varphi_i \in \Phi_E$ and $E_i'' \in V_{\varphi_i}^E$ such that $E_i'' \Rightarrow_{Z_i}^* E_i$ and thus $E_i'' \Rightarrow_{Z_i}^* E_{i+1}$. Thus $E_{i+1} \in H_{\varphi_i}^E$. If $E_{i+1} \in V_{\varphi_i}^E$, then the claim holds trivially. Suppose instead that $E_{i+1} \in U_{\varphi_i}^E$.

Let $Y = var(E) \backslash \Delta(E)$ and let $E_Y = \pi_Y(E)$. Recall that there exist a permutation $\sigma : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ and $y_1, y_2, \ldots, y_n$ with $Y = \{y_1, y_2, \ldots, y_n\}$ such either Statement 1 or Statement 2 of Lemma 6.10 holds. Suppose that Statement 1 holds (the case that Statement 2 holds is symmetric). Then we may write E_{i+1} as

$$\varphi_i(y_1)\varphi_i(y_2)\ldots\varphi_i(y_n) \doteq \delta_2\varphi_i(y_{\sigma(2)})\ldots\varphi_i(y_{\sigma(\iota-1)})\delta_1\varphi_i(y_{\sigma(\iota)})\ldots\varphi_i(y_{\sigma(n)})$$

where $\sigma(\iota) = 1$, $\delta_1\delta_2 = \varphi_i(y_{\sigma(1)})$ and $\delta_1, \delta_2 \neq \varepsilon$. Furthermore, we have $\hat{E} \in [E_Y]_\Rightarrow$ where $\hat{E}$ is the equation given by

$$y_1 y_2 \ldots y_n \doteq y_{\sigma(1)} y_{\sigma(2)} \ldots y_{\sigma(n)}.$$

It is straightforward to see that for $Z = L$, $\varphi_i(\hat{E}) \Rightarrow_Z^* E_{i+1}$ and since $\varphi_i(\hat{E}) \in V_{\varphi_i}^E$ by definition, the claim holds in this case.

It remains to consider the case that $Z = R$. By Lemma 6.3, E_Y is basic and therefore indecomposable. Thus $y_1 \neq y_{\sigma(1)}$. Let $\varphi_{i+1} : Y^* \to X^*$ be the morphism such that $\varphi_{i+1}(y_{\sigma(1)}) = \delta_2$, $\varphi_{i+1}(y_1) = \delta_1\varphi_i(y_1)$, and $\varphi_{i+1}(y_j) = \varphi_i(y_j)$ for $1 \leq j \leq n$ with $j \notin \{1, \sigma(1)\}$. Note that $\varphi_{i+1} \in \Phi_E$ since $\delta_1 \in \Delta(E)^*$ and $\delta_2 \in \Delta(E)^* y_{\sigma(1)}$. Let E''_{i+1} be the equation given by $\varphi_{i+1}(\hat{E})$, so that $E''_{i+1} \in V^E_{\varphi_{i+1}}$. Then we may write E''_{i+1} as:

$$\begin{aligned} & \delta_1\varphi_i(y_1)\varphi_i(y_2)\ldots\varphi_i(y_{\sigma(1)-1})\delta_2\varphi_i(y_{\sigma(1)+1})\ldots\varphi_i(y_n) \\ \doteq\ & \delta_2\varphi_i(y_{\sigma(2)})\ldots\varphi_i(y_{\sigma(\iota-1)})\delta_1\varphi_i(y_{\sigma(\iota)})\ldots\varphi_i(y_{\sigma(n)}). \end{aligned}$$

Consequently $E''_{i+1} \Rightarrow^*_R E_{i+1}$, so the claim holds for E_{i+1} and by induction, it holds for all $E' \in [E]_{\Rightarrow}$. □

It follows directly from Claim 6.17.1 that every vertex of $\mathscr{G}^{\rightrightarrows}_{[E]}$ belongs to H^E_φ for some $\varphi \in \Phi_E$ and is consequently also a vertex of some subgraph $\mathscr{H}^E_\varphi$. To see why the same holds for edges, note firstly that for every edge (E_1, E_2) in $\mathscr{G}^{\rightrightarrows}_{[E]}$, there exists $Z \in \{L, R\}$ such that $E_1 \Rightarrow_Z E_2$. By Claim 6.17.1 and since $E_1 \in [E]_{\Rightarrow}$, there exist $\varphi \in \Phi_E$ and $E' \in V^E_\varphi$ such that $E' \Rightarrow^*_Z E_1$. It follows that $E' \Rightarrow^*_Z E_2$, meaning that $E_1, E_2 \in H^E_\varphi$ (so they are both vertices of $\mathscr{H}^E_\varphi$). It follows by definition that (E_1, E_2) is an edge of $\mathscr{H}^E_\varphi$. □

The proof of Theorem 6.8 is completed by the following lemma which addresses the third statement of the theorem.

Lemma 6.18 *Let E be a basic RWE. Let $Y = var(E)\backslash\Delta(E)$ and let $E_Y = \pi_Y(E)$. Let $d = \max\{1, diam(\mathscr{G}^{\rightrightarrows}_{[E_Y]})\}$. Then $diam(\mathscr{G}^{\rightrightarrows}_{[E]}) \in O(d|E|^2)$.*

Proof Let $E', E'' \in [E]_{\Rightarrow}$. Then by Lemma 6.17, there exist $\varphi', \varphi'' \in \Phi_E$ such that $E' \in H_{\varphi'}$ and $E'' \in H_{\varphi''}$. For each $\varphi \in \Phi_E$, note that by Lemma 6.16 and Remark 4.6, there is path of length $O(d\text{Card}(\Delta(E)))$ between any two vertices in H^E_φ. Thus if $\varphi' = \varphi''$, then there is path of length $O(d\text{Card}(\Delta(E)))$ from E' to E''.

Suppose otherwise that $\varphi' \neq \varphi''$. Then it follows from Lemma 6.7, there exist $k \in O(\text{Card}(\Delta(E)))$ and $\varphi_1, \varphi_2, \ldots, \varphi_k \in \Phi_E$ such that $\varphi' = \varphi_1$, $\varphi'' = \varphi_k$ and φ_i, φ_{i+1} are close for $1 \leq i < k$. By Lemma 6.14, there exist $E_1, E_2, \ldots, E_{k-1}$ such that $E_i \in H^E_{\varphi_i} \cap H^E_{\varphi_{i+1}}$ for $1 \leq i < k$.

It follows that there exist paths from E' to E_0, from E_k to E'' and from E_i to E_{i+1} for $1 \leq i < k$ of length $O(d\text{Card}(\Delta(E)))$. Thus there is a path from E' to E'' of length $O(kd\text{Card}(\Delta(E))) = O(d\text{Card}(\Delta(E))^2) = O(d|E|^2)$. Since this is true for all E', E'', the statement of the lemma follows. □

7 Normal Forms and Block Decompositions

Having described the structure of $\mathscr{G}^{\rightrightarrows}_{[E]}$ for equations E which are not jumbled in the previous section, the current section focuses on the structure of $\mathscr{G}^{\rightrightarrows}_{[E]}$ in the case that E is jumbled. Our main result in this direction is the existence of specific normal forms, from which every vertex in $\mathscr{G}^{\rightrightarrows}_{[E]}$ is polynomial distance away. We present two

normal forms, with the second being a restriction on the first. Both are constructed based on reversed structures in such a way that they allow for taking full advantage of the invariant Υ_E from Section 5. A major advantage of this is that we are able to show later in Section 8 that the number of equations occurring as vertices in $\mathscr{G}^{\Rightarrow}_{[E]}$ in the second normal form is bounded by a polynomial in $|E|$, allowing us to prove that the diameter of $\mathscr{G}^{\Rightarrow}_{[E]}$ is also polynomial.

Since the results in this section mainly concern positive reachability statements, the technical content relies heavily on describing sequences of applications of $\Rightarrow$. Certain sequences will occur repeatedly, so it is convenient to define some shorthand notations given in terms of the following 'shortcut' relations.

Definition 7.1 ($\xrightarrow{u,v}$ and $\odot$) For each $u, v \in X$, we define the relation $\xrightarrow{u,v}$ over basic regular equations as $E_1 \xrightarrow{u,v} E_2$ if there exist $x, y \in X$ and $\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3 \in (X\backslash\{u, v, x, y\})^*$ such that E_1 may be written as $x\alpha_1 u\alpha_2 v\alpha_3 y \doteq y\beta_1 u\beta_2 v\beta_3 x$ and E_2 may be written as $x\alpha_1 v\alpha_3 u\alpha_2 y \doteq y\beta_1 v\beta_3 u\beta_2 x$. Additionally, we define $\odot = \bigcup_{u,v\in X} \xrightarrow{u,v}$.

Note that there exist $u, v \in X$ such that $E_1 \xrightarrow{u,v} E_2$ if and only if $E_1 \odot E_2$. The following lemma verifies that if $E_1 \odot E_2$, then we can reach E_2 from E_1 by a short sequence of applications of the rewriting transformation $\Rightarrow$, or equivalently, that there is a short path from E_1 to E_2 in $\mathscr{G}^{\Rightarrow}_{[E_1]}$.

Lemma 7.2 *Let $x, y, u, v \in X$ and $\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3 \in (X\backslash\{x, y, u, v\})^*$. Let E_1 be the basic RWE given by $x\alpha_1 u\alpha_2 v\alpha_3 y \doteq y\beta_1 u\beta_2 v\beta_3 x$ and let E_2 be the basic RWE given by $x\alpha_1 v\alpha_3 u\alpha_2 y \doteq y\beta_1 v\beta_3 u\beta_2 x$. Then there exist $n_1, n_2 < 4|E_1|$ such that $E_1 \Rightarrow^{n_1} E_2$ and $E_2 \Rightarrow^{n_2} E_1$.*

Proof Let E_3, E_4, E_5 be the equations given as follows:

$$\begin{aligned} E_3 &: \quad v\alpha_3 x\alpha_1 u\alpha_2 y \doteq y\beta_1 u\beta_2 v\beta_3 x \\ E_4 &: \quad x\alpha_1 v\alpha_3 u\alpha_2 y \doteq u\beta_2 y\beta_1 v\beta_3 x \\ E_5 &: \quad v\alpha_3 x\alpha_1 u\alpha_2 y \doteq u\beta_2 y\beta_1 v\beta_3 x. \end{aligned}$$

Then it follows directly from the definitions that $E_1 \Rightarrow^*_R E_3 \Rightarrow^*_L E_5 \Rightarrow^*_R E_4 \Rightarrow^*_L E_2$. Thus, by Remark 3.2, there exists $n_1 < 4|E_1|$ such that $E_1 \Rightarrow^{n_1} E_2$. By the same remark, we know that $\Rightarrow^*_L, \Rightarrow^*_R$ are symmetric, and thus we may similarly conclude that $E_2 \Rightarrow^*_L E_4 \Rightarrow^*_R E_5 \Rightarrow^*_L E_3 \Rightarrow^*_R E_1$ so there exists $n_2 < 4|E_1|$ such that $E_2 \Rightarrow^{n_2} E_1$. □

Corollary 7.3 *Let E_1, E_2 be basic RWEs. If $E_1 \odot^m E_2$ for some $m \in \mathbb{N}$, then $E_1 \Rightarrow^n E_2$ for some $n \in O(|E_1|m)$.*

The first of our two normal forms is defined as follows. Theorem 7.5 confirms the desired property that any basic RWE E can be transformed into an equation $\overline{E}$ which is in normal form in a small (i.e. polynomial in $|E|$) number of rewriting steps.

Definition 7.4 (Normal Form) Let E be a basic RWE. Then E is in *normal form* if it can be written as $x\alpha_1\alpha_2, \ldots \alpha_n y \doteq y\alpha_1^R\alpha_2^R \ldots \alpha_n^R x$ where $x, y \in X, \alpha_i \in X^+$ for $1 \leq i \leq n$, and $|\alpha_i| \leq 3$ for $1 \leq i < n$.

Theorem 7.5 *Let E be a jumbled basic RWE. Then there exists $\overline{E}$ which is in normal form and such that $E \Rightarrow^{n_1} \overline{E}$ and $\overline{E} \Rightarrow^{n_2} E$ for some $n_1, n_2 \in O(|E|^3)$.*

The main step in the proof of Theorem 7.5 is the following lemma, which we shall make use of again later and is therefore stated independently.

Lemma 7.6 *Let E be a jumbled basic RWE of the form $x\gamma_1\beta_1 y \doteq y\gamma_2\beta_2 x$ where $x, y \in X$, $\gamma_1, \gamma_2, \beta_1\beta_2 \in (X\backslash\{x, y\})^*$ and $var(\gamma_1) = var(\gamma_2)$. Then at least one of the following two statements holds:*

1. *$\beta_1 = \beta_2^R$, or*
2. *there exists $\alpha \in var(\beta_1)^*$ with $1 \leq |\alpha| \leq 3$, $\eta_1, \eta_2 \in var(\beta_1)^*$ and $n \in O(|E|)$ such that $E \odot^n x\gamma_1\alpha\eta_1 y \doteq y\gamma_2\alpha^R\eta_2 x$.*

Proof Throughout this proof, we shall use the fact that $E_1 \xrightarrow{u,v} E_2$ implies $E_1 \odot E_2$ and shall use the two notations interchangeably where convenient. Let E be a jumbled basic RWE of the form $x\gamma_1\beta_1 y \doteq y\gamma_2\beta_2 x$ where $x, y \in X$, $\gamma_1, \gamma_2, \beta_1\beta_2 \in (X\backslash\{x, y\})^*$ and $var(\gamma_1) = var(\gamma_2)$. Suppose that $\beta_1 \neq \beta_2^R$. Note that since E is basic and regular, $var(\beta_1) = var(\beta_2)$, and moreover we have that $\beta_1, \beta_2 \neq \varepsilon$. Hence we may write E in the form:

$$x\gamma_1 u\delta_1 y \doteq y\gamma_2\delta_2 u\delta_3 x \tag{1}$$

where $u \in X$, and $\delta_1, \delta_2, \delta_3 \in X^*$ such that $u\delta_1 = \beta_1$ and $\delta_2 u\delta_3 = \beta_2$. If $\delta_2 = \varepsilon$ then we can set $\alpha = u$ and we are done. Otherwise, the next step is to show that we can get to an equation of the form

$$x\gamma_1 u\delta_1' z_1 z_2 \ldots z_k \delta_2' y \doteq y\gamma_2 z_k z_{k-1} \ldots z_1 u\delta_3' x \tag{2}$$

where $z_1, z_2, \ldots, z_k \in X$, $\delta_1', \delta_2', \delta_3' \in X^*$. Suppose that our equation of the form (1) is not already of the form (2). Suppose firstly that there exist $v_1, v_2 \in X$ such that v_1 and v_2 occur in the same order in δ_1 and δ_2. In other words, suppose there exist $\delta_{1,1}, \delta_{1,2}, \delta_{1,3}, \delta_{2,1}, \delta_{2,2}, \delta_{2,3} \in X^*$ such that we can write $\delta_1 = \delta_{1,1}v_1\delta_{1,2}v_2\delta_{1,3}$ and $\delta_2 = \delta_{2,1}v_1\delta_{2,2}v_2\delta_{2,3}$. Then we have that $E \xrightarrow{v_1,v_2} E_{1,2}$ where $E_{1,2}$ is given by $x\gamma_1 u\hat{\delta_1} y \doteq y\gamma_2\hat{\delta_2} u\hat{\delta_3} x$ such that $|\hat{\delta_2}| < |\delta_2|$, with $\hat{\delta_1} = \delta_{1,1}v_2\delta_{1,3}v_1\delta_{1,2}$, $\hat{\delta_2} = \delta_{2,1}v_2\delta_{2,3}$ and $\hat{\delta_3} = \delta_3 v_1\delta_{2,2}$.

Iterating this, we may thus conclude that there exists $n_1 \leq |\delta_2|$ and a sequence $E = E_{1,1} \odot E_{1,2} \odot \ldots \odot E_{1,n_1}$ such that E_{1,n_1} has the form

$$x\gamma_1 u\hat{\delta}_1' y \doteq y\gamma_2 z_k z_{k-1} \ldots z_1 u\hat{\delta}_3' x$$

where $z_1, z_2, \ldots, z_k \in X$ and $\hat{\delta}_1' \in X^* z_1 X^* z_2 X^* \ldots X^* z_k X^*$. If all the internal X^* factors are the empty word (i.e. if $\hat{\delta}_1' \in X^* z_1 z_2 \ldots z_k X^*$), then E_{1,n_1} already has the

desired form described by (2). Otherwise, there exists $w \in X\backslash\{z_1, z_2, \ldots, z_k\}$ such that w occurs between z_1 and z_k in $\hat{\delta}'_1$. More precisely, we can write E_{1,n_1} as:

$$x\gamma_1 u\hat{\delta}_{1,1}z_1\hat{\delta}_{1,2}w\hat{\delta}_{1,3}z_k\hat{\delta}_{1,4}y \doteq y\gamma_2 z_k\hat{\delta}_{2,1}z_1\hat{\delta}_{3,1}w\hat{\delta}_{3,2}x$$

where $\hat{\delta}_{1,1}, \hat{\delta}_{1,2}, \hat{\delta}_{1,3}, \hat{\delta}_{1,4}, \hat{\delta}_{2,1}, \hat{\delta}_{3,1}, \hat{\delta}_{3,2} \in X^*$ such that $\hat{\delta}'_1 = \hat{\delta}_{1,1}z_1\hat{\delta}_{1,2}w\hat{\delta}_{1,3}z_k\hat{\delta}_{1,4}$, and $z_{k-1}z_{k-2}\ldots z_2 = \hat{\delta}_{2,1}$, and $\hat{\delta}'_3 = \hat{\delta}_{3,1}w\hat{\delta}_{3,2}$. In this case we have $E_{1,n_1} \xrightarrow{z_1,w} E_{2,1} \xrightarrow{z_k,z_1} E_{2,2}$ where $E_{2,1}$ is given by

$$x\gamma_1 u\hat{\delta}_{1,1}w\hat{\delta}_{1,3}z_k\hat{\delta}_{1,4}z_1\hat{\delta}_{1,2}y \doteq y\gamma_2 z_k\hat{\delta}_{2,1}w\hat{\delta}_{3,2}z_1u\hat{\delta}_{3,1}x$$

and $E_{2,2}$ is given by

$$x\gamma_1 u\hat{\delta}_{1,1}w\hat{\delta}_{1,3}z_1\hat{\delta}_{1,2}z_k\hat{\delta}_{1,4}y \doteq y\gamma_2 z_1u\hat{\delta}_{3,1}z_k\hat{\delta}_{2,1}w\hat{\delta}_{3,2}x$$

which is again of the desired form described by (2) for $k = 1$ and $z_1 = v_1$. In all cases, there exists $n_2 \leq |E|$ such that $E \odot^{n_2} E_{2,2}$ for some equation $E_{2,2}$ of the desired form (2).

Now suppose that $E_{2,2}$ has the form (2), and define $\delta'_1, \delta'_2, \delta'_3$ accordingly. Next, we note that there exists $n_3 \in \{0, 1\}$ such that $E_{2,2} \odot^{n_3} E_3$ where E_3 has the form

$$x\gamma_1 u'z_1z_2\ldots z_k\delta''_1 y \doteq y\gamma_2 z_k z_{k-1}\ldots z_1u'\delta''_2 x \tag{3}$$

where $u' \in X$ and $\delta''_1, \delta''_2 \in X^*$. Indeed, if $\delta'_1 = \varepsilon$, then this is trivial, simply taking $E_3 = E_{2,2}$. Otherwise, there exists $u' \in X$ and $\delta'_{1,1}, \delta'_{3,1}, \delta'_{3,2}$ such that $\delta'_1 = \delta'_{1,1}u'$ and $\delta'_3 = \delta'_{3,1}u'\delta'_{3,2}$. Then $E_{2,2}$ may be written as:

$$x\gamma_1 u\delta'_{1,1}u'z_1z_2\ldots z_k\delta'_2 y \doteq y\gamma_2 z_k z_{k-1}\ldots z_1u\delta'_{3,1}u'\delta'_{3,2}x$$

and $E_{2,2} \xrightarrow{u,u'} E_3$ where E_3 is given by

$$x\gamma_1 u'z_1z_2\ldots z_k\delta'_2 u\delta'_{1,1}y \doteq y\gamma_2 z_k z_{k-1}\ldots z_1u'\delta'_{3,2}u\delta'_{3,1}x$$

which is of the form (3) as required. Now, if $k \leq 2$, we may take $\alpha = u'z_1z_2\ldots z_k$, $\eta_1 = \delta'_2 u\delta'_{1,1}$ and $\eta_2 = \delta'_{3,2}u\delta'_{3,1}$ and we are done. Suppose otherwise that $k \geq 3$. Next, we observe that if $u\delta'_{1,1}$ and $u\delta'_{3,1}$ share a non-empty suffix, then we have an equation of the form $x\ldots sy \doteq y\ldots sx$. However, this implies that $(s, s) \in \Upsilon_{E_3}$, and by Theorem 5.3, $\Upsilon_{E_3} = \Upsilon_E$, meaning that E is not jumbled: a contradiction. Consequently, there must exist $s, t \in X$ with $s \neq t$ and $\beta'_{1,1}, \beta'_{1,2}, \beta'_{2,1}, \beta'_{2,2} \in X^*$ such that E_3 has the form

$$x\gamma_1 u'z_1z_2\ldots z_k\beta'_{1,1}s\beta'_{1,2}ty \doteq y\gamma_2 z_k z_{k-1}\ldots z_1u'\beta'_{2,1}t\beta'_{2,2}sx.$$

Then we have $E_3 \xrightarrow{z_2,s} E_{4,1} \xrightarrow{z_1,t} E_{4,2} \xrightarrow{z_k,z_1} E_{4,3} \xrightarrow{u',z_{k-1}} E_{4,4}$ where $E_{4,1}, E_{4,2}, E_{4,3}, E_{4,4}$ are given as follows:

$$\begin{aligned}
E_{4,1}: &\quad x\gamma_1 u'z_1s\beta'_{1,2}tz_2\ldots z_k\beta'_{1,1}y \doteq y\gamma_2 z_k z_{k-1}\ldots z_3sz_2z_1u'\beta'_{2,1}t\beta'_{2,2}x\\
E_{4,2}: &\quad x\gamma_1 u'tz_2\ldots z_k\beta'_{1,1}z_1s\beta'_{1,2}y \doteq y\gamma_2 z_k z_{k-1}\ldots z_3sz_2t\beta'_{2,2}z_1u'\beta'_{2,1}x\\
E_{4,3}: &\quad x\gamma_1 u'tz_2\ldots z_{k-1}z_1s\beta'_{1,2}z_k\beta'_{1,1}y \doteq y\gamma_2 z_1u'\beta'_{2,1}z_k z_{k-1}\ldots z_3sz_2t\beta'_{2,2}x\\
E_{4,4}: &\quad x\gamma_1 z_{k-1}z_1s\beta'_{1,2}z_k\beta'_{1,1}u'tz_2\ldots z_{k-2}y \doteq y\gamma_2 z_1z_{k-1}z_{k-2}\ldots z_3sz_2t\beta'_{2,2}u'\beta'_{2,1}z_kx.
\end{aligned}$$

Now, $E_{4,4}$ has the required form with $\alpha = z_{k-1}z_1$, $\eta_1 = s\beta'_{1,2}z_k\beta'_{1,1}u't z_2 \dots z_{k-2}$ and $\eta_2 = z_{k-2}\dots z_3 s z_2 t \beta'_{2,2}u'\beta'_{2,1}z_k$. Moreover, we have that $E \odot^n E_{4,4}$ with $n \leq n_2 + n_3 + 4 \leq |E| + 5 \in O(|E|)$ as claimed. □

We can now prove Theorem 7.5 with a simple induction based on Lemma 7.6.

Theorem 7.5. By Lemma 6.13, we have that $E \Rightarrow^{n_1} E'$ and $E' \Rightarrow^{n'_1} E$ where E' is a basic regular equation of the form $x\beta_1 y \doteq y\beta_2 x$ such that $x, y \in X$ and $\beta_1, \beta_2 \in (X\backslash\{x, y\})^*$ with $n_1, n'_1 \in O(|E|^2)$. By Theorem 5.3, since E is jumbled, E' is also jumbled. By a simple induction using Lemma 7.6 (starting with the case that $\gamma_1 = \gamma_2 = \varepsilon$) we can therefore infer that $E' \odot^{n_2} \overline{E}$ for some $\overline{E}$ in normal form and $n_2 \in O(|E|^2)$. It follows directly from the definitions that $\odot$, is symmetric, so we also have that $\overline{E} \odot^{n_2} E'$. Thus, by Corollary 7.3 we have that $E' \Rightarrow^{n_3} \overline{E}$ and $\overline{E} \Rightarrow^{n'_3} E'$ for some $n_3, n'_3 \in O(|E|^3)$, and therefore also that $E \Rightarrow^{n} \overline{E}$ and $\overline{E} \Rightarrow^{n'} E$ for some $n, n' \in O(|E|^3)$ as claimed. □

The idea behind the first normal form is to divide the RWE into pairs (α_i, α_i^R) which are regular-reversed word equations (although solutions to the full equation E are not necessarily solutions to these smaller equations), and for which all but one belong to a finite number of cases (i.e. three cases depending on the length of α_i). Forcing the sub-equations to be regular-reversed gives us the most control when working with the invariant Υ_E. Some intuition behind this fact can be derived from the observation that if we know that a (complete) basic RWE E is regular-reversed, we can uniquely reconstruct it from the leftmost two variables on the LHS and Υ_E. Indeed, any regular-reversed basic RWE E can be written in the form $x_1x_2\dots x_n \doteq x_n x_{n-1}\dots x_1$, meaning that $\Upsilon_E = \{(x_{i-1}, x_{i+1}) \mid 2 \leq i \leq n\} \cup \{(x_{n-1}, x_2)\}$, and if we know x_1, then we may infer from Υ_E all the odd-index variables $(x_3, x_5, \dots)$ and if we know x_2 then we may infer all the even-index variables $(x_4, x_6, \dots)$.

Rather than looking at the pairs (α_i, α_i^R) in isolation, in order to take full advantage of the invariant Υ_E, we actually need to consider pairs of the form

$$(\alpha_i\alpha_{i+1}\dots\alpha_j, \alpha_i^R\alpha_{i+1}^R\dots\alpha_j^R)$$

for well-chosen values i and j. We shall call such pairs *blocks*, which we define formally below.

Definition 7.7 (Blocks) We define 3 variations of blocks which may each have up to two types.

1. A *standard block* is a pair $(\alpha_1\alpha_2\dots\alpha_j, \alpha_1^R\alpha_2^R\dots\alpha_j^R)$ such that $j \geq 1$, $\alpha_i \in X^*$ for $1 \leq i \leq j$, $|\alpha_1| \in \{1, 3\}$, and for each i, $1 < i \leq j$, $|\alpha_i| = 2$. It is *Type A* if $|\alpha_1| = 1$ and *Type B* if $|\alpha_1| = 3$.
2. An *initial block* is a pair $(x\alpha_1\dots\alpha_j, y\alpha_1^R\dots\alpha_j^R)$ with $j \geq 0$, $x, y \in X$ with $x \neq y$, and $\alpha_i \in (X\backslash\{x, y\})^*$ where $|\alpha_i| = 2$ for $1 \leq i \leq j$. All initial blocks are *Type A*.
3. A *final block* is a pair $(\gamma_1\delta y, \gamma_2\delta^R x)$ where $x, y \in X$ with $x \neq y$, and $\gamma_1, \gamma_2, \delta \in X^*$ with $|\delta| \geq 1$ such that (γ_1, γ_2) is a block (initial or standard). It is *Type A* if (γ_1, γ_2) is Type A, and *Type B* otherwise.

Given an equation which is in normal form, we may decompose it uniquely into blocks in the following manner. The intuition behind this decomposition is that if we fix the invariant property Υ_E, then each block (with the exception of the final block) is determined entirely by the block preceding it along with its first (leftmost in the first element) variable. This gives us a crucial degree of control when considering which equations in normal form may appear in $\mathscr{G}^{\Rightarrow}_{[E]}$.

Definition 7.8 (Block Decomposition) Let E be a basic RWE in normal form. Then E may be written as $x\alpha_1\alpha_2\ldots\alpha_n y \doteq y\alpha_1^R\alpha_2^R\ldots\alpha_n^R x$ where $x, y \in X, \alpha_i \in X^+$ for $1 \leq i \leq n$, and $|\alpha_i| \leq 3$ for $1 \leq i < n$. Let $I = \{i_1, i_2, \ldots, i_k\} = \{i \mid 1 \leq i < n \text{ and } |\alpha_i| \neq 2\}$ with $1 \leq i_1 < i_2 < \ldots < i_k < n$. If $I = \emptyset$, let $\mathfrak{B} = (E)$. Otherwise, let $\mathfrak{B} = (B_0, B_1, \ldots, B_k)$ where for $0 \leq j \leq k$, the B_j are blocks such that:

1. $B_0 = (x\alpha_1\ldots\alpha_{i_1-1}, y\alpha_1^R\ldots\alpha_{i_1-1}^R)$,
2. $B_k = (\alpha_{i_k}\ldots\alpha_n y, \alpha_{i_k}^R\ldots\alpha_n^R x)$, and
3. for $1 \leq j < k$, $B_j = (\alpha_{i_j}\ldots\alpha_{i_{j+1}-1}, \alpha_{i_j}^R\ldots\alpha_{i_{j+1}-1}^R)$.

Then $\mathfrak{B}$ is the *block decomposition* of E.

As an example, consider the basic RWE E given as follows:

$$x\overbrace{z_1z_2}^{\alpha_1}\overbrace{z_3}^{\alpha_2}\overbrace{z_4z_5z_6}^{\alpha_3}\overbrace{z_7z_8}^{\alpha_4}\overbrace{z_9}^{\alpha_5}\overbrace{z_{10}z_{11}z_{12}z_{13}}^{\alpha_6}y \doteq y\overbrace{z_2z_1}^{\alpha_1^R}\overbrace{z_3}^{\alpha_2^R}\overbrace{z_6z_5z_4}^{\alpha_3^R}\overbrace{z_8z_7}^{\alpha_4^R}\overbrace{z_9}^{\alpha_5^R}\overbrace{z_{13}z_{12}z_{11}z_{10}}^{\alpha_6^R}x$$

Note that E is in normal form. Then $I = \{2, 3, 5\}$ and the block decomposition of E is (B_0, B_1, B_2, B_3) where:

$$\begin{aligned} B_0 &= (xz_1z_2,\ yz_2z_1) \\ B_1 &= (z_3,\ z_3) \\ B_2 &= (z_4z_5z_6z_7z_8,\ z_6z_5z_4z_8z_7) \\ B_3 &= (z_9z_{10}z_{11}z_{12}y,\ z_9z_{12}z_{11}z_{10}x). \end{aligned}$$

Another example illustrating the block decomposition of an equation in normal form is given in Fig. 6. The next fact follows directly from the definitions.

Fact 7.9 For every basic RWE in normal form, there exists a unique block decomposition $(B_0, B_1, \ldots, B_k)$ where $k \leq \text{Card}(var(E))$, B_k is a final block, and if $k > 0$, then B_0 is an initial block.

Since the blocks are fixed by their first variable, it is natural to ask for which variables we can find an equation in our graph $\mathscr{G}^{\Rightarrow}_{[E]}$ such that the block begins with that variable. In particular, can we find an equation in normal form in $\mathscr{G}^{\Rightarrow}_{[E]}$ for which the first variable of each block is lexicographically minimal when reading from left to right? The answer to the question is "nearly". In other words, if we relax the notion slightly to account for some specific exceptions, then we can always guarantee the existence of such an equation. This leads to the notion of Lex Normal Form defined below.

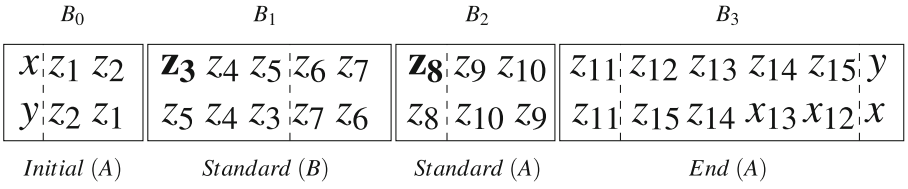

Fig. 6 A depiction of the equation E given by $xz_1z_2z_3z_4z_5z_6z_7z_8z_9z_{10}z_{11}z_{12}z_{13}z_{14}z_{15}y \doteq yz_2z_1z_5z_4z_3z_7z_6z_8z_{10}z_9z_{11}z_{15}z_{14}z_{13}z_{12}x$ where x, y and z_i for $1 \leq i \leq 15$ are variables. The LHS and RHS of the equation are aligned vertically. The block decomposition $\mathfrak{B} = (B_0, B_1, B_2, B_3)$ of E is shown with solid rectangles and with the variety and type of the block written beneath. The additional divisions into the factors α_i, α_i^R required by the definition of normal form are indicated by dashed lines (so that, i.e. $\alpha_1 = z_1z_2$, $\alpha_2 = z_3z_4z_5$, $\alpha_3 = z_6z_7$, $\alpha_4 = z_8, z_5$, $\alpha_5 = z_9z_{10}$, $\alpha_6 = z_{11}$ and $\alpha_7 = z_{12}z_{13}z_{14}z_{15}$). In order for the equation to satisfy the definition of Lex Normal Form, the variables highlighted in bold must be lexicographically minimal with respect to the appropriate sets Γ_i^E. For $i = 1$, we have that $\Gamma_1^E = \{z_i \mid 3 \leq i \leq 15\} \backslash \{z_4\}$. In particular, Γ_1^E consists of the first variable in the block B_1 (x_3) along with (nearly) all variables on the LHS of the equation occurring to the right of z_3, excluding the rightmost variable (y), and since B_1 is Type B, also excluding the second variable in the block B_1 (namely z_4). On the other hand, since B_2 is Type A, for $i = 2$, we do not need to exclude the second variable in the block B_2, so $\Gamma_2^E = \{z_i \mid 8 \leq i \leq 15\}$. Assuming an underlying lexicographic order for which z_{i+1} is greater than z_i, we can conclude that E is in Lex Normal Form

Definition 7.10 (Lex Normal Form) Let E be a basic RWE in normal form. Then there exist $x, y \in X$ and $\alpha, \beta \in (X \backslash \{x, y\})^*$ such that E has the form $x\alpha y \doteq y\beta x$. Let $(B_0, B_1, \ldots, B_k)$ be the block decomposition of E. For each i, $0 \leq i \leq k$, let $\gamma_i, \gamma_i' \in X^*$ such that $B_i = (\gamma_i, \gamma_i')$, let $S_i = \{\gamma_i[2], y\}$ whenever B_i is Type B and $S_i = \{y\}$ otherwise, and let $\Gamma_i^E = \left(\bigcup_{i \leq j \leq k} var(\gamma_j) \right) \backslash S_i$. A block B_i is *lex-minimal* if $\gamma_i[1]$ is lexicographically minimal in Γ_i^E. The equation E is in Lex Normal Form (LNF) if, for each i, $0 < i < k$, B_i is lex-minimal.

Lex Normal Form (see also Fig. 6 for an example) describes the class of equations for which the first variable of each blocks is lexicographically minimal *whenever possible*. We can, in general, guarantee the existence of an equation E' in $\mathscr{G}_{[E]}^{\Rightarrow}$ such that the first variable of each block is lexicographically minimal with the following exceptions. Firstly, we must exclude the first and last blocks (the first block is fixed completely by Υ_E). Secondly, we must only compare the first variable to other variables occurring further right in the LHS of the equation, and excluding the rightmost variable on the LHS of the equation (y in the definition above) and, for blocks of Type B, the second variable in the block. The sets Γ_i^E in the definition account for these exclusions.

The main result of this section is that every vertex in $\mathscr{G}_{[E]}^{\Rightarrow}$ is never more than a polynomial distance away from a vertex corresponding to an equation in LNF.

Theorem 7.11 *Let E be a jumbled basic RWE. Then there exists E' such that E' is in Lex Normal Form, and such that $E \Rightarrow^{n_1} E'$ and $E' \Rightarrow^{n_2} E$ for some $n_1, n_2 \in O(|E|^4)$.*

Although Theorem 7.11 does not provide as detailed a description of the graphs $\mathscr{G}^{\rightrightarrows}_{[E]}$ in the jumbled case as Theorem 6.8 does in the non-jumbled case, it does allow us to study them as the polynomial-distance neighbourhoods of the highly restricted set of vertices corresponding to equations in Lex Normal Form. Section 8 gives a strong example of the benefits of this approach, allowing us to show firstly that the cardinality of the set of vertices in Lex Normal Form is bounded by a polynomial in $|E|$ (in contrast to the fact that the total number of vertices will typically be exponential, as shown in Section 9), and consequently, that the diameter of $\mathscr{G}^{\rightrightarrows}_{[E]}$ is also bounded by a polynomial in $|E|$.

Proof of Theorem 7.11 The rest of this section is devoted to proving Theorem 7.11. To do so, we essentially provide a strategy for rewriting any jumbled basic regular word equation E into an equation in Lex Normal Form. The overall structure is similar to that of Theorem 7.5 in the sense that we transform the equation in steps from left to right so that after each step, the prefixes of the LHS and RHS having the desired form are longer. Since each side of the equation stays the same length under the transformations, we eventually reach a state where the entire equation is in the correct form.

The first step in this strategy is to first ensure that E is in normal form (which we can do due to Theorem 7.5). We can then decompose E into blocks according to Definition 7.8 (see also Fig. 6). In each subsequent step, we apply transformations which increase the number of blocks satisfying the requirements for Lex Normal Form. In particular, if the first j blocks satisfy the requirements for Lex Normal Form, then we apply a sequence of transformations which either preserve the first $j-1$ blocks and turn the j^{th} block into a final block, or which preserves the first j blocks, and which result in an equation which is also in normal form, and for which the $j+1^{th}$ block also satisfies the requirements for Lex Normal Form. Note that Lex Normal Form does not impose any additional constraints on the initial or final blocks, so we can start with $j=1$ and we are done whenever we produce a final block.

There are two cases depending on whether the $j+1^{th}$ block is Type A or Type B. The case that it is Type A is substantially the easier of the two and is considered directly in the proof of Lemma 7.17. Lemmas 7.12-7.16 focus on the case that the block is Type B. In this case, there exist $x, y, a, b, c, \in X$ and $\mu_1, \mu_1', \mu_2, \mu_2' \in X^*$ such that our equation may be written as

$$x\mu_1 abc\mu_2 y \doteq y\mu_1' cba\mu_2' x$$

where $var(\mu_1) = var(\mu_1')$ and $var(\mu_2) = var(\mu_2')$, the prefixes $x\mu_1$ and $y\mu_1'$ constitute the first j blocks (the ones satisfying the requirements for LNF), and such that the $j+1^{th}$ block, which does not satisfy the requirements for LNF, has the form $(abc\gamma, cba\gamma')$ for prefixes γ, γ' of μ_2, μ_2' respectively. Our aim is to transform the equation above into an equation either of the form:

$$x\mu_1\beta y \doteq y\mu_1'\beta^R x$$

in which case the j^{th} block becomes final (and all other blocks are preserved), or of the form:

$$x\mu_1 zbw\eta y \doteq y\mu_1' wbz\eta' x$$

where $w, z \in X$ and $\eta, \eta' \in X^*$, such that either $\eta' = \eta^R$ (meaning $(zbw\eta y, wbz\eta' x)$ is a final block), or z is lexicographically minimal in $\Gamma^E_{j+1} = var(\mu_2) \cup \{a, c\}$.

In the case that $\eta' = \eta^R$, then the new equation is in normal form and will have a block decomposition with $j + 1$ blocks, such that the first j blocks are the same as before, and thus satisfy the requirements for LNF. The $j + 1^{th}$ block is final, and trivially satisfies the requirements for LNF, so the whole equation is in LNF. In the second case, we can apply Lemma 7.6 to further transform our equation into one in normal form without changing the prefixes $x\mu_1 zbw$ and $y\mu'_1 wbz$. In the resulting block decomposition, the first j blocks will remain unchanged, while the $j + 1^{th}$ block will have the form $(zbw\gamma, wbz\gamma')$ for some $\gamma, \gamma' \in {\Gamma^E_{j+1}}^*$. Since Γ^E_{j+1} will also remain unchanged, z is lexicographically minimal in $\Gamma^{E'}_{j+1}$ for our new equation E', so the $j + 1^{th}$ block also satisfies the requirements for LNF as intended.

The following lemma shows us how, under the rewriting transformation $\odot$, we can replace the factors abc and cba with factors dbe and ebd, providing that $d, e \in X$ occur in the appropriate positions (namely directly left of y and x on the LHS and RHS respectively).

Lemma 7.12 *Let E, E' be basic RWEs given by*

$$\begin{aligned} E: &\quad x\mu_1 abc\mu_2 d\mu_3 ey \doteq y\mu'_1 cba\mu'_2 e\mu'_3 dx \\ E': &\quad x\mu_1 ebd\mu_3 c\mu_2 ay \doteq y\mu'_1 dbe\mu'_3 a\mu'_2 cx \end{aligned}$$

where $x, y, a, b, c, d, e \in X$ and $\mu_1, \mu_2, \mu_3, \mu'_1, \mu'_2, \mu'_3 \in X^*$. Then $E \odot^3 E'$.

Proof It follows from the definitions that:

$$\begin{aligned} & \overbrace{x\ \mu_1\ abc\ \mu_2\ d\ \mu_3\ ey \doteq y\ \mu'_1\ cba\ \mu'_2\ e\ \mu'_3\ dx}^{E} \\ \xrightarrow{b,e} \quad & x\ \mu_1\ aebc\ \mu_2\ d\ \mu_3\ y \doteq y\ \mu'_1\ ce\ \mu'_3\ dba\ \mu'_2\ x \\ \xrightarrow{c,d} \quad & x\ \mu_1\ aebd\ \mu_3\ c\ \mu_2\ y \doteq y\ \mu'_1\ dba\ \mu'_2\ ce\ \mu'_3\ x \\ \xrightarrow{a,e} \quad & \underbrace{x\ \mu_1\ ebd\ \mu_3\ c\ \mu_2\ ay \doteq y\ \mu'_1\ dbe\ \mu'_3\ a\mu'_2 cx.}_{E'} \end{aligned}$$

Thus the statement follows by Lemma 7.2. □

Of course, the variable d occurring to the left of y on the RHS will in general not be the lexicographically minimal element z of Γ^E_{j+1}. In order to take advantage of Lemma 7.12, we also need to find a sequence of transformations which, for any $z \in \{c\} \cup var(\nu)$, results in an equation of the form $x\mu_1 a'bc'\eta zy \doteq y\mu'_1 c'ba'\eta' x$ with $a', c' \in X$ and $\eta, \eta' \in X^*$. To achieve this, we need Lemmas 7.13 and 7.14 as follows.

Lemma 7.13 *Let E be a basic RWE given by $x\mu_1\alpha\mu_2 y \doteq y\mu'_1\alpha^R\mu'_2 x$ with $\alpha, \mu_1, \mu_2, \mu'_1, \mu'_2 \in X^*$, $2 \leq |\alpha| \leq 3$, $|\mu_2| \geq 1$ and $var(\mu_2) = var(\mu'_2)$. Let*

$v = \alpha[|\alpha| - 1]$. *Then for each* $z \in var(\alpha\mu_2)\backslash\{v\}$, *there exists* $n \leq 3$ *and* $\eta, \eta' \in X^*$ *such that* $E \odot^n x\mu_1\eta zy \doteq y\mu_1'\eta' x$.

Proof Let $z \in var(\alpha\mu_2)\backslash\{v\}$. If z is a suffix of μ_2 then the statement holds trivially. Suppose that z is not a suffix of μ_2. We shall consider two cases separately. Firstly, suppose that $z \in var(\mu_2) \cup \{\alpha[|\alpha|]\}$. Then there exists $w \in X$ such that zw is a factor of $\alpha\mu_2$. Moreover, $w \in var(\mu_2) = var(\mu_2')$, so there exist $\nu_1, \nu_2, \nu_1', \nu_2' \in X^*$ such that $\mu_2 = \nu_1 w \nu_2$ and $\mu_2' = \nu_1' w \nu_2'$ where $\nu_1 = \varepsilon$ if $z = \alpha[|\alpha|]$, and $\nu_1[|\nu_1|] = z$ otherwise. Furthermore, there exists $u \in X$ and $\alpha' \in X^*$ such that $\alpha = u\alpha'$ and $\alpha^R = \alpha'^R u$. Thus we may write E as $x\mu_1 u\alpha'\nu_1 w\nu_2 y \doteq y\mu_1'\alpha'^R u\nu_1' w\nu_2' x$, and thus $E \xrightarrow{u,w} x\mu_1 w\nu_2 u\alpha'\nu_1 y \doteq y\mu_1'\alpha'^R w\nu_2' u\nu_1' x$. Since z is a suffix of $\alpha'\nu_1$, the statement of the lemma follows.

Now suppose that $z \notin var(\mu_2) \cup \{\alpha[|\alpha|]\}$. Then the only possibility is that $|\alpha| = 3$ and $z = \alpha[1]$. In this case, due to the fact that $\odot$ is symmetric, the statement follows directly from Lemma 7.12. □

Lemma 7.14 *Let* E *be a basic RWE given by* $x\mu_1 v\mu_2 y \doteq y\mu_1' v\mu_2' x$ *with* $v \in X$ *and* $\mu_1, \mu_2, \mu_1', \mu_2' \in X^*$ *such that* $var(\mu_2) = var(\mu_2')$. *Then for every* $z \in var(v\mu_2)$, *there exist* $v' \in X$ *and* $\eta, \eta' \in X^*$ *and* $n \leq 1$ *such that* $E \odot^n x\mu_1 v'\eta zy \doteq y\mu_1' v'\eta' x$.

Proof Let $z \in var(v\mu_2)$. If z is a suffix of μ_2, then the statement holds trivially. Otherwise, there exists $w \in X$ such that zw is a factor of $v\mu_2$. Moreover, since $w \neq v$, $w \in var(\mu_2) = var(\mu_2')$, so there exist $\nu_1, \nu_2, \nu_1', \nu_2' \in X^*$ such that $v\mu_2 = \nu_1 w\nu_2$ and $v\mu_2' = \nu_1' w\nu_2'$. Thus we may write E as $x\mu_1\nu_1 w\nu_2 y \doteq y\mu_1'\nu_1' w\nu_2' x$ such that v is a prefix of ν_1 and ν_1', and such that z is a suffix of ν_1. Thus, $E \xrightarrow{v,w} x\mu_1 w\nu_2\nu_1 y \doteq y\mu_1' w\nu_2'\nu_1' x$, and since z is a suffix of ν_1, the statement of the lemma follows. □

Recall that our strategy for transforming an equation of the form $x\mu_1 abc\mu_2 y \doteq y\mu_1' cba\mu_2' x$ into one of the form $x\mu_1 zbw\eta y \doteq y\mu_1' wbzv' x$ is first to 'move' the lexicographically minimal variable z from Γ_{j+1}^E into the correct position (to the left of y on the LHS) and then to apply Lemma 7.14. We can consider three cases for z separately. The first, that $z = a$ is trivial, and we do not need to change our original equation at all. The case that $z = c$ is the most involved and is considered in the proof of Lemma 7.16. All other choices of z (namely when $z \in var(\mu_2)$), are addressed in Lemma 7.15 below.

Note that in the statement of Lemma 7.15, the factors μ_2, μ_2' are replaced by $\mu_2\delta$ and $\mu_2'\delta^R$ respectively. We may make this change w.l.o.g. since our equation is in normal form, and since the case that $\mu_2 = \mu_2' = \varepsilon$ is trivial (the j^{th} block will be final in this case). Moreover, if $|\delta| = 1$, then $(\delta, \delta) \in \Upsilon_E$, so it follows from the definitions that the equation is not jumbled. Since we are only interested in this section in jumbled equations, we may therefore also assume that $|\delta| \geq 2$, which is necessary for the proof of the lemma.

Lemma 7.15 *Let E be a basic RWE in normal form given by*

$$x\mu_1 abc\mu_2\delta y \doteq y\mu_1' cba\mu_2'\delta^R x$$

with $a, b, c \in X$ and $\delta, \mu_1, \mu_2, \mu_1', \mu_2' \in X^*$ such that $|\delta| \geq 2$, and $var(\mu_2) = var(\mu_2')$. Then at least one of the following two statements is true.

1. *There exist $n \in O(|E|)$, $a', c' \in X$, and $\beta \in X^+$ such that $E \odot^n x\mu_1 a'bc'\beta y \doteq y\mu_1' c'ba'\beta^R x$, or*
2. *for every $z \in var(\mu_2\delta)$, there exist $a', c' \in X$, $\eta, \eta' \in X^*$, and $n \in O(|E|^2)$ such that $E \odot^n x\mu_1 a'bc'\eta z y \doteq y\mu_1' c'ba'\eta' x$.*

Proof Suppose that the first statement does not hold and notice that this implies $|\mu_2| \geq 1$. We shall now prove that the second statement holds. We divide our reasoning into three cases based on the prefixes of μ_2 and μ_2'. In particular, since E is in normal form, there exists a prefix α_i of μ_2 such that α_i^R is a prefix of μ_2' and such that $1 \leq |\alpha_i| \leq 3$. Firstly suppose that $|\alpha_i| = 1$, or in other words that μ_2 and μ_2' have a common prefix $v \in X$. Then the statement follows directly from Lemma 7.14.

It remains to consider the cases that $|\alpha_i| = 2$ and $|\alpha_i| = 3$. Before we consider these cases explicitly, it is convenient to define the following equation E' such that $E \odot^{n'} E$; for some $n' \in O(|E|)$. In particular, note that there exist $u, v \in X$ such that $\delta = u\delta' v$. It follows by Lemma 7.12 that there exist $\nu_1, \nu_1 \in X^+$ with $var(\nu_1) = var(\nu_1')$ such that $E \odot^3 x\mu_1 vbu\nu_1 y \doteq y\mu_1' ubv\nu_1' x$. Moreover, by Lemma 7.6, there exist $\nu_2, \nu_2' \in X^*$, $\beta \in X^+$ and $n' \in O(|E|)$ such that $E \odot^{n'} E'$ where E' is given by

$$E': \quad x\mu_1 vbu\beta\nu_2 y \doteq y\mu_1' ubv\beta^R\nu_2' x$$

where $1 \leq |\beta| \leq 3$ (recall by our assumption that the first statement of the lemma does not hold, that $\nu_2 \neq \varepsilon$). Note that since $E \odot^* E'$, we have $E \Rightarrow^* E'$ and thus by Theorem 5.3, $\Upsilon_{E'} = \Upsilon_E = \Upsilon$.

We are now ready to consider the second case, that $|\alpha_i| = 2$. In this case, there exist $d, e \in X$ such that $\alpha_i = de$, so de is a prefix of μ_2 and ed is a prefix of μ_2'. If $z \in var(\mu_2\alpha)\backslash\{d\}$, then the second statement of the lemma follows directly from Lemma 7.13. Suppose instead that $z = d$. In this case, we shall show that the (second statement of the) lemma holds for E'. Since $E \odot^{n'} E'$, it follows that the lemma also holds for E.

If $|\beta| = 1$, then the second statement of the lemma follows from Lemma 7.14 along with the fact that $E \odot^{n'} E'$. Similarly, if $|\beta| \in \{2, 3\}$ and $z \neq \beta[|\beta| - 1]$, the statement follows from Lemma 7.13. Finally, we must consider the case that $\beta \in \{2, 3\}$ and $z = \beta[|\beta| - 1]$. If $|\beta| = 2$, then there exists $z' \in X$ such that $\beta = zz'$. It follows that zz' is a factor of the LHS of E' and vz' is a factor of the RHS of E', so $(z, v) \in \Upsilon$. Furthermore, by our assumption that $z = d$, $ze = de$ is a factor of the LHS of E and ae is a factor of the RHS of E, so $(z, a) \in \Upsilon$. However, since $a \neq v$, this contradicts Remark 5.2. We can proceed similarly when $|\beta| = 3$. In particular, if $|\beta| = 3$, then there exist $z', z'' \in X$ such that $\beta = z'zz''$. It follows that $(z, v), (u, z) \in \Upsilon$. Furthermore, since $z = d$, we also have that $(z, a) \in \Upsilon$. However, since $v \neq a$ we again get a contradiction to Remark 5.2. Thus $d \neq \beta[|\beta| - 1]$ and we are done with the case that $|\alpha_i| = 2$.

Suppose now that $|\alpha_i| = 3$, meaning there exist $d, e, f \in X$ such that $\alpha_i = def$ is a prefix of μ_2 and fed is a prefix of μ_2'. As before, if $z \in var(\mu_2\alpha)\backslash\{e\}$, the second statement of the lemma follows from Lemma 7.13 (applied to E). Suppose instead that $z = e$. We shall again proceed by showing that the second statement of the lemma holds for E'. If $|\beta| = 1$, it follows directly from Lemma 7.14. Similarly, if $|\beta| \in \{2, 3\}$ and $z \neq \beta[|\beta| - 1]$, the statement again follows from Lemma 7.13. Finally, suppose for contradiction that $\beta \in \{2, 3\}$ and $z = \beta[|\beta| - 1]$. We again have to consider two cases based on $|\beta|$. If $|\beta| = 2$, then there exists $z' \in X$ such that $\beta = zz'$. It follows that $(z, v) \in \Upsilon$. Furthermore, since $z = e$, we also have that $(z, a) \in \Upsilon$, a contradiction to Remark 5.2. Similarly, if $|\beta| = 3$, then there exist $z', z'' \in X$ such that $\beta = z'zz''$. It follows that $(z, v), (u, z) \in \Upsilon$. Furthermore, since $z = e$, we also have that $(z, a), (c, z) \in \Upsilon$. However, since $u \neq c$, $v \neq a$ we again get a contradiction to Remark 5.2. Thus $d \neq \beta[|\beta| - 1]$ and the statement holds as required. □

We are now ready to prove the following lemma, which is the main technical step in the proof of Theorem 7.11, showing that we can replace the factors abc and cba at the start of the $j + 1^{th}$ block (which occur whenever the block is Type B) with factors zbw and wbz where z is any variable from Γ^E_{j+1}, and hence that we can do the same for the lexicographically minimal choice of z. This, combined with Lemma 7.6, allows us to transform the equation into one with the $j + 1^{th}$ block satisfying the requirements for Lex Normal Form.

It is also worth noting that the variable b and whether the block is Type A or Type B remain unchanged (see Section 8 for more information on why we cannot change them). Aside from these parameters, we can essentially produce all other possibilities for the variable in the first position in the block. In other words, we do not use anything about the lexicographic order other than it permits us to make some well-defined choice at each stage which is consistent across all equations. Consequently, there is a high degree of symmetry in the set of equations in normal form occurring in the graph $\mathscr{G}_{\overrightarrow{[E]}}$.

Lemma 7.16 *Let E be a basic RWE in normal form given by*

$$x\mu_1 abc\mu_2\delta y \doteq y\mu_1' cba\mu_2'\delta^R x$$

with $a, b, c \in X$ and $\delta, \mu_1, \mu_2, \mu_1', \mu_2' \in X^*$ such that $|\delta| \geq 2$ and $var(\mu_2) = var(\mu_2')$. Let $\Gamma = var(\mu_2\delta) \cup \{a, c\}$. Then at least one of the following two statements is true.

1. *There exist $n \in O(|E|)$, $a', c' \in X$, and $\beta \in X^+$ such that $E \odot^n x\mu_1 a'bc'\beta y \doteq y\mu_1' c'ba'\beta^R x$, or*
2. *for each $z \in \Gamma$, there exist $w \in X$, $\eta, \eta' \in X^*$, and $n \in O(|E|^2)$ such that $E \odot^n x\mu_1 zbw\eta y \doteq y\mu_1' wbz\eta' x$.*

Proof Assume that the first statement does not hold and notice that this implies $|\mu_2| \geq 1$. We shall now prove that the second statement holds. The case that $z = a$ is trivial. Next, consider the case that $z \notin \{a, c\}$. Then $z \in var(\mu_2\delta)$. By Lemma 7.15,

and by our assumption that Statement 1 of the lemma does not hold, we get that there exist $a', c' \in X$ $\nu, \nu' \in X^*$ and $n' \in O(|E|^2)$ such that

$$E \odot^{n'} x\mu_1 a' b c' \nu z y \doteq y\mu_1' c' b a' \nu' x.$$

Since E is basic and regular, and since $var(\mu_1) = var(\mu_1')$, we may conclude that $var(\nu') = var(\nu z)$. Thus, by Lemma 7.12, there exist $\eta, \eta' \in X^*$ such that

$$x\mu_1 a' b c' \nu z y \doteq y\mu_1' c' b a' \nu' x \odot^3 x\mu_1 z b w \eta y \doteq y\mu_1' w b z \eta' x$$

where $w = \nu'[|\nu'|]$. Consequently, we have that $E \odot^n x\mu_1 z b w \eta y \doteq y\mu_1' w b z \eta' x$ for some $n \in O(|E|)$ and the second statement holds as claimed.

It remains to consider the case that $z = c$. Then since $|\delta| \geq 2$, there exist $u, v \in X\backslash\{a, b, c\}$ such that $\delta = u\delta' v$ for some $\delta' \in X^*$. Thus, by Lemma 7.12, there exist $\nu_1, \nu_1' \in X^*+$ such that $E \odot^3 x\mu_1 v b u \nu_1 y \doteq y\mu_1' u b v \nu_1' x$. Moreover, since E is basic and regular, and since $var(\mu_1) = var(\mu_1')$, we may conclude that $var(\nu_1) = var(\nu_1')$. Thus, by Lemma 7.6, there exist $\nu_2, \nu_2' \in X^*$ and $\beta \in X^+$ and $n_1 \in O(|E|)$ such that $E \odot^{n_1} E'$ where E' is given by $x\mu_1 v b u \beta \nu_2 y \doteq y\mu_1' u b v \beta^R \nu_2' x$ and such that $1 \leq |\beta| \leq 3$ whenever $\nu_2 \neq \varepsilon$. By our assumption that the first statement of the lemma is not true, we must in fact have that $\nu_2 \neq \varepsilon$.

Additionally, note that $var(\nu_2) = var(\nu_2')$ and $c \in var(\beta\nu_2)$. Thus, by Lemma 7.15, along with our assumption that the first statement of the lemma does not hold, it follows that there exist $n_2 \in O(|E|^2)$, $a', c', d \in X$ and $\eta, \eta' \in X^*$ such that $E' \odot^{n_2} E''$ where E'' is given by $x\mu_1 a' b c' \eta c y \doteq y\mu_1' c' b a' \eta' d x$. As before, since E (and therefore also E'') is basic and regular, and since $var(\mu_1) = var(\mu_1')$, we may further conclude that $var(\eta c) = var(\eta' d)$. Similarly, since E is jumbled and $E \odot^* E''$ (meaning also that $E \Rightarrow^* E''$) it follows that E'' is also jumbled and consequently that $d \neq c$. Hence we may write E'' as $x\mu_1 a' b c' \eta_1 d \eta_2 c y \doteq y\mu_1' c' b a' \eta_1' c \eta_2' d x$ where $\eta_1, \eta_1', \eta_2, \eta_2' \in X^*$ and the second statement of the lemma follows from Lemma 7.12. □

Having described the main technical elements to the proof of Theorem 7.11, we are now ready to give the main intuitive statement as to why it holds, which also constitutes the main induction step, forming the backbone of the proof.

Lemma 7.17 *Let E be a jumbled basic RWE in normal form with block decomposition $(B_0, B_1, \ldots, B_k)$. Let $\iota \in \mathbb{N}$ with $0 < \iota < k$. Then at least one of the following two statements is true.*

1. *There exists a (final) block C_ι, $\hat{E} \in [E]_\Rightarrow$ and $n \in O(|E|)$ such that $E \odot^n \hat{E}$ and such that $\hat{E}$ has a block decomposition $(B_0, B_1, \ldots, B_{\iota-1}, C_\iota)$, or*
2. *there exist blocks $C_\iota, C_{\iota+1}, \ldots C_\ell$, $\hat{E} \in [E]_\Rightarrow$ and $n \in O(|E|^2)$ such that $E \odot^n \hat{E}$ and such that $\hat{E}$ has a block decomposition $(B_0, B_1, \ldots, B_{\iota-1}, C_\iota, C_{\iota+1}, \ldots C_\ell)$ and such that C_ι is lex-minimal.*

Proof Let E be given by

$$x\alpha_1\alpha_2 \ldots \alpha_m y \doteq y\alpha_1^R\alpha_2^R \ldots \alpha_m^R x$$

such that $x, y \in X$, $\alpha_i \in X^+$ for $1 \leq i \leq n$, and $|\alpha_i| \leq 3$ for $1 \leq i < m$. Let $I_E = \{i_1, i_2, \ldots, i_k\} = \{i \mid 1 \leq i < m \text{ and } |\alpha_i| \neq 2\}$ with $1 \leq i_1 < i_2 < \ldots < i_k < m$. If $I_E = \emptyset$, then the statement holds trivially. Thus we may assume that $I_E \neq \emptyset$. Note that the block decomposition $\mathfrak{B}$ of E is given by $(B_0, B_1, \ldots, B_k)$ where

$$\begin{aligned} B_0 &= (x\alpha_1\alpha_2 \ldots \alpha_{i_1-1}, y\alpha_1^R\alpha_2^R \ldots \alpha_{i_1-1}^R) \\ B_j &= (\alpha_{i_j}\alpha_{i_j+1} \ldots \alpha_{i_{j+1}-1}, \alpha_{i_j}^R\alpha_{i_j+1}^R \ldots \alpha_{i_{j+1}-1}^R) \\ B_k &= (\alpha_{i_k}\alpha_{i_k+1} \ldots \alpha_n y, \alpha_{i_k}^R\alpha_{i_k+1}^R \ldots \alpha_n^R x) \end{aligned}$$

for $0 < j < k$.

Now, let $\iota \in \mathbb{N}$ with $0 < \iota < k$. If B_ι is lex-minimal, the second statement holds trivially for $\ell = k$ and $C_j = B_j$ for $\iota \leq j \leq k$. Suppose instead that B_ι is not lex-minimal. We shall consider the cases that B_ι is Type A and Type B separately. Suppose firstly that B_ι is Type A. Then $|\alpha_{i_\iota}| = 1$. Thus we can write E as

$$x\mu_1 v\mu_2 y \doteq y\mu_1' v\mu_2' x$$

where $v = \alpha_{i_\iota} \in X$, $\mu_1 = \alpha_1\alpha_2 \ldots \alpha_{i_\iota-1}$, $\mu_1' = \alpha_1^R\alpha_2^R \ldots \alpha_{i_\iota-1}^R$, $\mu_2 = \alpha_{i_\iota+1}\alpha_{i_\iota+2} \ldots \alpha_m$ and $\mu_2' = \alpha_{i_\iota+1}^R\alpha_{i_\iota+2}^R \ldots \alpha_m^R$. Moreover, $\Gamma_\iota^E = var(v\mu_2)$. Let z be the lexicographically minimal element of Γ_ι^E. Then by our assumption that B_ι is not lex-minimal, we have that $z \neq v$. Thus there exist $\nu_1, \nu_2, \nu_1', \nu_2' \in X^*$ such that $\mu_2 = \nu_1 z \nu_2$ and $\mu_2' = \nu_1' z \nu_2'$. Consequently, $E \xrightarrow{v,z} x\mu_1 z\nu_2 v\nu_1 y \doteq y\mu_1' z\nu_2' v\nu_1' x$ and since $var(\mu_1 z) = var(\mu_1' z)$, by Lemma 7.6, we have that $E \odot^n E'$ where E' is given by:

$$x\alpha_1\alpha_2 \ldots \alpha_{i_\iota-1} z\alpha_{i_\iota+1}'\alpha_{i_\iota+2}' \ldots \alpha_{m'}' y \doteq y\alpha_1^R\alpha_2^R \ldots \alpha_{i_\iota-1}^R z\alpha_{i_\iota+1}'^R\alpha_{i_\iota+2}'^R \ldots \alpha_{m'}'^R x$$

for some $n \in O(|E|^2)$ and $\alpha_{i_\iota+1}', \alpha_{i_\iota+2}', \ldots, \alpha_{m'}' \in X^+$ with $1 \leq |\alpha_j'| \leq 3$ for $i_\iota + 1 \leq j < m'$. Let $I_{E'} = \{i_1', i_2', \ldots, i_\ell'\} = \{i \mid 1 \leq i < i_\iota \text{ and } |\alpha_i| \neq 2\} \cup \{i_\iota\} \cup \{i \mid i_\iota < i < m' \text{ and } |\alpha_i'| \neq 2\}$ with $1 \leq i_1' < i_2' < \ldots < i_\ell' < m$.

Let $\mathfrak{B}' = (B_0', B_1', \ldots, B_\ell')$ be the block decomposition of E'. Then since $I_E \cap \{1, 2, \ldots, i_\iota\} = I_{E'} \cap \{1, 2, \ldots, i_\iota\}$, we have $B_j = B_j'$ for $0 \leq j \leq \iota - 1$. Moreover, since z is minimal in $\Gamma_\iota^E = \Gamma_\iota^{E'}$, B_ι' is lex-minimal and the second statement holds.

Now suppose that B_ι is Type B. Then $|\alpha_{i_\iota}| = 3$, so there exist $a, b, c \in X$ such that $\alpha_{i_\iota} = abc$. Thus we can write E as

$$x\mu_1 abc\mu_2\delta y \doteq y\mu_1' cba\mu_2'\delta^R x$$

where $\mu_1 = \alpha_1\alpha_2 \ldots \alpha_{i_\iota-1}$, $\mu_1' = \alpha_1^R\alpha_2^R \ldots \alpha_{i_\iota-1}^R$, $\mu_2 = \alpha_{i_\iota+1}\alpha_{i_\iota+2} \ldots \alpha_{m-1}$, $\mu_2' = \alpha_{i_\iota+1}^R\alpha_{i_\iota+2}^R \ldots \alpha_{m-1}^R$ and $\delta = \alpha_m$. Moreover, $\Gamma_\iota^E = var(\mu_2\delta) \cup \{a, c\}$. Let z be the lexicographically minimal element of Γ_ι^E. Then by our assumption that B_ι is not lex-minimal, $z \neq a$. Moreover, since E is jumbled, we may conclude that $|\delta| \neq 1$ (otherwise we would have $(\delta, \delta) \in \Upsilon_E$, a contradiction).

By Lemma 7.16, we have two cases. The first is that there exists $n \in O(|E|^2)$, $a', c' \in X$ and $\beta \in X^+$ such that $E \odot E'$ where E' is given by

$$x\alpha_1\alpha_2 \ldots \alpha_{i_\iota-1} a'bc'\beta y \doteq y\alpha_1^R\alpha_2^R \ldots \alpha_{i_\iota-1}^R c'ba'\beta^R x.$$

Let $I_{E'} = \{i'_1, i'_2, \ldots, i'_\ell\} = \{i \mid 1 \leq i < i_\iota \text{ and } |\alpha_i| \neq 2\} \cup \{i_\iota\}$. Let $\mathfrak{B}' = (B'_0, B'_1, \ldots, B'_\ell)$ be the block decomposition of E'. Then since $I_E \cap \{1, 2, \ldots, i_\iota\} = I_{E'} \cap \{1, 2, \ldots, i_\iota\}$, we have $B_j = B'_j$ for $0 \leq j \leq \iota - 1$. Moreover, since $I_{E'}$ does not contain any elements greater than i_ι, B'_ι is the final block, so the first statement holds for $C_\iota = B_\iota$.

The second case is that there exist $n' \in O(|E|^2)$, $w \in X$ and η, η' such that $E \odot^{n'} x\mu_1 zbw\eta y \doteq y\mu'_1 wbz\eta' x$. By Lemma 7.6, there exist $n'' \in O(|E|^2)$ and $\alpha'_{i_\iota+1}, \alpha'_{i_\iota+2}, \ldots, \alpha'_{m'} \in X^+$ with $|\alpha'_j| \leq 3$ for $i_\iota < j < m'$ such that $E \odot E'$ where E' is given by

$$x\alpha_1\alpha_2 \ldots \alpha_{i_\iota-1} zbw\alpha'_{i_\iota+1}\alpha'_{i_\iota+2} \ldots \alpha'_{m'} y \doteq y\alpha_1^R\alpha_2^R \ldots \alpha_{i_\iota-1}^R wbz\alpha_{i_\iota+1}'^R\alpha_{i_\iota+2}'^R \ldots \alpha_{m'}'^R x.$$

Let $I_{E'} = \{i'_1, i'_2, \ldots, i'_\ell\} = \{i \mid 1 \leq i < i_\iota \text{ and } |\alpha_i| \neq 2\} \cup \{i_\iota\} \cup \{i \mid i_\iota + 1 \leq i < m' \text{ and } |\alpha'_i| \neq 2\}$ with $1 \leq i'_1 < i'_2 < \ldots < i'_\ell < m'$. Let $\mathfrak{B}' = (B'_0, B'_1, \ldots, B'_\ell)$ be the block decomposition of E'. Then since $I_E \cap \{1, 2, \ldots, i_\iota\} = I_{E'} \cap \{1, 2, \ldots, i_\iota\}$, we have $B_j = B'_j$ for $0 \leq j \leq \iota - 1$. Moreover, since z is minimal in $\Gamma_\iota^E = \Gamma_\iota^{E'}$, B'_ι is lex-minimal and the second statement of the lemma statement holds for $C_j = B'_j$ for $\iota \leq j \leq \ell$. □

Finally, for the sake of completeness, we provide a formal summary of the proof of Theorem 7.11 based on Lemma 7.17 using the arguments which have so-far been described informally.

Theorem 7.11 Let E be a jumbled basic RWE. By Theorem 7.5, we may assume that E is in normal form. Let $\mathfrak{B} = (B_0, B_1, \ldots, B_k)$ be its block decomposition. If B_i is lex-minimal for $0 < i < k$, then E is in LNF and we are done (this also covers the case that $k \leq 1$). Otherwise, suppose that $k > 1$ and let $\iota = \min_{0<j<k} \{j \mid B_j \text{ is not lex-minimal}\}$. Then by Lemma 7.17, we have two possibilities. Either:

1. there exists a block C_ι, $n \in O(|E|)$ and $\hat{E}$ such that $E \odot^n \hat{E}$ and $\hat{E}$ has the block decomposition $(B_0, B_1, \ldots, B_{\iota-1}, C_\iota)$, or
2. there exist blocks $C_\iota, C_{\iota+1}, \ldots, C_\ell$, $n \in O(|E|^2)$ and $\hat{E}$ such that $E \odot^n \hat{E}$ and such that $\hat{E}$ has the block decomposition $(B_0, B_1, \ldots, B_{\iota-1}, C_\iota, C_{\iota+1}, \ldots, C_\ell)$ and such that C_ι is lex-minimal.

In the first case, by definition of ι, B_j is lex-minimal for $0 < j < \iota$, meaning $\hat{E}$ is in LNF and we are done. In the second case, we have an equation $\hat{E}$ such that $E \odot^{n'} \hat{E}$ where $n' \in O(|E|)^2$ and such that the block decomposition of $\hat{E}$ has a longer initial sequence of lex-minimal blocks than the block decomposition of E.

Furthermore, it follows from the definitions that any block decomposition cannot have more blocks than the number of variables occurring in the equation. Recall that the set of variables occurring in an equation is invariant under $\Rightarrow^*$ (and therefore also $\odot$). Thus with at most $O(|E|)$ applications of Lemma 7.17, we may conclude that $E \odot^{n''} E'$ for an equation E' and with block decomposition $(B'_0, B'_1, B'_2, \ldots, B'_{k'})$ such that B'_j is lex-minimal for $0 < j < k'$ (meaning E' is in LNF) and such that $n'' \in O(|E|^3)$. It follows directly from the definitions that $\odot$ is symmetric, and

therefore we also have $E' \odot^{n''} E$. By Corollary 7.3, we may therefore conclude that $E' \Rightarrow^{n_1} E$ and $E \Rightarrow^{n_2} E'$ for some $n_1, n_2 \in O(|E|^4)$. □

8 Diameter

It was mentioned in the previous section that the choices for the blocks in a block decomposition of an equation in normal form are restricted by the invariant Υ_E. We shall now make full use of that fact to show that the number of equations in Lex Normal Form in a single graph $\mathscr{G}^{\Rightarrow}_{[E]}$ is bounded by a polynomial in $|E|$ (Theorem 7.11), and as a consequence that the diameter of $\mathscr{G}^{\Rightarrow}_{[E]}$ is also bounded by a polynomial in $|E|$ (Theorem 8.11). By combining this result with Theorems 6.8 and 4.8, we can extend it from jumbled basic regular word equations to all regular word equations. Consequently, we can conclude that satisfiability of regular word equations is NP-complete (Theorem 8.12).

Since each equation in Lex Normal Form has a unique block decomposition, it is sufficient to count the possible block decompositions satisfying the conditions for Lex Normal Form for a given value of Υ_E. We shall focus on conditions which force two blocks to be the same. We shall consider the cases of initial, standard and final blocks separately, but first we need the following lemmas which take advantage of the invariant Υ_E in order to limit the equations in normal form occurring in a single equivalence class $[E]_{\Rightarrow}$.

The first of these lemmas, and the resulting corollary provide some intuition behind the definition of the block decomposition and to why the blocks are often fixed by the invariant Υ_E (along with the leftmost variable which, aside from exceptional cases, is fixed by Lex Normal Form). Essentially, they show that the length-two factors α_i (and thus α_i^R) occurring as per the definition of normal form are fixed exactly by the variables preceding them along with the invariant Υ_E.

Lemma 8.1 *Let $u, v, a, b \in X$ and let $\alpha_1, \alpha_2, \beta_1, \beta_2, \alpha_1', \alpha_2', \beta_1', \beta_2', \gamma \in X^*$ such that $1 \leq |\gamma| \leq 3$. Let E_1 and E_2 be jumbled basic RWEs given by*

$$\begin{aligned} E_1 : &\quad \alpha_1 uab\alpha_2 \doteq \beta_1 vba\beta_2 \\ E_2 : &\quad \alpha_1' u\gamma\alpha_2' \doteq \beta_1' v\gamma^R\beta_2'. \end{aligned}$$

If $\Upsilon_{E_1} = \Upsilon_{E_2}$ then $\gamma = ab$.

Proof Let $\gamma = c_1c_2 \ldots c_n$ with $c_i \in X$, $1 \leq i \leq n$. Suppose that $\Upsilon_{E_1} = \Upsilon_{E_2} = \Upsilon$. Note that $(a, v), (u, b) \in \Upsilon$. If $|\gamma| = 1$, then $(u, v) \in \Upsilon$, which by Remark 5.2, implies $a = u$, a contradiction to the assumption that E_1 is regular. Similarly, if $|\gamma| = 3$, then $(c_2, v), (u, c_2) \in \Upsilon$ which by Remark 5.2 implies $c_2 = a = b$, again a contradiction to the assumption that E_1 is regular. Thus, it follows that $|\gamma| = 2$. In this case, we have that $(c_1, v), (u, c_2) \in \Upsilon$. By Remark 5.2, it follows that $c_1 = a$ and $c_2 = b$ so $\gamma = ab$ as required. □

Corollary 8.2 *Let $k \in \mathbb{N}$. For $1 \leq i \leq 4$ and $1 \leq j \leq k$, let $\mu_i, \mu_i', \alpha_j, \beta_j \in X^*$ such that $|\alpha_j| = |\beta_j| = 2$. Let E_1 and E_2 be the jumbled basic RWEs given by*

$$E_1 : \quad \mu_1 u \alpha_1 \alpha_2 \ldots \alpha_k \mu_2 \doteq \mu_3 v \alpha_1^R \alpha_2^R \ldots \alpha_k^R \mu_4$$
$$E_2 : \quad \mu_1' u \beta_1 \beta_2 \ldots \beta_k \mu_2' \doteq \mu_3' v \beta_1^R \beta_2^R \ldots \beta_k^R \mu_4'.$$

Suppose that $\Upsilon_{E_1} = \Upsilon_{E_2}$. Then $\alpha_j = \beta_j$ for $1 \leq j \leq k$.

Any initial block has the form $(x\alpha_1\alpha_2 \ldots \alpha_i, y\alpha_1^R\alpha_2^R \ldots \alpha_i^R)$ where $x, y \in X$ and $\alpha_j \in X^*$ with $|\alpha_j| = 2$ for $1 \leq j \leq i$. Since x, y are fixed by Υ_E, it follows from Corollary 8.2 that all the α_j factors, for $1 \leq j \leq i$ are fixed exactly by the invariant Υ_E. With a little additional effort, we can conclude the slightly more general statement that initial blocks occurring in the block decomposition of some equation E in normal form are fixed exactly by Υ_E. Recall from the definitions that in a block decomposition $(B_0, B_1, \ldots, B_k)$ of an equation in normal form, B_0 will be an initial block provided $k \geq 1$ (if $k = 0$ then $B_0 = B_k$ will be a final block).

Lemma 8.3 *Let E_1, E_2 be jumbled basic RWEs in normal form such that $\Upsilon_{E_1} = \Upsilon_{E_2}$. Let $(B_0, B_1, \ldots, B_k)$ and $(C_0, C_1, \ldots, C_\ell)$ be the block decompositions of E_1 and E_2 respectively. Suppose that $k, \ell \geq 1$. Then $B_0 = C_0$.*

Proof Since E_1 is in normal form, we may write it as $x\alpha_1\alpha_2 \ldots \alpha_n y \doteq y\alpha_1^R\alpha_2^R \ldots \alpha_n^R x$ with $x, y \in X$ and $\alpha_i \in X^+$ for $1 \leq i \leq n$ such that $|\alpha_i| \leq 3$ for $1 \leq i < n$. Similarly, we may write E_2 as $x'\alpha_1'\alpha_2' \ldots \alpha_m' y' \doteq y'\alpha_1'^R\alpha_2'^R \ldots \alpha_m'^R x'$ with $x', y' \in X$ and $\alpha_i' \in X^+$ for $1 \leq i \leq m$ such that $|\alpha_i'| \leq 3$ for $1 \leq i < m$. Suppose that $\Upsilon_{E_1} = \Upsilon_{E_2} = \Upsilon$ and note that this implies $var(E_1) = var(E_2)$. Similarly, is easily verified (either from the definition of $\Rightarrow$, or from Remark 5.2) that $x = x'$ and $y = y'$.

Since $k, \ell \geq 1$, there must exist $p = \min\{i \mid 1 \leq i < n \text{ and } |\alpha_i| \neq 2\}$ and $q = \min\{i \mid 1 \leq i < m \text{ and } |\alpha_i'| \neq 2\}$. It follows that $B_0 = (x\alpha_1\alpha_2 \ldots \alpha_{p-1}, y\alpha_1^R\alpha_2^R \ldots \alpha_{p-1}^R)$ and $C_0 = (x\alpha_1'\alpha_2' \ldots \alpha_{q-1}', y\alpha_1'^R\alpha_2'^R \ldots \alpha_{q-1}'^R)$. By Corollary 8.2, it follows that $\alpha_i = \alpha_i'$ for $1 \leq i < \min\{p, q\}$.

Suppose for contradiction that $p \neq q$. W.l.o.g. suppose that $p > q$. Then we may write E_1 and E_2 as $\mu_1 uab\mu_2 \doteq \mu_3 vba\mu_4$ and $\mu_1' uab\gamma\mu_2' \doteq \mu_3' v\gamma^R\mu_4'$ respectively where $\mu_1, \mu_2, \mu_3, \mu_4, \mu_1', \mu_2', \mu_3', \mu_4', \gamma \in X^*$, $u, v, a, b \in X$, and $|\gamma| \in \{1, 3\}$ (in particular, this is true for $ab = \alpha_q$ and $\gamma = \alpha_q'$). However in this case, it follows from Lemma 8.1 that $\Upsilon_{E_1} \neq \Upsilon_{E_2}$, a contradiction. Thus we must have that $p = q$, and the fact that $B_0 = C_0$ follows immediately. □

Similarly to initial blocks, we can use Corollary 8.2 to restrict standard blocks which are Type A. These blocks will have the form $(z\alpha_1\alpha_2 \ldots \alpha_i, z\alpha_1^R\alpha_2^R \ldots \alpha_i^R)$ where $z \in X$ and $\alpha_j \in X^*$ with $|\alpha_j| = 2$ for $1 \leq j \leq i$. Hence the factors α_j, $1 \leq j \leq i$ are fixed completely by Υ_E and z. For Type B blocks, which instead have the form $(abc\alpha_1\alpha_2 \ldots \alpha_i, cba\alpha_1^R\alpha_2^R \ldots \alpha_i^R)$ with $a, b, c \in X$, we need the following additional observation.

Lemma 8.4 *Let $u, v, a, b, c, \in X$ and let $\alpha_1, \alpha_2, \beta_1, \beta_2, \alpha_1', \alpha_2', \beta_1', \beta_2', \gamma \in X^*$ such that $1 \leq |\gamma| \leq 3$. Let E_1 and E_2 be the basic regular word equations given by*

$$E_1: \quad \alpha_1 uabc\alpha_2 \doteq \beta_1 vcba\beta_2$$
$$E_2: \quad \alpha_1' u\gamma\alpha_2' \doteq \beta_1' v\gamma^R\beta_2'.$$

If $\Upsilon_{E_1} = \Upsilon_{E_2}$ then there exist $a', c' \in X$ such that $\gamma = a'bc'$. Moreover, if $a' = a$, then $c' = c$.

Proof Let $\gamma = e_1e_2 \dots e_n$ with $e_i \in X$, $1 \leq i \leq n$. Suppose that $\Upsilon_{E_1} = \Upsilon_{E_2} = \Upsilon$. Note that $(u, b), (a, c), (b, v) \in \Upsilon$. If $|\gamma| = 1$, then $(u, v) \in \Upsilon$, and by Remark 5.2 we have that $u = b$, a contradiction to the assumption that E is regular. Thus we assume $n \geq 2$. Then $(u, e_2), (e_{n-1}, v) \in \Upsilon$. Hence, we have $e_2 = e_{n-1} = b$, and since E is regular, this implies that $n = 3$ so the statement holds with $a' = e_1, b' = e_3$. Finally, we note that since $(a', c') \in \Upsilon$, by Remark 5.2, if $a = a'$ then $c = c'$ as claimed. □

In what follows we shall show that for two jumbled basic regular equations E_1, E_2 in Lex Normal Form with $\Upsilon_{E_1} = \Upsilon_{E_2}$ and block decompositions of the same length, all blocks except the final blocks must be identical (Corollary 8.7). We have already shown in Lemma 8.3 that this is true for the initial blocks, The next step is to show that if the previous blocks in both block decompositions are identical, then the next blocks will have the same type.

Lemma 8.5 *Let E_1, E_2 be jumbled basic regular word equations in normal form such that $\Upsilon_{E_1} = \Upsilon_{E_2}$. Let $(B_0, B_1, \dots, B_k)$ and $(C_0, C_1, \dots, C_\ell)$ be block decompositions of E_1 and E_2 respectively. Suppose that $i, j \in \mathbb{N}_0$ with $i < k$, $j < \ell$ such that $B_i = C_j$. Then B_{i+1} and C_{j+1} have the same type.*

Proof Since there are two types, it is sufficient to prove that B_{i+1} is Type B if and only if C_{j+1} is Type B. Suppose that B_{i+1} is Type B and suppose for contradiction that C_{j+1} is Type A. Then there exist $\gamma_1, \gamma_2, \gamma_3, \gamma_4 \in X^*$, and $a, b, c, d \in X$ such that $B_{i+1} = (abc\gamma_1, cba\gamma_2)$ and $C_{j+1} = (d\gamma_3, d\gamma_4)$. Note that there exist $u, v \in X$ such that $B_i = C_j = (\delta_1 u, \delta_2 v)$ where $\delta_1, \delta_2 \in X^*$. Hence there exist $\alpha_1, \alpha_2, \beta_1, \beta_2, \alpha_1', \alpha_2', \beta_1', \beta_2' \in X^*$ such that E_1 is may be written as $\alpha_1 uabc\alpha_2 \doteq \beta_1 vcba\beta_2$ and E_2 may be written as $\alpha_1' ud\alpha_2' \doteq \beta_1' vd\beta_2'$. However, by Lemma 8.4, this implies $\Upsilon_{E_1} \neq \Upsilon_{E_2}$, a contradiction. Consequently, C_{j+1} is Type B if B_{i+1} is Type B. The proof that B_{i+1} is Type B if C_{j+1} is Type B is symmetric and can be obtained by simply swapping E_1 and E_2. □

We are now ready to show that standard blocks in a block decomposition are fixed entirely by the preceding block, the invariant Υ_E, and the leftmost letter of the block. This is the primary motivation for the definition of Lex Normal Form, which restricts the choice for the leftmost letter of the block where possible, and thus restricts the possibilities for the standard blocks. In particular, it follows directly by a straightforward induction that for two jumbled basic RWEs in Lex Normal Form with the same invariant Υ_E, if their block decompositions have the same length, then all but the final blocks will be identical.

Lemma 8.6 *Let E_1, E_2 be jumbled basic RWEs in normal form such that $\Upsilon_{E_1} = \Upsilon_{E_2}$. Let $(B_0, B_1, \ldots, B_k)$ and $(C_0, C_1, \ldots, C_\ell)$ be their respective block decompositions and let $k, \ell > 0$. Suppose that $B_i = C_j$, for some $i < k-1$, $j < \ell - 1$. Let $B_{i+1} = (\gamma_1, \gamma_2)$ and $C_{j+1} = (\delta_1, \delta_2)$ with $\gamma_1, \gamma_2, \delta_1, \delta_2 \in X^*$. If $\gamma_1[1] = \delta_1[1]$, then $B_{i+1} = C_{j+1}$.*

Proof Note that since $0 < i+1 < k$ and $0 < j+1 < \ell$, the blocks B_{i+1} and C_{j+1} are both standard blocks. Note also that by Lemma 8.5, B_{i+1} and C_{j+1} have the same type. Hence, by definition, there exist $\alpha_1, \alpha_2, \ldots, \alpha_n$, $\beta_1, \beta_2, \ldots, \beta_m \in X^+$ such that $B_{i+1} = (\alpha_1\alpha_2\ldots\alpha_n, \alpha_1^R\alpha_2^R\ldots\alpha_n^R)$ and $C_{j+1} = (\beta_1\beta_2\ldots\beta_m, \beta_1^R\beta_2^R\ldots, \beta_m^R)$, where $|\alpha_1| = |\beta_1| \in \{1,3\}$ and $|\alpha_p|, |\beta_q| = 2$ for $2 \leq p \leq n$ and $2 \leq q \leq m$. Since $B_i = C_j$, there exist $u, v \in X$ and $\mu_1, \mu_2, \nu_1, \nu_2, \mu_1', \mu_2', \nu_1', \nu_2', \eta, \eta' \in X^*$ with $|\eta|, |\eta'| \in \{1,3\}$ and such that E_1 is given by $\mu_1 u\alpha_1\alpha_2\ldots\alpha_n\eta\mu_2 \doteq \nu_1 v\alpha_1^R\alpha_2^R\ldots\alpha_n^R\eta^R\nu_2$ and E_2 is given by $\mu_1' u\beta_1\beta_2\ldots\beta_n\eta'\mu_2' \doteq \nu_1' v\beta_1^R\beta_2^R\ldots\beta_n^R\eta'^R\nu_2'$.

By the assumption that $\gamma[1] = \delta[1]$, we have that $\alpha_1[1] = \beta_1[1]$ meaning if $|\alpha_1| = |\beta_1| = 1$ then $\alpha_1 = \beta_1$ holds trivially. Similarly, if $|\alpha_1| = |\beta_1| = 3$, then it follows from Lemma 8.4 that $\alpha_1 = \beta_1$. In both cases, it follows from Corollary 8.2 that additionally, $\alpha_p = \beta_p$ for $2 \leq p \leq \min\{n, m\}$. It follows from Lemma 8.1 that $n = m$. Hence we have $B_{i+1} = C_{j+1}$ as required. □

Note that if the first i blocks are identical in the block decompositions of two jumbled basic RWEs in Lex Normal Form with the same invariant set Υ_E, it follows that the set Γ_{i+1}^E is also the same in both cases. Consequently, by definition of Lex Normal Form, if the $i+1^{th}$ blocks are not final blocks, the leftmost variable will be the same in each case (namely the lexicographically minimal element of Γ_{i+1}^E). Consequently, by Lemma 8.6, the $i+1^{th}$ blocks will also be identical. By a simple induction, we can thus conclude the following.

Corollary 8.7 *Let E_1, E_2 be jumbled basic RWEs in Lex Normal Form such that $\Upsilon_{E_1} = \Upsilon_{E_2}$. Let $(B_0, B_1, \ldots, B_k)$ and $(C_0, C_1, \ldots, C_\ell)$ be their respective block decompositions and suppose that $k, \ell > 0$. Then $B_i = C_i$ for $0 \leq i < \min(k, \ell)$.*

Consequently, two equations in Lex Normal Form in the graph $\mathscr{G}_{[E]}^{\rightrightarrows}$ with block decompositions containing the same number of blocks may differ only in the final block. Clearly, the number of blocks in a block decomposition is at most $\mathrm{Card}(var(E))$. Thus, in order to bound the number of equations in Lex Normal Form in $\mathscr{G}_{[E]}^{\rightrightarrows}$, it suffices to count the possibilities for the final block.

Recall from the definition of normal form that the last (rightmost) α_i factor is the only one which may have length greater than 3. Consequently, we need a counterpart to Lemmas 8.1 and 8.4 for this case, given by the following.

Lemma 8.8 *Let $u, v, x, y, x', y' \in X$ and let $\alpha, \beta, \alpha', \beta', \gamma, \gamma' \in X^*$ such that $|\gamma| \geq 1$. Let E_1 and E_2 be the basic regular word equations given by $x\alpha u\gamma y \doteq y\beta v\gamma^R x$ and $x'\alpha' u\gamma' y' \doteq y'\beta' v\gamma'^R x'$ respectively. If $\Upsilon_{E_1} = \Upsilon_{E_2}$ and $\gamma[1] = \gamma'[1]$ then $\gamma = \gamma'$.*

Proof Let $z_1, z_2, \ldots, z_n, w_1, w_2, \ldots, w_m \in X$ be variables such that $\gamma = z_1 z_2 \ldots z_n$ and $\gamma' = w_1 w_2 \ldots w_m$ and suppose that $z_1 = w_1$. Suppose also that $\Upsilon_{E_1} = \Upsilon_{E_2} = \Upsilon$. Note that for $1 \leq i \leq \min\{n, m\} - 2$, we have $(z_i, z_{i+2}), (w_i, w_{i+2}) \in \Upsilon$. Moreover, if $n, m \geq 2$, we also have that $(u, z_2), (u, w_2) \in \Upsilon$. Consequently, by Remark 5.2, we have that $w_i = z_i$ for $1 \leq i \leq \min\{n, m\}$. If $n = m$ we are done. Otherwise, suppose that $n \neq m$, and note in particular that since E_1, E_2 are regular, this implies $z_n \neq w_m$. However, $(z_n, z_1), (w_n, w_1) \in \Upsilon$, and since $w_1 = z_1$, by Remark 5.2 we have that $z_n = w_m$, a contradiction. Thus we must have $n = m$ and $\gamma = \gamma'$ as claimed. □

The following lemma establishes conditions under which two final blocks must be identical, forming the basis for our bound on the number of possible final blocks in a block decomposition of an equation in Lex Normal Form, and consequently, a bound on the number of equations in Lex Normal Form itself.

Lemma 8.9 *Let E_1, E_2 be jumbled basic RWEs in normal form such that $\Upsilon_{E_1} = \Upsilon_{E_2}$. Let $(B_0, B_1, \ldots, B_k)$ and $(C_0, C_1, \ldots, C_\ell)$ be their respective block decompositions. Suppose that $k, \ell > 0$ and that $B_{k-1} = C_{\ell-1}$. Let $B_k = (\alpha_1 \alpha_2 \ldots \alpha_n y, \alpha_1^R \alpha_2^R \ldots \alpha_n^R x)$ and $C_\ell = (\beta_1 \beta_2 \ldots \beta_m y, \beta_1^R \beta_2^R \ldots, \beta_m^R x)$, where $x, y \in X$, $\alpha_1, \alpha_2, \ldots, \alpha_n$, $\beta_1, \beta_2, \ldots, \beta_m \in X^+$, $|\alpha_1| = |\beta_1| \in \{1, 3\}$ and $|\alpha_i|, |\beta_j| = 2$ for $2 \leq i < n$ and $2 \leq j < m$. Then if $\alpha_1[1] = \beta_1[1]$, $n = m$, and $\alpha_n[1] = \beta_m[1]$, we have $B_k = C_\ell$.*

Proof Suppose that all the conditions of the lemma are met. Note that B_k and C_ℓ are both end blocks. Note also that by Lemma 8.5, B_k and C_ℓ have the same type.

Since $B_{k-1} = C_{\ell-1}$, there exist $u, v \in X$ and $\mu_1, \mu_2, \mu_1', \mu_2' \in X^*$ such that E_1 and E_2 are given by:

$$\begin{aligned} E_1 : &\quad x\mu_1 u \alpha_1 \alpha_2 \ldots \alpha_n y \doteq y \mu_2 v \alpha_1^R \alpha_2^R \ldots \alpha_n^R x \\ E_2 : &\quad x\mu_1' u \beta_1 \beta_2 \ldots \beta_n y \doteq y \mu_2' v \beta_1^R \beta_2^R \ldots \beta_n^R x. \end{aligned}$$

By the assumption that $\alpha_1[1] = \beta_1[1]$, we have that if $|\alpha_1| = |\beta_1| = 1$ then trivially $\alpha_1 = \beta_1$, and if $|\alpha_1| = |\beta_1| = 3$, then $\alpha_1 = \beta_1$ by Lemma 8.4. In both cases, it follows from Corollary 8.2 that $\alpha_i = \beta_i$ for $1 \leq i < \min\{n, m\}$. It follows from Lemma 8.1 that $n = m$, and from Lemma 8.8 that $\alpha_n = \beta_m$. Consequently, we have $B_k = C_\ell$ as claimed. □

Lemma 8.9 reveals that the options for last block are dependent only on the choices of three parameters: $\alpha_1[1]$, $\alpha_n[1]$, and n. Since each of these can take at most $|E|$ possible values, there are $|E|^3$ possibilities altogether. Thus for each possible number of blocks, there are at most $|E|^3$ possible block decompositions, and therefore only $|E|^4$ possible block decompositions respecting the invariant Υ_E in total. Since every equation in Lex Normal Form permits a unique block decomposition, this gives us our desired polynomial bound.

Theorem 8.10 *Let E be a jumbled basic RWE. Let S be the set of basic regular equations E' in Lex Normal Form for which $\Upsilon_E = \Upsilon_{E'}$. Then $Card(S) \leq |E|^4$.*

Proof We shall count possible block decompositions of equations E' for which $\Upsilon_{E'} = \Upsilon_E = \Upsilon$. Since the block decomposition uniquely determines the equation, this count is an upper bound on the number of equations in S. Note that $\Upsilon_{E'} = \Upsilon_E$, implies $var(E') = var(E)$.

It is straightforward from the definitions that any block decomposition of an equation E' can have at most $\text{Card}(var(E')) = \text{Card}(var(E)) < |E|$ blocks, so it is sufficient to count how many block decompositions with exactly N blocks are possible for each $N \leq \text{Card}(var(E))$.

We start with the case that the block decomposition consists of exactly one block ($N = 1$). Suppose we have two basic regular word equations E_1, E_2 in Lex Normal Form, such that $\Upsilon_{E_1} = \Upsilon_{E_2} = \Upsilon$ (and so additionally $var(E_1) = var(E_2) = var(E)$). Suppose that (B_0) and (C_0) are the block decompositions of E_1 and E_2 respectively. By definition $B_0 = E_1$ and $C_0 = E_2$. It follows that $B_0 = (x\alpha_1\alpha_2 \ldots \alpha_o y, y\alpha_1^R\alpha_2^R \ldots \alpha_o^R x)$ and $C_0 = (x'\alpha_1'\alpha_2' \ldots \alpha_m' y', y'\alpha_1', \alpha_2', \ldots, \alpha_m' x')$ where $x, x', y, y' \in X$ and $\alpha_i, \alpha_j' \in X^+$ for $1 \leq i \leq o$, $1 \leq j \leq m$ and such that $|\alpha_i|, |\alpha_j| = 2$ for $1 \leq i < o$ and $1 \leq j < m$. It is easily verified (either from the definition of $\Rightarrow$, or from Remark 5.2) that $x = x'$ and $y = y'$. Moreover, we clearly must have $o, m < \text{Card}(var(E))$. Now suppose that $o = m$. Then by Corollary 8.2, we may conclude that $\alpha_i = \alpha_i'$ for $1 \leq i < n$. Similarly, it follows from Lemma 8.9 that $\alpha_n = \alpha_n'$, and thus $B_0 = C_0$. Hence, for each possible value of o, there is at most one possible block decomposition, meaning there are fewer than $\text{Card}(var(E)) < |E|$ possible block decompositions containing only one block.

Now consider the cases that there is more than one block in the block decomposition ($1 < N \leq \text{Card}(var(E))$). Suppose we have two basic regular word equations E_1, E_2 in Lex Normal Form, such that $\Upsilon_{E_1} = \Upsilon_{E_2} = \Upsilon$. Suppose that $(B_0, B_1, B_2, \ldots, B_n)$ and $(C_0, C_1, \ldots, C_n)$ are the block decompositions of E_1 and E_2 respectively, and that they have the same number of blocks $1 < n \leq \text{Card}(var(E))$. By Corollary 8.7, we have that $B_i = C_i$ for $0 \leq i \leq n-1$. By Lemma 8.9, there are at most $|E|^3$ possibilities for the end block C_n. Thus there are at most $|E|^3$ block decompositions overall with exactly n blocks for $1 < n \leq |E|$. Thus at most $|E|^4$ possible block decompositions in total, and the statement of the theorem follows. □

For a jumbled basic RWE E, since every vertex in $\mathscr{G}^{\Rightarrow}_{[E]}$ is a small (i.e. bounded by a polynomial in $|E|$) distance from a vertex in Lex Normal Form, and since there are only a small number of such vertices, it is straightforward to show that the diameter of $\mathscr{G}^{\Rightarrow}_{[E]}$ must also be small: indeed if we have a sufficiently long path between two vertices, then we must have a long path between two vertices which are close to the same vertex in Lex Normal Form. Since they are close to the same vertex, we can find a shortcut between them, and the initial long path is not minimal. Knowing that the diameter of $\mathscr{G}^{\Rightarrow}_{[E]}$ is bounded by a polynomial in $|E|$ when E is jumbled and basic, it follows from Theorems 6.8 and 4.8 (see also Remark 4.6) and Proposition 3.5 that the diameter of $\mathscr{G}^{\Rightarrow NT}_{[E]}$ is bounded by a polynomial in $|E|$ whenever E is regular.

Theorem 8.11 *Let E be a basic RWE. Then $diam(\mathscr{G}^{\Rightarrow}_{[E]}) \in O(|E|^{10})$. Consequently, for any RWE E, $diam(\mathscr{G}^{\Rightarrow_{NT}}_{[E]}) \in O(|E|^{12})$.*

Proof We shall first consider the case of $diam(\mathscr{G}^{\Rightarrow}_{[E]})$ when E is jumbled, basic and regular. Let $S = \{E' \in [E]_{\Rightarrow} \mid E' \text{ is in Lex Normal Form}\}$. By Theorem 5.3, $\Upsilon_{E_1} = \Upsilon_{E_2}$ for all $E_1, E_2 \in [E]_{\Rightarrow}$. Thus, by Theorem 8.10, we have that $\text{Card}(S) \leq |E|^4$. Moreover, by Theorem 7.11, for every $E' \in [E]_{\Rightarrow}$, there exists some $\hat{E}' \in S$ such that E' is at most distance $O(|E|^4)$ from $\hat{E}'$, and $\hat{E}'$ is at distance at most $O(|E|^4)$ from E' in the graph $\mathscr{G}^{\Rightarrow}_{[E]}$. From this, we may conclude that $diam(\mathscr{G}^{\Rightarrow}_{[E]}) \in O(|E|^8)$ as follows: suppose for contradiction that, for an appropriate constant c, there exist $\overline{E}_1, \overline{E}_2 \in [E]_{\Rightarrow}$ such that the minimal path between them in $\mathscr{G}^{\Rightarrow}_{[E]}$ has length at least $2c|E|^8 + 1$. Let that path be $E_1, E_2, \ldots, E_n$ where $E_1 = \overline{E}_1$, $E_n = \overline{E}_2$, and $E_i \Rightarrow E_{i+1}$ for $1 \leq i \leq n$ and such that $n > 2c|E|^8 + 1$. Now, to each E_i, $1 \leq i \leq n$, we may associate some $\hat{E}_i \in S$ such that the distance from E_i to $\hat{E}_i$ is at most $c|E|^4$. Since $\text{Card}(S) \leq |E|^4$ and $n > 2c|E|^8 + 1$, we must have that there exists $\hat{E} \in S$ such that $\hat{E} = \hat{E}_i$ for at least $2c|E|^4 + 1$ different values of i. This implies in particular that there exist i_1, i_2 with $i_1 - i_2 > 2c|E|^4$ such that $\hat{E}_{i_1} = \hat{E}_{i_2}$. It follows that the length of the path $E_{i_1}, E_{i_1+1}, \ldots E_{i_2}$ is at least $2c|E|^4 + 1$, and moreover, since $E_1, E_2, \ldots, E_n$ is the shortest path between E_1 and E_2, $E_{i_1}, E_{i_1+1}, \ldots E_{i_2}$ must also be the shortest path between E_{i_1} and E_{i_2}. However, we have that E_{i_1} is distance at most $c|E|^4$ from $\hat{E}$, and that $\hat{E}$ is at most distance E_{i_2} at most $c|E|^4$ from E_{i_2}. Consequently, E_{i_1} is distance at most $2c|E|^4$ from E_{i_2}, a contradiction to the fact that $E_{i_1}, E_{i_1+1}, \ldots E_{i_2}$ is the shortest possible path. Consequently, if E is jumbled basic and regular, then $diam(\mathscr{G}^{\Rightarrow}_{[E]}) \in O(|E|^8)$.

Now we shall consider the case that E of $diam(\mathscr{G}^{\Rightarrow}_{[E]})$ when E is basic and regular, but not necessarily jumbled. Suppose that E is given by $\alpha \doteq \beta$. Let $Y = var(E) \backslash \Delta(E)$ and let E' be the equation $\pi_Y(\alpha) \doteq \pi_Y(\beta)$. Clearly, E' is basic, regular and $|E'| \leq |E|$. By Theorem 6.8, we have that $diam(\mathscr{G}^{\Rightarrow}_{[E]}) \in O(diam(\mathscr{G}^{\Rightarrow}_{[E']})|E|^2)$. Moreover, by Lemma 6.3, E' is jumbled. Thus by our previous claim, it follows that $diam(\mathscr{G}^{\Rightarrow}_{[E]}) \in O(|E'|^8|E|^2) = O(|E|^{10})$.

Finally, we consider the case of $diam(\mathscr{G}^{\Rightarrow_{NT}}_{[E]})$ for arbitrary regular equations E. Let E be any regular word equation. Then by Proposition 3.5, $diam(\mathscr{G}^{\Rightarrow_{NT}}_{[E]}) \leq 1 + (|E| + 1)m$ where

$$m = \max\{diam(\mathscr{G}^{\Rightarrow}_{[E']}) \mid E \Rightarrow^*_{NT} E'\}.$$

Now fix E' be such that $E \Rightarrow^*_{NT} E'$ and $diam(\mathscr{G}^{\Rightarrow}_{[E']}) = m$. Then since $E \Rightarrow^*_{NT} E'$, E' is also regular and $|E'| \leq |E|$. Moreover by Theorem 4.8, there exists a basic regular equation E'' such that $|E''| \leq |E|$ and such that $\mathscr{G}^{\Rightarrow}_{[E'']}$ is isomorphic to an isolated path compression of order $|E'|$ of $\mathscr{G}^{\Rightarrow}_{[E']}$. Thus (cf. Remark 4.6), we have $m \leq |E'|diam(\mathscr{G}^{\Rightarrow}_{[E'']})$. Since E'' is basic and regular, we have that $diam(\mathscr{G}^{\Rightarrow}_{[E'']}) \in O(|E''|^{10})$. Since $|E''|, |E'| \leq |E|$, we therefore have $m \in O(|E|^{11})$ and $diam(\mathscr{G}^{\Rightarrow_{NT}}_{[E]}) \in O(|E|^{12})$. □

Due to Proposition 3.4, we may infer directly from Theorem 8.11 that the satisfiability problem for regular word equations is in NP. It was already shown in [8] that this problem is NP-hard, and thus we obtain matching upper and lower bounds for its complexity.

Theorem 8.12 *The satisfiability problem for RWEs is NP-complete.*

Proof Directly from Theorem 8.11 and Proposition 3.4. □

9 Size

While the diameter of $\mathscr{G}^{\Rightarrow}_{[E]}$ is one important parameter, being directly related to the complexity of the satisfiability problem, it is by no means the only interesting one. The overall size of the graphs will also play a central role in the practical performance of the algorithm described in Section 3.

For basic RWEs, we are able to give tight upper and lower bounds on the number of vertices in the graphs $\mathscr{G}^{\Rightarrow}_{[E]}$, as well as identifying the cases in which these bounds are reached. Recalling Theorem 4.8, we are also able to translate these bounds into the case of general (i.e. not basic) RWEs. In particular, when moving to a general RWE from the corresponding basic one, the effect on the graph $\mathscr{G}^{\Rightarrow}_{[E]}$ is that 'isolated paths' of length linear in $|E|$ are collapsed. In fact, an inspection of the proofs (in particular of Lemma 4.7) yields a tighter bound, namely that collapsed paths will have at most $\max(T_1, T_2)$ internal vertices where T_1 and T_2 are the number of occurrences of terminal symbols and single-occurrence variables in the LHS and RHS respectively.

Corollary 9.1 *Let E be an RWE given by $\alpha \doteq \beta$. Let E_{basic} be the corresponding basic equation as per Theorem 4.8. Let $n = Card(qv(E))$ and let $M = \max\{|\alpha| - n, |\beta| - n\}$. Then*

$$Card([E_{basic}]_{\Rightarrow}) \leq Card([E]_{\Rightarrow}) \leq MCard([E_{basic}]_{\Rightarrow}).$$

We begin with the upper bounds, which occur in the case of basic regular-rotated word equations.

Lemma 9.2 *Let E be a basic regular word equation. Let $n = Card(var(E))$ and suppose that $n \geq 2$. Let V be the number of vertices in $\mathscr{G}^{\Rightarrow}_{[E]}$. Then $V \leq \frac{n!}{2}$. Moreover, $V = \frac{n!}{2}$ if and only if there exists $E' \in [E]_{\Rightarrow}$ such that E' is regular rotated.*

Proof Let E be a basic regular word equation. Let $n = \text{Card}(var(E))$ and suppose that $n \geq 2$. Let $V = \text{Card}([E]_{\Rightarrow})$ be the number of vertices in $\mathscr{G}^{\Rightarrow}_{[E]}$. We shall begin with the claim that $V \leq \frac{n!}{2}$. To do this, we recall that from Theorem 5.3, the set $S_\Upsilon = \{E' \mid E' \text{ is a basic regular equation such that } \Upsilon_{E'} = \Upsilon_E\}$ is a (not necessarily strict) superset of $[E]_{\Rightarrow}$. We shall show that the cardinality of S_Υ is at most $\frac{n!}{2}$. Let $\Upsilon = \Upsilon_E$ and let E' be a regular basic equation such that $\Upsilon_{E'} = \Upsilon$. Now, it follows from the definition of Υ that $var(E') = var(E)$ and that the rightmost variables the

LHS (resp. RHS) of E and E' are the same. More precisely, there exist $x, y \in var(E)$ and $\alpha, \alpha', \beta, \beta' \in X^*$ such that E may be written $\alpha x \doteq \beta y$ and E' may be written as $\alpha' x \doteq \beta' y$. Clearly, there are at most $(n-1)!$ possibilities for α'. Moreover, since $\Upsilon_{E'} = \Upsilon$ is fixed, we can, given α', for each $u \in var(\beta') \backslash \{\alpha'[1], \beta'[1]\}$, determine uniquely the predecessor of u in $\beta' y$. More precisely, there exist factors vu and $v'u$ of $\alpha' x$ and $\beta' y$ respectively where $v, v' \in var(E)$. Thus $(v, v') \in \Upsilon$, so if v is fixed (i.e. by α') then v' is also fixed by Υ. It follows directly that for each choice of α', there exists a unique suffix γ of $\beta' y$ having $\alpha'[1]$ as a prefix. Moreover, once the variable occurring immediately to the left of γ (i.e. the predecessor of $\gamma[1]$ in $\beta' y$) is fixed, then $\beta' y$ is fixed entirely, meaning that there are $n - |\gamma|$ possible choices for $\beta' y$ once α' is fixed.

Next, we shall show that for each k, $1 \leq k \leq n-1$, there are exactly $(n-2)!$ choices of α' such that the corresponding γ has length exactly k. For other values of k, there are no possible choices of α' due to the fact that every equation in S_Υ is basic and regular (note in particular that the case $k = n$ would result in an equation which is decomposable and therefore not basic). It follows from this that the cardinality of S_Υ is at most $\frac{n!}{2}$:

$$\mathrm{Card}(S_\Upsilon) \leq \sum_{k=1}^{n-1} k(n-2)! = (n-2)! \sum_{k=1}^{n-1} k = (n-2)! \frac{n(n-1)}{2} = \frac{n!}{2}.$$

To see why there are exactly $(n-2)!$ choices of α' such that the corresponding γ has length k, we shall take a slightly different approach to constructing/selecting α' and β'. In particular, we shall first choose γ and then see how many choices there are for α'. Let $k \in \mathbb{N}$ such that $1 \leq k < n$.

By definition of Υ_E, we must have that if $\gamma = v_1 v_2 \ldots v_{k-1} y$, then there exist $u_1, u_2, \ldots, u_{k-1} \in var(E)$ such that $\alpha'[1] = v_1$ and $(u_i, v_i) \in \Upsilon$ for $1 \leq i \leq k-2$, $(u_{k-1}, y) \in \Upsilon$, and such that $u_{k-1} y$ is a factor of $\alpha' x$ and $u_i v_{i+1}$ are factors of $\alpha' x$ for $1 \leq i \leq k-2$. Since E' is regular, it follows that $v_i \neq x$ for $1 \leq i \leq k-1$. Consequently, there are $\binom{n-2}{k-1}(k-1)! = \frac{(n-2)!}{(n-k-1)!}$ possible ways of choosing γ. Once γ is fixed, then, since $u_{k-1} y$ is a factor of $\alpha' x$ and $u_i v_{i+1}$ are factors of $\alpha' x$ for $1 \leq i \leq k-2$, we may infer that α' is uniquely determined by the relative order of the variables in $var(E) \backslash \{x, y, v_1, v_2, \ldots, v_{k-1}\}$, and thus there are $(n-k-1)!$ possible choices for α' for each choice of γ. Altogether we have $(n-k-1)! \frac{(n-2)!}{(n-k-1)!} = (n-2)!$ possible choices for α' as claimed, and it follows that $V \leq \frac{n!}{2}$.

It remains to consider the claim that $V = \frac{n!}{2}$ if and only if there exists $E' \in [E]_\Rightarrow$ such that E' is regular rotated. Note that since $n > 1$, and since E' is basic (and therefore indecomposable) for all $E' \in [E]_\Rightarrow$, E' is not regular ordered for all $E' \in [E]_\Rightarrow$.

We shall begin with the 'if' direction. Let $V = \mathrm{Card}([E]_\Rightarrow)$ be the number of vertices in $\mathscr{G}^{\Rightarrow}_{[E]}$. Then we may assume w.l.o.g. that E is regular rotated and thus we can write E as $y_1 y_2 \ldots y_k x_1 y_{k+1} y_{k+2} \ldots y_\ell x_2 \doteq y_{k+1} y_{k+2} \ldots y_\ell x_2 y_1 y_2 \ldots y_k x_1$ where $x_1, x_2, y_1, y_2, \ldots y_\ell \in X$, $\ell = n-2$ and $k \leq \ell$. Then $\Delta(E) = \{y_1, y_2, \ldots, y_\ell\}$. Consequently, by Theorem 6.8, the set of equations

$$S = \{\alpha x_1 \beta x_2 \doteq \beta x_2 \alpha x_1 \mid |\alpha\beta|_y = 1 \text{ if } y \in \Delta(E) \text{ and } |\alpha\beta|_z = 0 \text{ otherwise}\}$$

is a subset of $[E]_{\Rightarrow}$. Now, for each i, $1 \leq i \leq \ell = \mathrm{Card}(\Delta(E))$, let the set $S_i \subset S$ be the set

$$S_i = \{\alpha x_1 \beta x_2 \doteq \beta x_2 \alpha x_1 \mid |\alpha| = i \wedge |\alpha\beta|_y = 1 \text{ if } y \in \Delta(E) \text{ and } |\alpha\beta|_z = 0 \text{ otherwise}\}.$$

Clearly, we have $S = \bigcup_{0 \leq i \leq \ell} S_i$. Moreover, we have that $\mathrm{Card}(S_i) = \ell! = (n-2)!$ for each $i, 0 \leq i \leq \ell$. Finally, note that for each $i, 0 \leq i \leq \ell$, if $E' \in S_i$, then for $T_{E'} = \{E'' \mid E' \Rightarrow_R^* E''\}$, we have that $\mathrm{Card}(T_{E'}) = i + 1$ It is straightforward from the definitions that for $E_1, E_2 \in S$, if $E_1 \neq E_2$, then $T_{E_1} \cap T_{E_2} = \emptyset$. Consequently, we may conclude that

$$V \geq \sum_{E' \in S} \mathrm{Card}(T_{E'}) = \sum_{0 \leq i \leq \ell} (i+1)\mathrm{Card}(S_i) = \frac{(\ell+1)(\ell+2)}{2}(n-2)! = \frac{n!}{2}.$$

We have already shown that $V \leq \frac{n!}{2}$, so $V = \frac{n!}{2}$ as required.

Suppose now that E' is not regular rotated for all $E' \in [E]_{\Rightarrow}$. To see that $V < \frac{n!}{2}$, it suffices to notice that we can decrease the bound on $\mathrm{Card}(S_\Upsilon)$ if not all the previously considered possibilities for the left-hand-sides $\alpha' y$ are actually possible.

Recall from the Theorem 5.3 that $\Delta(E) = \Delta(E')$ for all $E' \in [E]_{\Rightarrow}$. Moreover, it follows from the definitions that the rightmost variables on each side of the equation are not contained in $(\Delta(E))$ and thus $\mathrm{Card}(\Delta(E)) \leq n-2$. Next, suppose (for contradiction) that $\mathrm{Card}(\Delta(E)) = n-2$. Then there exist $z_1, z_2, \ldots, z_n \in X$ and $i, 1 \leq i < n$ such that z_n is a suffix of the LHS of E and z_i is a suffix of the RHS of E, meaning that $\Delta(E) = \{z_j \mid 1 \leq j < n, j \neq i\}$. Consequently, there exists $j, i < j \leq n$ such that E may be written $z_1 z_2 \ldots z_n \doteq z_{j+1} \ldots z_{n-1} z_n z_{i+1} \ldots z_{j-1} z_j z_1 \ldots z_{i-2} z_{i-1} z_i$. Thus $E \Rightarrow_L^* E'$ where E' is given by $z_1 z_2 \ldots z_n \doteq z_{i+1} \ldots z_{j-1} z_j z_{j+1} \ldots z_{n-1} z_n z_1 \ldots z_{i-2} z_{i-1} z_i$. However, E' is regular-rotated, a contradiction.

Hence, we may assume that $\mathrm{Card}(\Delta(E)) < n-2$, and consequently, there exist pairwise distinct variables $u, v, x, y \in var(E)$ such that $(u, v), (x, y) \in \Upsilon_E$. However, if this is the case, then the LHS of any equation in $[E]_{\Rightarrow}$ cannot contain both the factors uv and xy. Suppose for contradiction that both factors were present in the LHS, then by definition of Υ_E, there must exist $z \in X$ such that either uz is a factor of the LHS and vz is a factor of the RHS, or xz is a factor of the LHS and yz is a factor of the RHS. W.l.o.g. we may assume the first case that uz is a factor of the LHS and vz is a factor of the RHS. However, by the assumption that uv is also a factor of the LHS, we have $z = v$, and consequently vv is a factor of the RHS, a contradiction to the fact that E is regular. It follows in this case that $\mathrm{Card}(S_\Upsilon) < \frac{n}{2}$, and thus that $V < \frac{n}{2}$. □

We can use Corollary 9.1 to adapt Lemma 9.2 to general RWEs as follows. Let E be a RWE given by $\alpha \doteq \beta$, let $n = \mathrm{Card}(qv(E))$, and let $T = \max\{|\alpha| - n, |\beta| - n\}$. Let E_{basic} be the corresponding basic RWE as per Theorem 4.8. Clearly for $\mathrm{Card}([E]_{\Rightarrow})$ to be maximal, E should be indecomposable. Now, by Corollary 9.1, we have that $\mathrm{Card}([E]_{\Rightarrow}) \leq T\mathrm{Card}([E_{basic}]_{\Rightarrow}) \leq T\frac{n!}{2} \leq \frac{(n+T)!}{2} = \frac{(\max\{|\alpha|,|\beta|\})!}{2}$.

Note also that if E is not regular-rotated, then either E_{basic} is not regular-rotated, or E is decomposable and E_{basic} is regular-rotated but with fewer variables. In either

case it follows that the second inequality becomes strict. Similarly, if $T \neq 0$, then the third inequality becomes strict. Hence we get the following.

Corollary 9.3 *Let E be a RWE given by $\alpha \doteq \beta$. Let $M = \max\{|\alpha|, |\beta|\}$. Let V be the number of vertices in $\mathscr{G}^{\rightrightarrows}_{[E]}$. Then $V \leq \frac{M!}{2}$. Moreover, $V = \frac{M!}{2}$ if and only if E is basic and there exists $E' \in [E]_{\Rightarrow}$ such that E' is regular rotated.*

For upper bounds on the number of vertices in $\mathscr{G}^{\rightrightarrows}_{[E]}$, we consider the class of regular-reversed equations. We shall eventually prove a statement similar to that of Lemma 9.2, but first we need some additional definitions and lemmas. Our reasoning in this case revolves primarily around a particular binary-tree like structure arising locally in the graphs $\mathscr{G}^{\rightrightarrows}_{[E]}$. The binary trees do not occur directly as subgraphs of $\mathscr{G}^{\rightrightarrows}_{[E]}$, but rather can be obtained by treating certain short paths as edges. The relation defining the 'edges' of the tree is given by $\triangleright$, introduced formally below. By showing that these binary trees always occur in the graphs $\mathscr{G}^{\rightrightarrows}_{[E]}$, and by verifying that they are balanced and have height proportional to the number of edges, we are able to produce the lower bound on the number of vertices in $\mathscr{G}^{\rightrightarrows}_{[E]}$ given in Lemma 9.11.

Definition 9.4 ($\rightarrow_R$, $\rightarrow_L$, $\triangleright$, $W(E)$) Let E be a basic RWE such that $\mathrm{Card}(var(E)) \geq 2$. Then we may write E in the form

$$x\gamma_0 z_1\gamma_1 z_2\gamma_2 \ldots z_k\gamma_k y\alpha \doteq y\delta_0 w_1\delta_1 w_2\delta_2 \ldots w_k\delta_k x\beta$$

with $x, y, z_1, z_2, \ldots, z_k, w_1, w_2, \ldots, w_k \in X$ such that $\{z_1, z_2, \ldots, z_k\} = \{w_1, w_2, \ldots, w_k\}$, and $\alpha, \beta, \gamma_1, \gamma_2, \ldots, \gamma_k, \delta_1, \delta_2, \ldots, \delta_k \in (X\backslash\{x, y, z_1, z_2, \ldots, z_k\})^*$ such that for each i, j, $0 \leq i, j \leq k$, we have $var(\gamma_i) \cap var(\delta_j) = \emptyset$. Note that this decomposition is unique. We define $W(E) = \{x, y, z_1, z_2, \ldots, z_k\}$. Moreover, there exist i, j such that $w_i = z_k$ and $z_j = w_k$. We define the relations $\rightarrow_L$ and $\rightarrow_R$ such that

$$\begin{aligned} &x\gamma_0 z_1\gamma_1 z_2\gamma_2 \ldots z_k\gamma_k y\alpha \doteq y\delta_0 w_1\delta_1 w_2\delta_2 \ldots w_k\delta_k x\beta \\ \rightarrow_L\ &x\gamma_0 z_1\gamma_1 z_2\gamma_2 \ldots z_k\gamma_k y\alpha \doteq w_i\delta_i w_{i+1}\delta_{i+1} \ldots w_k\delta_k y\delta_0 w_1\delta_1 w_2\delta_2 \ldots w_{i-1}\delta_{i-1} x\beta \end{aligned}$$

and

$$\begin{aligned} &x\gamma_0 z_1\gamma_1 z_2\gamma_2 \ldots z_k\gamma_k y\alpha \doteq y\delta_0 w_1\delta_1 w_2\delta_2 \ldots w_k\delta_k x\beta \\ \rightarrow_R\ &z_j\gamma_i z_{j+1}\gamma_{j+1} \ldots z_k\gamma_k x\gamma_0 z_1\gamma_1 z_2\gamma_2 \ldots z_{j-1}\gamma_{j-1} y\alpha \doteq y\delta_0 w_1\delta_1 w_2\delta_2 \ldots w_k\delta_k x\beta \end{aligned}$$

Additionally, for convenience, we define $\triangleright = \rightarrow_L \cup \rightarrow_R$.

The tree-structure we are interested in is the set $S = \{E' \mid E \triangleright^* E'\}$ for a given basic RWE E with at least two variables (the one-variable case being trivial). An example is given by Fig. 7. The following fact can be verified directly from the definition, and confirms that the set S is indeed contained in $\mathscr{G}^{\rightrightarrows}_{[E]}$.

Fact 9.5 Let E_1, E_2 be basic RWEs with $\mathrm{Card}(var(E_1)), \mathrm{Card}(var(E_2)) \geq 2$. Let $Z \in \{L, R\}$. If $E_1 \rightarrow_Z E_2$, then $E_1 \Rightarrow^*_Z E_2$. Conversely, if $E_1 \Rightarrow_Z E_2$, then either $E_1 \rightarrow^*_Z E_2$ or $E_2 \rightarrow^*_Z E_1$.

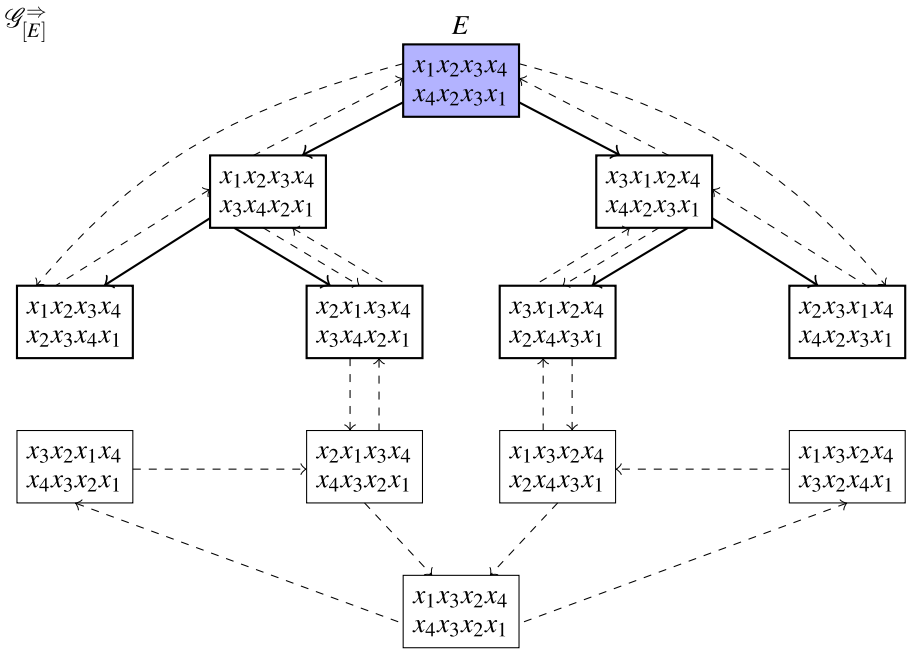

Fig. 7 The set $S = \{E' \mid E \triangleright^* E'\}$ occurring as a subset of the vertices of the graph $\mathscr{G}^{\Rightarrow}_{[E]}$ in the case that E is given by $x_1x_2x_3x_4 \doteq x_4x_2x_3x_1$. In order to conserve space, for each vertex, the equation is arranged vertically with the LHS above and the RHS below. The vertices belonging to S are highlighted in bold, and E is shaded (blue). The tree structure induced by the relation $\triangleright$ is given by the bold solid edges, while the edges of $\mathscr{G}^{\Rightarrow}_{[E]}$ are dashed. Note that the edges due to $\triangleright$ do not necessarily coincide with edges due to $\Rightarrow$, but for every $\triangleright$-edge, there is a corresponding path using $\Rightarrow$-edges, guaranteeing that $S \subseteq [E]_{\Rightarrow}$. In this case we have that $W(E) = var(E) = \{x_1, x_2, x_3, x_4\}$, so S forms a tree of height $2^{4-2} - 1 = 3$, and contains exactly $2^{4-1} - 1 = 7$ equations

In what follows, in order to understand the number of equations occurring in $S = \{E' \mid E \triangleright^* E'\}$, we shall show that when combined with the relation $\triangleright$, it becomes a balanced binary tree of height $\mathrm{Card}(W(E)) - 1$. We proceed by noting two more facts following directly from the definition. Fact 9.6 provides the first step towards understanding why $\triangleright$ induces a binary tree like structure on S: the leaf nodes are equations for which $\mathrm{Card}(W(E)) = 2$, while all other equations have exactly two children w.r.t. $\triangleright$.[6]

Fact 9.6 Let E be a basic RWE with $\mathrm{Card}(var(E)) \geq 2$. Then the following statements are equivalent.

1. $\mathrm{Card}(W(E)) > 2$,
2. there exists E' such that $E \rightarrow_L E'$,
3. there exists E' such that $E \rightarrow_R E'$.

[6]It is worth noting that since basic RWEs are indecomposable, $\mathrm{Card}(W(E)) \geq 2$ whenever $\mathrm{Card}(var(E)) \geq 2$.

Fact 9.7 allows us to infer exactly the height of the tree by establishing a natural ordering (namely the cardinality of $W(E)$) on equations. Note that by Fact 9.6, whenever we move from a an equation to one of its children w.r.t. $\triangleright$, we decrease $\mathrm{Card}(W(E))$ by exactly one.

Fact 9.7 Let E_1, E_2 be basic RWEs with $\mathrm{Card}(var(E_1)), \mathrm{Card}(var(E_2)) \geq 2$. Let $Z \in \{L, R\}$ and suppose that $E_1 \rightarrow_Z E_2$. Suppose that $x, y \in X$ and let $\alpha_1, \alpha_2, \beta_1, \beta_2 \in (X \backslash \{x, y\})^*$ such that E_1 may be written $x\alpha_1 y\alpha_2 \doteq y\beta_1 x\beta_2$. If $Z = L$, then $W(E_2) = W(E_1) \backslash \{y\}$ and if $Z = R$, then $W(E_2) = W(E_1) \backslash \{x\}$.

Facts 9.7 and 9.6 are sufficient to observe that the set $\{E' \mid E \triangleright^* E'\}$ combined with $\triangleright$ forms a DAG of bounded height. However, this is not sufficient for our purposes of providing a lower bound on the number of equations contained in $\{E' \mid E \triangleright^* E'\}$. The following lemma shows that this DAG is in fact a tree by confirming that for each equation (which is not a leaf node), the two 'subtrees' rooted at the two children of that equation do not share any vertices.

Lemma 9.8 *Let E, E_1, E_2 be basic regular word equations such that $E \rightarrow_L E_1$ and $E \rightarrow_R E_2$. Let $S_1 = \{E'_1 \mid E_1 \triangleright^* E'_1\}$ and let $S_2 = \{E'_2 \mid E_2 \triangleright^* E'_2\}$. Then $S_1 \cap S_2 = \emptyset$ and $E \notin S_1 \cup S_2$.*

Proof The fact that $E \notin S_1 \cup S_2$ follows from the fact that, by Fact 9.7, for all $E' \in S_1 \cup S_2$, we have $\mathrm{Card}(W(E')) \leq \mathrm{Card}(W(E_1)) = \mathrm{Card}(W(E_2)) < \mathrm{Card}(W(E))$. We shall next consider the claim that $S_1 \cap S_2 = \emptyset$. Notice that it follows from the definitions of $\rightarrow_R$ and $\rightarrow_L$ that if $E' \triangleright E''$ and $w \in var(E') \backslash W(E')$, then firstly $w \in var(E'') \backslash W(E'')$, and secondly $Q_{E'}(w) = Q_{E''}(w)$ where $Q_{E'}, Q_{E''}$ are the functions defined in accordance with Definition 5.1. Now, if $\mathrm{Card}(W(E)) \leq 2$, then the statement follows trivially. Otherwise let $x, y, z_1, z_2, \ldots, z_k, w_1, w_2, \ldots, w_k \in X$ such that $\{z_1, z_2, \ldots, z_k\} = \{w_1, w_2, \ldots, w_k\}$, and $\alpha, \beta, \gamma_1, \gamma_2, \ldots, \gamma_k, \delta_1, \delta_2, \ldots, \delta_k \in (X \backslash \{x, y, z_1, z_2, \ldots, z_k\})^*$ such that $var(\gamma_i) \cap var(\delta_j) = \emptyset$ for $0 \leq i \leq k$ and such that E may be written as:

$$x\gamma_0 z_1 \gamma_1 z_2 \gamma_2 \ldots z_k \gamma_k y\alpha \doteq y\delta_0 w_1 \delta_1 w_2 \delta_2 \ldots w_k \delta_k x\beta.$$

From Fact 9.7, it follows that $y \notin W(E_1)$, so we may conclude that $Q_{E'}(y) = Q_{E_1}(y)$ for all $E' \in S_1$. Similarly, it follows from Fact 9.7 that $x \notin W(E_2)$, and we may hence conclude that $Q_{E'}(x) = Q_{E_2}(x)$ for all $E' \in S_2$. Now, let u, v be the rightmost variables in $z_k\gamma_k$ and $w_k\delta_k$ respectively. Then $Q_{E_1}(y) = Q_{E_2}(x) = (u, v)$. However, since E' is regular, $x \neq y$, so by properties of the functions $Q_{E'}$ (namely that by Remark 5.2 they are injective), we cannot have that $Q_{E'}(x) = (u, v)$ for any $E' \in S_1$ and likewise we cannot have $Q_{E'}(y) = (u, v)$ for any $E' \in S_2$. Consequently, $S_1 \cap S_2 = \emptyset$. □

Lemma 9.8, along with Facts 9.6 and 9.7, are sufficient to confirm our claim that the set $\{E' \mid E \triangleright^* E'\}$ forms a balanced binary tree of height $\mathrm{Card}(W(E)) - 2$. Thus we are now in a position to state the cardinality of $\{E' \mid E \triangleright^* E'\}$ precisely as follows.

Lemma 9.9 *Let E be a basic regular word equation such that $Card(W(E)) \geq 2$. Let $S = \{E' \mid E \triangleright^* E'\}$. Then $Card(S) = 2^{Card(W(E))-1} - 1$.*

Proof We shall prove the claim by induction on $\mathrm{Card}(W(E))$. If $\mathrm{Card}(W(E)) = 2$ then $S = \{E\}$ and the statement is immediate. Now suppose that the claim holds for all basic regular word equations E such that $\mathrm{Card}(W(E)) \leq n$ for some $n \geq 2$. Let E be a basic regular word equation such that $\mathrm{Card}(W(E)) = n + 1$. Then $\mathrm{Card}(W(E)) > 2$, so by Fact 9.6, there exist $E_1, E_2 \in [E]_{\Rightarrow}$ such that $E \rightarrow_L E_1$ and $E \rightarrow_R E_2$. From the definitions, we have that $S = \{E\} \cup S_1 \cup S_2$ where $S_1 = \{E_1' \mid E_1 \triangleright^* E_1'\}$ and $S_2 = \{E_2' \mid E_2 \triangleright^* E_2'\}$. By Lemma 9.8, it follows that $\mathrm{Card}(S) = 1 + \mathrm{Card}(S_1) + \mathrm{Card}(S_2)$. Moreover, since $\mathrm{Card}(W(E_1)) = \mathrm{Card}(W(E_2)) = n$, we have from our induction hypothesis that $\mathrm{Card}(S_1) = \mathrm{Card}(S_2) = 2^{n-1} - 1$. Thus we have $\mathrm{Card}(S) = 2(2^{n-1} - 1) + 1 = 2^{(n+1)-1} - 1$ as required. □

Lemma 9.9 together with Fact 9.5 are sufficient to provide lower bounds on the number of vertices of $\mathscr{G}^{\Rightarrow[E]}$, and we are nearly ready to provide the counterpart to Lemma 9.2. The final step before we do so is the following lemma which characterises the basic RWEs E for which the set of vertices of $\mathscr{G}^{\Rightarrow}_{[E]}$ is exactly $W(E)$. Since by Fact 9.5, $W(E)$ is always a subset of the vertices of $\mathscr{G}^{\Rightarrow}_{[E]}$, this naturally leads us to the extremal case in which the lower bound is obtained.

Lemma 9.10 *Let E be a basic regular word equation. Let $S = \{E' \mid E \triangleright^* E'\}$. Then $S = [E]_{\Rightarrow}$ if and only if E is regular reversed.*

Proof Let E be a basic regular word equation. If $\mathrm{Card}(var(E)) = 1$ then E can be written as $x = x$, for some $x \in X$, meaning that E is regular reversed, and moreover, that $S = [E]_{\Rightarrow} = \{E\}$, so the statement holds trivially. Suppose henceforth that $\mathrm{Card}(var(E)) \geq 2$.

Consider first the case that E' is not regular reversed for all $E' \in [E]_{\Rightarrow}$. Then by Lemma 6.13, there exists $E_1 \in [E]_{\Rightarrow}$ such that E_1 has the form $x\alpha y \doteq y\beta x$ where $x, y \in X$ and $\alpha, \beta \in (X\backslash\{x, y\})^*$. By our assumption, E_1 is not regular reversed. Hence we may write E_1 as:

$$x\alpha_1 u\alpha_2 v\alpha_3 y \doteq y\beta_1 u\beta_2 v\beta_3 x$$

where $x, y, u, v \in X$ and $\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3 \in (X\backslash\{x, y, u, v\})^*$. Thus, by Lemma 7.2 we have that $E_2 \in [E]_{\Rightarrow}$ where E_2 is given by $x\alpha_1 v\alpha_3 u\alpha_2 y \doteq y\beta_1 v\beta_3 u\beta_2 x$. However, $\mathrm{Card}(W(E_1)) = \mathrm{Card}(W(E_2)) = n$. Since by Fact 9.7, $E' \triangleright E''$ implies $\mathrm{Card}(W(E'')) < \mathrm{Card}(W(E'))$, and hence $\mathrm{Card}(W(E')) < \mathrm{Card}(W(E))$ for all $E' \in S\backslash\{E\}$, we may immediately conclude that at least one of $E_1, E_2 \notin S$, and hence $S \neq [E]_{\Rightarrow}$.

Now suppose that E is regular reversed. We have the following claim:

Claim 9.10.1 Let $E' \in S$ be given by $\alpha \doteq \beta$. Then the equation $\pi_{W(E')}(\alpha) \doteq \pi_{W(E')}(\beta)$ is regular reversed.

Proof We shall prove the claim by induction on $\mathrm{Card}(W(E'))$. In particular note that if $\mathrm{Card}(W(E')) = \mathrm{Card}(W(E))$, then by Fact 9.7, we have $E' = E$ and the statement holds trivially. Now suppose for some n that the claim holds for all $E' \in S$ with $\mathrm{Card}(W(E')) \geq n$. Let $E' \in S$ such that $\mathrm{Card}(W(E')) = n - 1$. By definition, since $E' \neq E$, there exists $E'' \in S$ such that $E'' \triangleright E'$. By Fact 9.7, we have also that $\mathrm{Card}(W(E'')) = n$. Assume w.l.o.g. that $E'' \rightarrow_R E'$. Then by the induction hypothesis, there exist $x, y, z_1, z_2, \ldots, z_{n-2} \in X$, and $\alpha, \beta, \gamma_0, \gamma_1, \gamma_2, \ldots, \gamma_k, \delta_0, \delta_1, \delta_2, \ldots, \delta_k \in (X \backslash \{x, y, z_1, z_2, \ldots, z_k\})^*$ such that $var(\gamma_i) \cap var(\delta_j) = \emptyset$ for $0 \leq i \leq k$ and such that E'' is given by

$$x\gamma_0 z_1 \gamma_1 z_2 \ldots z_k \gamma_k y\alpha \doteq y\delta_0 z_k \delta_1 z_{k-1} \delta_2 \ldots z_1 \delta_k x\beta$$

and E' is given by

$$z_1 \gamma_1 z_2 \ldots z_k \gamma_k x \gamma_0 y\alpha \doteq y\delta_0 z_k \delta_1 z_{k-1} \delta_2 \ldots z_1 \delta_k x\beta.$$

Note that $W(E') = W(E'') \backslash \{x\} = \{y, z_1, z_2, \ldots, z_k\}$. Erasing all the variables not in $W(E')$ from E' yields

$$z_1 z_2 \ldots z_k y \doteq y z_k z_{k-1} \ldots z_1$$

which is regular reversed so the statement of the claim holds for E'. By induction, it holds for all $E' \in S$ as required. □

Now suppose for contradiction that $[E]_{\Rightarrow} \neq S$. This implies that there exists $E' \in [E]_{\Rightarrow}$ such that $E' \notin S$. Now, by Fact 9.5, this implies that there exists a sequence $E_1, E_2, \ldots E_n$ such that $E_1 = E$, $E_n \notin S$ and such that either $E_i \triangleright E_{i+1}$ or $E_{i+1} \triangleright E_i$ for each $i, 1 \leq i < n$. Let us take the shortest such sequence. Note that this implies that $E_i \in S$ for all $i, 1 \leq i < n$, and consequently, that $E_i \triangleright E_{i+1}$ for all $i, 1 \leq i < n - 1$, and that $E_{n-1} \not\triangleright E_n$, meaning that $E_n \triangleright E_{n-1}$ instead. It follows from the fact that $W(E) = \mathrm{Card}(var(E))$, and by Fact 9.7 that there does not exist $E' \in [E]_{\Rightarrow}$ such that $E' \triangleright E$. Hence we may additionally conclude that $n > 2$. Moreover, since $E_{n-2} \in S$ and $E_n \notin S$, we have that $E_{n-2} \neq E_n$. Thus we must necessarily have that either $E_{n-2} \rightarrow_L E_{n-1}$ and $E_n \rightarrow_R E_{n-1}$, or symmetrically $E_{n-2} \rightarrow_R E_{n-1}$ and $E_n \rightarrow_L E_{n-1}$. W.l.o.g. we may assume the first case holds. Then it follows from the definitions that there exist $x_1, x_2, y_1, y_2, z_1, z_2, \alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3, \gamma_1, \gamma_2, \gamma_3, \delta_1, \delta_2, \delta_3$ such that $var(\alpha_1\alpha_2) \subseteq var(\beta_3)$, $var(\beta_1\beta_2) \subseteq var(\alpha_3)$, $var(\gamma_1\gamma_2) \subseteq var(\delta_3)$ and $var(\delta_1\delta_2) \subseteq var(\gamma_3)$, and such that E_{n-2} is given by $x_1\alpha_1 z_1 \alpha_2 y_1 \alpha_3 \doteq y_1 \beta_1 z_1 \beta_2 x_1 \beta_3$, E_n is given by $x_2\gamma_1 z_2 \gamma_2 y_2 \gamma_3 \doteq y_2 \delta_1 z_2 \delta_2 x_2 \delta_3$, and therefore that E_{n-1} can be written both as

$$z_1\alpha_2 x_1 \alpha_1 y_1 \alpha_3 \doteq y_1 \beta_1 z_1 \beta_2 x_1 \beta_3 \quad \text{and as} \quad x_2\gamma_1 z_2 \gamma_2 y_2 \gamma_3 \doteq z_2 \delta_2 y_2 \delta_1 x_2 \delta_3.$$

It follows that $x_2 = z_1$, $z_2 = y_1$, and thus that $\gamma_1 = \alpha_2 x_1 \alpha_1$, $\alpha_3 = \gamma_2 y_2 \gamma_3$, $\beta_1 = \delta_2 y_2 \delta_1$, and $\delta_3 = \beta_2 x_1 \beta_3$. Consequently, we may write E_{n-2} as:

$$x_1\alpha_1 z_1 \alpha_2 y_1 \gamma_2 y_2 \gamma_3 \doteq y_1 \delta_2 y_2 \delta_1 z_1 \beta_2 x_1 \beta_3.$$

Now, let E'_{n-1} be the equation

$$x_1\alpha_1 z_1 \alpha_2 y_1 \gamma_2 y_2 \gamma_3 \doteq y_2 \delta_1 z_1 \beta_2 y_1 \delta_2 x_1 \beta_3.$$

Since $var(\delta_2) \subseteq var(\gamma_3) \subseteq var(\alpha_3)$, we have that $var(\delta_2) \cap var(\alpha_1\alpha_2\gamma_2) = \emptyset$, and consequently, $E'_{n-1} \rightarrow_L E_{n-2}$. However, since $z_1, y_1 \in W(E'_{n-1})$, we can infer from Claim 1.10.1 that $E_{n-1} \notin S$. However, this contradicts our earlier assumption that the sequence $E_1, E_2, \ldots, E_n$ is minimal, since $E_1, E_2, \ldots E_{n-2}, E'_{n-1}$ also satisfies that $E_1 = E$, $E'_{n-1} \notin S$ and $E_i \triangleright E_{i+1}$ or $E_{i+1} \triangleright E_i$ for $1 \leq i < n-2$ and $E'_{n-1} \triangleright E_{n-2}$. Thus, we must have that $[E]_{\Rightarrow} = S$ as required. □

We are now ready to give the tight lower bounds on the number of vertices in $\mathscr{G}^{\Rightarrow}_{[E]}$, and to characterise those equations for which the lower bounds are achieved. The final step is to move from the bounds depending on $\mathrm{Card}(W(E))$ given by Lemma 9.9 to bounds depending on $\mathrm{Card}(var(E))$ by noting that by Lemma 6.13, there is always an equation in $\mathscr{G}^{\Rightarrow}_{[E]}$ for which $\mathrm{Card}(var(E)) = \mathrm{Card}(W(E))$.

Lemma 9.11 *Let E be a basic regular word equation. Let $n = Card(var(E))$ and suppose that $n \geq 2$. Let V be the number of vertices in $\mathscr{G}^{\Rightarrow}_{[E]}$. Then $V \geq 2^{n-1} - 1$. Moreover, $V = 2^{n-1} - 1$ if and only if E is regular reversed.*

Proof Let E be a basic regular word equation and let $n = \mathrm{Card}(var(E)) \geq 2$. Let $V = \mathrm{Card}([E]_{\Rightarrow})$ be the number of vertices in $\mathscr{G}^{\Rightarrow}_{[E]}$. W.l.o.g. by Lemma 6.13, we may assume that E has the form $x\alpha y \doteq y\beta x$ for some $x, y \in X$ and $\alpha, \beta \in (X\backslash\{x, y\})^*$. Thus $\mathrm{Card}(W(E)) = n$. Let $S = \{E' \mid E \triangleright^* E'\}$. Then by Fact 9.5, $S \subseteq [E]_{\Rightarrow}$. By Lemma 9.9, $\mathrm{Card}(S) = 2^{n-1} - 1$. Hence we have that $V \geq 2^{n-1} - 1$. Moreover, by Lemma 9.10, $S = [E]_{\Rightarrow}$ if and only if E is regular reversed. Hence $V = 2^{n-1} - 1$ if and only if there exists $E' \in [E]_{\Rightarrow}$ such that E' is regular reversed. □

It is worth noting that the lower bound given by Lemma 9.11 is already exponential in the number of variables, which, since we consider basic RWEs, is proportional to the length of the equation. In order to interpret these bounds in the more general (i.e. not basic) case we recall from Section 4 that for any RWE $\alpha \doteq \beta$, there exist prefixes α', β' of α and β respectively such that E' given by $\alpha' \doteq \beta'$ is indecomposable, and such that $\mathscr{G}^{\Rightarrow}_{[E]}$ is isomorphic to $\mathscr{G}^{\Rightarrow}_{[E']}$. In this case, the lower bound on the number of vertices in $\mathscr{G}^{\Rightarrow}_{[E]}$ becomes $2^{m-1} - 1$ where $m = \mathrm{Card}(qv(E'))$.

We conclude this section with the following theorem summarising the bounds on the number of vertices in $\mathscr{G}^{\Rightarrow}_{[E]}$.

Theorem 9.12 *Let E be a basic RWE and let $n = Card(var(\alpha))$. Suppose that $n > 1$. Let V be the number of vertices in $\mathscr{G}^{\Rightarrow}_{[E]}$. Then:*

1. $2^{n-1} - 1 \leq V \leq \frac{n!}{2}$,
2. *$V = 2^{n-1} - 1$ if and only if there exists $E' \in [E]_{\Rightarrow}$ such that E' is regular reversed,*
3. *$V = \frac{n!}{2}$ if and only if there exists $E' \in [E]_{\Rightarrow}$ such that E' is regular rotated.*

Proof Directly from Lemmata 9.2 and 9.11. □

10 DAG-Width

In addition to the size we are also able to give some insights about the connectedness of the graphs, which, as discussed in Section 3.3, are of interest when solving RWEs modulo additional constraints. We show firstly that there exist classes of equations E for which $dgw(\mathscr{G}_{[E]}^{\Rightarrow_{NT}})$ may be arbitrarily large.

Theorem 10.1 *Let $x, y, z_0, z_1, z_2, \ldots, z_n \in X$. Let E be the equation given by*

$$xz_0z_1z_2 \ldots z_n y \doteq yz_0z_nz_{n-1} \ldots z_1x.$$

Then $dgw(\mathscr{G}_{[E]}^{\Rightarrow_{NT}}) > n$.

To prove Theorem 10.1, we make use of the k-cops and robber games for directed graphs as introduced by [5]. The following definition is taken directly from [5].

Definition 10.2 (Cops and robber game [5]) Given a directed graph $G = (V, E)$, the k-cops and robber game on G is played between two players, the cop and the robber player. Positions of this game are pairs (X, r) where $X \in V^{\leq k}$ are the vertices occupied by the cops and $r \in V$ is the vertex occupied by the robber. The game is played as follows:

- At the beginning, the cop player chooses $X_0 \in V^{\leq k}$, and the robber player chooses a vertex $r_0 \in V$, giving position (X_0, r_0).
- From position (X_i, r_i), if $r_i \notin X_i$, then the cop player chooses $X_{i+1} \in V^{\leq k}$, and the robber player chooses a vertex $r_{i+1} \in V$ such that there is a directed path from r_i to r_{i+1} in the graph $G \backslash (X_i \cap X_{i+1})$.
- A play in the game is a maximal (finite or infinite) sequence $\pi = (X_0, r_0), (X_1, r_1), (X_2, r_2), \ldots$ of positions given by the rules above.
- A play π is winning for the cop player if and only if it is finite. (Note that, by the rules above, this implies that $r_m \in X_m$ for the last position (X_m, r_m) of this play.) A play π is winning for the robber player if and only if it is infinite.
- A (k-cop) strategy for the cop player is a function f from $V^{\leq k} \times V$ to $V^{\leq k}$. A play $(X_0, r_0), (X_1, r_1), \ldots$ is consistent with a strategy f if $X_{i+1} = f(X_i, r_i)$ for all i. The strategy f is called a winning strategy if every play consistent with the strategy is winning for the cop player.
- The cop number of a directed graph G is the least k such that the cop player has a strategy to win the k-cops and robber game on G.

It is shown in [5] (Theorem 16) that for any directed graph G, there is a DAG-decomposition of G of width at most k only if the cop player has a winning strategy in the k-cops and robber game on G. Thus, to show that a graph G has DAG-width greater than n, it is sufficient to show that there is no n-cop winning strategy in the n-cops and robber game on G. This equivalently amounts to providing a winning strategy for the robber. We shall use this fact to prove Theorem 10.1 as follows. Figure 8 provides an example and depicts how the winning strategy for the robber works.

Theorem 10.1. Note that it is sufficient to show that the DAG-width of $\mathscr{G}^{\Rightarrow}_{[E]}$ is greater than n, since $\mathscr{G}^{\Rightarrow}_{[E]}$ is a subgraph of $\mathscr{G}^{\Rightarrow^{NT}}_{[E]}$. For $0 \leq i \leq n$, let E_i be the (basic regular) equation given by:

$$xz_iz_{i+1}\ldots z_nz_0z_1z_2\ldots z_{i-1}y \doteq yz_iz_{i-1}\ldots z_1z_0z_nz_{n-1}\ldots z_{i+1}x$$

where $x, y, z_0, z_1, \ldots, z_n \in X$. Note that $E = E_0$. Let $V = [E]_{\Rightarrow}$. Before describing a winning strategy for the robber in the n-cops and robber game on $\mathscr{G}^{\Rightarrow}_{[E]}$, we define some useful subsets of vertices of $\mathscr{G}^{\Rightarrow}_{[E]}$ as follows. For each $i, 0 \leq i \leq n$ and each $j, 0 \leq i \leq n$ with $j > i$, let:

$$\begin{aligned}
T_i^j = \{&z_iz_{i+1}\ldots z_nz_0z_1\ldots z_{i-1}xy \doteq yz_iz_{i-1}\ldots z_0z_nz_{n-1}\ldots z_{i+1}x,\\
&z_{i+1}\ldots z_nz_0z_1\ldots z_{i-1}xz_iy \doteq yz_iz_{i-1}\ldots z_0z_nz_{n-1}\ldots z_{i+1}x,\\
&\qquad\vdots\\
&z_jz_{j+1}\ldots z_nz_0z_1\ldots z_{i-1}xz_iz_{i+1}\ldots z_{j-1}y \doteq yz_iz_{i-1}\ldots z_0z_nz_{n-1}\ldots z_{i+1}x\}\\
\cup\{&z_jz_{j+1}\ldots z_nz_0z_1\ldots z_{i-1}xz_iz_{i+1}\ldots z_{j-1}y \doteq\\
&\qquad z_iz_{i-1}\ldots z_0z_nz_{n-1}\ldots z_{j+1}yz_jz_{j-1}\ldots z_{i+1}x\}\\
\cup\{&z_{j+2}\ldots z_nz_0z_1\ldots z_{i-1}xz_jz_{j+1}z_iz_{i+1}\ldots z_{j-1}y \doteq\\
&\qquad z_iz_{i-1}\ldots z_0z_nz_{n-1}\ldots z_{j+1}yz_jz_{j-1}\ldots z_{i+1}x,\\
&\qquad\vdots\\
&z_{i-1}xz_jz_{j+1}\ldots z_nz_0z_1\ldots z_{i-2}z_iz_{i+1}\ldots z_{j-1}y \doteq\\
&\qquad z_iz_{i-1}\ldots z_0z_nz_{n-1}\ldots z_{j+1}yz_jz_{j-1}\ldots z_{i+1}x\}.
\end{aligned}$$

Similarly, for each $i, 0 \leq i \leq n$ and each $j, 0 \leq j \leq n$ with $j < i$, let:

$$\begin{aligned}
T_i^j = \{&z_iz_{i+1}\ldots z_nz_0z_1\ldots z_{i-1}xy \doteq yz_iz_{i-1}\ldots z_0z_nz_{n-1}\ldots z_{i+1}x,\\
&z_{i+1}\ldots z_nz_0z_1\ldots z_{i-1}xz_iy \doteq yz_iz_{i-1}\ldots z_0z_nz_{n-1}\ldots z_{i+1}x,\\
&\qquad\vdots\\
&z_jz_{j+1}\ldots z_{i-1}xz_iz_{i+1}\ldots z_nz_0z_1\ldots z_{j-1}y \doteq yz_iz_{i-1}\ldots z_0z_nz_{n-1}\ldots z_{i+1}x\}\\
\cup\{&z_jz_{j+1}\ldots z_{i-1}xz_iz_{i+1}\ldots z_nz_0z_1\ldots z_{j-1}y \doteq\\
&\qquad z_iz_{i-1}\ldots z_0z_nz_{n-1}\ldots z_{j+1}yz_jz_{j-1}\ldots z_{i+1}x\}\\
\cup\{&z_{j+2}\ldots z_{i-1}xz_jz_{j+1}z_iz_{i+1}\ldots z_nz_0z_1\ldots z_{j-1}y \doteq\\
&\qquad z_iz_{i-1}\ldots z_{j+1}yz_jz_{j-1}\ldots z_0z_nz_{n-1}\ldots z_{i+1}x,\\
&\qquad\vdots\\
&z_{i-1}xz_jz_{j+1}\ldots z_{i-2}z_iz_{i+1}\ldots z_nz_0z_1\ldots z_{j-1}y \doteq\\
&\qquad z_iz_{i-1}\ldots z_{j+1}yz_jz_{j-1}\ldots z_0z_nz_{n-1}\ldots z_{i+1}x\}.
\end{aligned}$$

For each $i, 0 \leq i \leq n$, let $S_i^{out} = \bigcup_{0\leq j\leq n, i\neq j} T_i^j$ and let

$$\begin{aligned}
S_i^{in} = \{&xz_iz_{i+1}\ldots z_nz_0z_1\ldots z_{i-1}y \doteq z_jz_{j-1}\ldots z_0z_nz_{n-1}\ldots z_{i+1}yz_iz_{i-1}\ldots z_{j+1}x \mid j \leq i\}\\
\cup\{&xz_iz_{i+1}\ldots z_nz_0z_1\ldots z_{i-1}y \doteq z_jz_{j-1}\ldots z_{i+1}yz_iz_{i-1}\ldots z_0z_nz_{n-1}\ldots z_{j+1}x \mid j > i\}.
\end{aligned}$$

Note that $S_i^{in} = \{E' \mid E' \Rightarrow_L^* E_i\}\backslash\{E_i\}$. Moreover, we shall now show that for each E_i, E_j with $i \neq j$, there exist $F_1, F_2, \ldots, F_k \in S_i^{out}$ and $G_1, G_2, \ldots G_\ell \in S_j^{in}$ such that

$$E_i \Rightarrow F_1 \Rightarrow F_2 \Rightarrow \ldots F_k \Rightarrow G_1 \Rightarrow G_2 \Rightarrow \ldots \Rightarrow G_\ell \Rightarrow E_j. \tag{4}$$

Indeed, observe that

$$\begin{aligned}
E_i &\Rightarrow z_i z_{i+1} \ldots z_n z_0 z_1 \ldots z_{i-1} xy \doteq y z_i z_{i-1} \ldots z_0 z_n z_{n-1} \ldots z_{i+1} x \\
&\Rightarrow z_{i+1} \ldots z_n z_0 z_1 \ldots z_{i-1} x z_i y \doteq y z_i z_{i-1} \ldots z_0 z_n z_{n-1} \ldots z_{i+1} x \\
&\qquad\qquad \vdots \\
&\Rightarrow z_j z_{j+1} \ldots z_n z_0 z_1 \ldots z_{i-1} x z_i z_{i+1} \ldots z_{j-1} y \doteq y z_i z_{i-1} \ldots z_0 z_n z_{n-1} \ldots z_{i+1} x \\
&\Rightarrow z_j z_{j+1} \ldots z_n z_0 z_1 \ldots z_{i-1} x z_i z_{i+1} \ldots z_{j-1} y \doteq \\
&\qquad z_i z_{i-1} \ldots z_0 z_n z_{n-1} \ldots z_{j+1} y z_j z_{j-1} \ldots z_{i+1} x \\
&\Rightarrow z_{j+2} \ldots z_n z_0 z_1 \ldots z_{i-1} x z_j z_{j+1} z_i z_{i+1} \ldots z_{j-1} y \doteq \\
&\qquad z_i z_{i-1} \ldots z_0 z_n z_{n-1} \ldots z_{j+1} y z_j z_{j-1} \ldots z_{i+1} x \\
&\qquad\qquad \vdots \\
&\Rightarrow z_{i-1} x z_j z_{j+1} \ldots z_n z_0 z_1 \ldots z_{i-2} z_i z_{i+1} \ldots z_{j-1} y \doteq \\
&\qquad z_i z_{i-1} \ldots z_0 z_n z_{n-1} \ldots z_{j+1} y z_j z_{j-1} \ldots z_{i+1} x \\
&\Rightarrow x z_j z_{j+1} \ldots z_n z_0 z_1 \ldots z_{i-1} z_i z_{i+1} \ldots z_{j-1} y \doteq \\
&\qquad z_i z_{i-1} \ldots z_0 z_n z_{n-1} \ldots z_{j+1} y z_j z_{j-1} \ldots z_{i+1} x \in S_j^{in}.
\end{aligned}$$

Thus, there exist $F_1, F_2, \ldots, F_k \in S_i^{out}$ and $G_1 \in S_j^{in}$ such that $E_i \Rightarrow F_1 \Rightarrow F_2 \Rightarrow \ldots \Rightarrow F_k \Rightarrow G_1$. By definition, $S_j^{in} = \{E' \mid E' \Rightarrow_L^* E_j\}\backslash\{E_j\}$, so it follows directly that there exist $G_2, \ldots, G_\ell \in S_j^{in}$ such that $G_1 \Rightarrow G_2 \Rightarrow \ldots \Rightarrow E_j$ as claimed.

Consequently, we may conclude that $S_i^{in} \cup S_i^{out}\{E_i\} \subset [E]_{\Rightarrow}$ for all $i, 0 \leq i \leq n$. Clearly, each E_i, $0 \leq i \leq n$ is not contained in any S_j^Z for $0 \leq j \leq n$ and $Z \in \{in, out\}$. Furthermore, since the RHS of every equation in S_i^{out} has either yz_i or z_i as a prefix, $S_i^{out} \cap S_j^{out} = \emptyset$ whenever $i \neq j$. Similarly since the LHS of every equation in S_i^{in} has xz_i as a prefix, $S_i^{in} \cap S_j^{in} = \emptyset$ whenever $i \neq j$. Since the LHS of all equations in S_i^{in} has x as a prefix, and since the LHS all equations in S_j^{out} does not have x as a prefix, we may conclude further that $S_i^Z \cap S_j^{Z'} = \emptyset$ for all $i \neq j$ and $Z, Z' \in \{in, out\}$.

We are now ready to give the strategy for the robber in the n-cops and robber game on $\mathscr{G}_{[E]}^{\Rightarrow}$. We shall say that E_i is a 'safe' vertex if $S_i^{in} \cup S_i^{out} \cup \{E_i\}$ contains no vertex with a cop on it. Since there are only n cops, it follows from the fact that the sets $S_i^{in} \cup S_i^{out} \cup \{E_i\}$ are pairwise disjoint that, at any given time, there must be at least one $i, 0 \leq i \leq n$ such that E_i is safe. By definition, if the robber is on a safe vertex, then there is no cop also on that vertex, so the play continues.

Clearly, if the cop player chooses an initial placement $X_0 \in [E]_{\Rightarrow}^{\leq n}$, then the robber may be placed on a safe vertex $r_0 = E_{i_1}$ for some $i_1, 0 \leq i_1 \leq n$. Now, suppose after k steps in the game the position is (X_k, r_k) where r_k is a safe vertex. Then we

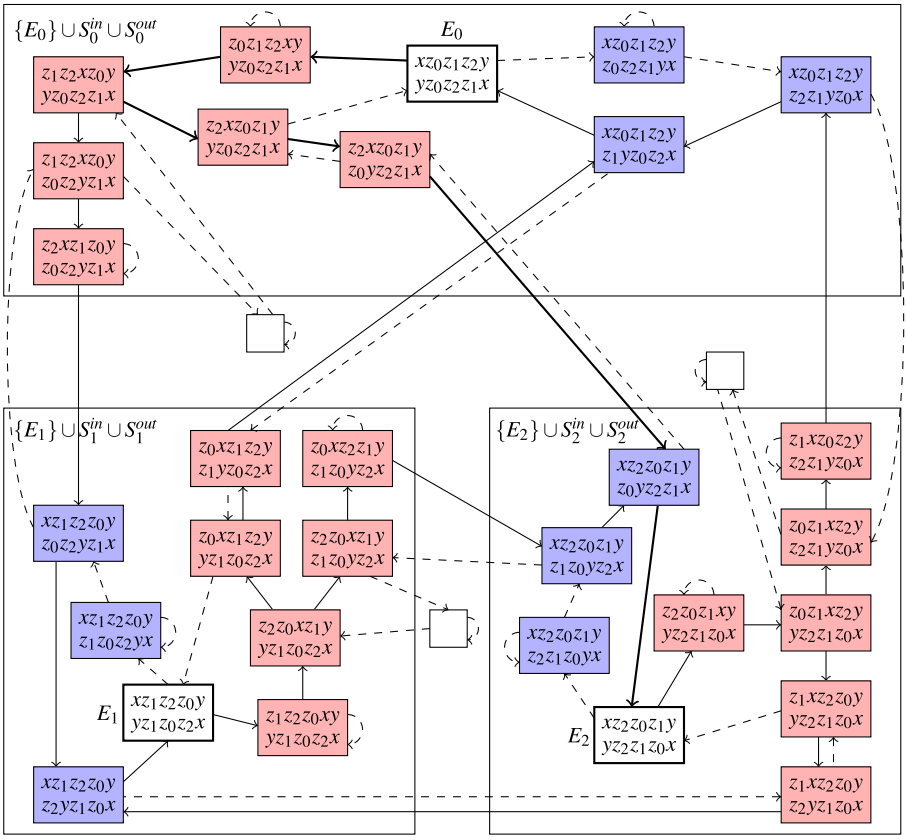

Fig. 8 A depiction of the graph $\mathscr{G}_{[E]}^{\Rightarrow}$ in the case that $E = xz_0z_1z_2y \doteq yz_0z_2z_1x$. Thus this is an example of Theorem 10.1 for the case $n = 2$. The graph is divided into sections corresponding to the (disjoint) sets $\{E_i\} \cup S_i^{in} \cup S_i^{out}$ for $0 \leq i \leq 2$. The vertices E_i are highlighted in bold while vertices from S_i^{in} are coloured blue and vertices from S_i^{out} are coloured red. In order to conserve space, vertices belonging to one of these sets are displayed with the LHS and RHS of the equation arranged vertically while for other vertices the equations are omitted. Since there are three values for i, if there are two cops, there will always be at least one i such that no vertex in $\{E_i\} \cup S_i^{in} \cup S_i^{out}$ has a cop on it. The strategy of the robber is to always be on E_i for such a choice of i. This is due to the fact that for each i and j, there is an path from E_i to E_j visiting only vertices from S_i^{out} and S_j^{in} which can be used as an escape-route (an example for $i = 1$ and $j = 3$ is highlighted in bold in the figure). Thus, if at any given stage in the game, a cop moves to a vertex in $\{E_i\} \cup S_i^{in} \cup S_i^{out}$, the robber can use the escape route to safely move to some E_j for which no vertex in $\{E_j\} \cup S_j^{in} \cup S_j^{out}$ has a cop on it. The edges making up the escape-route paths needed for this strategy are given by solid arrows, while the other edges which are not used by the robber are dashed

shall show that, whatever the cop player chooses for X_{k+1}, the robber may choose r_{k+1} such that r_{k+1} is safe. Indeed, if $r_k = E_{i_k}$ for some $i_k, 0 \leq i_k \leq n$ is safe, then $(S_{i_k}^{out} \cup \{E_{k_i}\}) \cap X_k = \emptyset$. Moreover, since there are only n cops, whatever the choice of X_{k+1}, there exists $r_{k+1} = E_{i_{k+1}}$ for some $i_{k+1}, 0 \leq i_{k+1} \leq n$ such that $E_{i_{k+1}}$ is safe, meaning that $X_{k+1} \cap (S_{i_{k+1}}^{in} \cup \{E_{i_{k+1}}\}) = \emptyset$. It follows that $S_{i_k}^{out} \cup S_{i_{k+1}}^{in} \cup \{E_{i_k}, E_{i_{k+1}}\} \subset [E]_{\Rightarrow} \setminus (X_{k+1} \cap X_k)$. We have already shown (Equation 4) that there is a directed path in $\mathscr{G}_{[E]}^{\Rightarrow}$ using only vertices from $S_{k_i}^{out} \cup S_{k_{i+1}}^{in} \cup \{E_{i_k}, E_{i_{k+1}}\}$ from $r_k (= E_{i_k})$ to

$r_{k+1}(= E_{i_{k+1}})$, and hence (X_{k+1}, r_{k+1}) is a valid next position satisfying the rules of the game. Since r_{k+1} is also safe, this proves our claim, and by a simple induction, it follows that for any n-cop strategy, there is an infinite play (i.e. robber wins). It follows that there is no winning n-cop strategy, so the DAG-width of $\mathscr{G}^{\Rightarrow}_{[E]}$ is greater than n as required. □

Since high connectivity can be seen as an obstacle to deciding the satisfiability problem with additional constraints, it is also worth noting classes for which the DAG-width is bounded by a small constant. If all variables occur at most once in an equation E, then it is not difficult to see that the graph $\mathscr{G}^{\Rightarrow_{NT}}_{[E]}$ will be a DAG. However, when variables may occur more than once, the graphs of even very simple equations such as $x\mathtt{ab} \doteq \mathtt{ba}x$ will contain cycles, and will therefore have DAG-width at least two. The following theorem describes an infinite class of equations for which the DAG-width of $\mathscr{G}^{\Rightarrow_{NT}}_{[E]}$ is at most two. It is worth pointing out that the NP-hardness result for the satisfiability problem for regular word equations from [8] applies to this class, and so, by Theorem 8.12, this class also has an NP-complete satisfiability problem.

Theorem 10.3 *Let $\alpha_1, \alpha_2, \ldots, \alpha_n, \beta_1, \beta_2, \ldots, \beta_n \in X^*$ such that*

1. $|\alpha_i| = |\beta_i| \in \{1, 2, 3\}$ *for* $1 \leq i \leq n$, *and*
2. $var(\alpha_i) = var(\beta_i)$ *for* $1 \leq i \leq n$, *and*
3. $var(\alpha_i) \cap var(\alpha_j) = \emptyset$ *for* $1 \leq i, j \leq n$ *with* $i \neq j$.

Let E be the RWE $\alpha_1\alpha_2 \ldots \alpha_n \doteq \beta_1\beta_2 \ldots \beta_n$. Then $dgw(\mathscr{G}^{\Rightarrow_{NT}}_{[E]}) \leq 2$.

Proof Let E be of the form described in the theorem. By Proposition 3.5,

$$dgw(\mathscr{G}^{\Rightarrow_{NT}}_{[E]}) = \max\{m \mid E \Rightarrow^*_{NT} E' \text{ and } m = dgw(\mathscr{G}^{\Rightarrow}_{[E']})\}.$$

Let $\mathscr{C}$ be the subclass of RWEs of the form $\alpha_1\alpha_2 \ldots \alpha_k \doteq \beta_1\beta_2 \ldots \beta_k$ where $k \in \mathbb{N}_0$ such that:

1. $\alpha_i, \beta_i \in X^*$with $|\alpha_i| = |\beta_i| \in \{1, 2, 3\}$ for $1 \leq i \leq k$, and
2. $var(\alpha_i) = var(\beta_i)$ for $1 \leq i \leq k$, and
3. $var(\alpha_i) \cap var(\alpha_j) = \emptyset$ for all $i \neq j$, $1 \leq i, j \leq k$.

Clearly, we have $E \in \mathscr{C}$. Since k is not restricted, we may also assume w.l.o.g. that for any word equation in $\mathscr{C}$, the 'sub-equations' $\alpha_i \doteq \beta_i$ are indecomposable. Moreover, if E' is not the equation $\varepsilon \doteq \varepsilon$, we may also assume that $|\alpha_1| \geq 1$. Under these assumptions, it follows from Corollary 4.4 that for any $E' \in \mathscr{C}$, the graph $\mathscr{G}^{\Rightarrow}_{[E']}$ is isomorphic to the graph $\mathscr{G}^{\Rightarrow}_{[\alpha_1 \doteq \beta_1]}$. There are four possibilities for $\alpha_1 \doteq \beta_1$ (up to a renaming of the variables, which does not alter the structure of the graph $\mathscr{G}^{\Rightarrow}_{[\alpha_1 \doteq \beta_1]}$), namely $x \doteq x$, $xy \doteq yx$, $xyz \doteq zyx$, $xyz \doteq yzx$ and $xyz \doteq zxy$. It is easily verified by hand that in all cases the DAG-width is at most two (it is exactly two in the cases where $|\alpha_1| = |\beta_1| = 3$). Moreover, it follows from the definitions that if $E_1 \in \mathscr{C}$ and $E_1 \Rightarrow_{NT} E_2$ for some E_2, then $E_2 \in \mathscr{C}$. Consequently, we have that

$$dgw(\mathscr{G}^{\Rightarrow_{NT}}_{[E]}) = \max\{m \mid E \Rightarrow^*_{NT} E' \text{ and } m = dgw(\mathscr{G}^{\Rightarrow}_{[E']})\} \leq 2.$$

□

11 Extension to Systems of Equations

So far, we have considered individual equations. However, it is often the case that there is not just one equation to be solved, but a system of several equations which should be satisfied concurrently. However, while constructions exist which transform a system of equations into a single equation (see e.g. [17]), the resulting equation will generally not be quadratic/regular. We extend the definition of regular equations to regular systems as follows.

Definition 11.1 (Regular systems) Let $\Theta = \{\alpha_1 \doteq \beta_1, \alpha_2 \doteq \beta_2, \ldots, \alpha_n \doteq \beta_n\}$ be a system of word equations. An *orientation* of Θ is any element of $\{\alpha_1 \doteq \beta_1, \beta_1 \doteq \alpha_1\} \times \{\alpha_2 \doteq \beta_2, \beta_2 \doteq \alpha_2\} \times \ldots \times \{\alpha_n \doteq \beta_n, \beta_n \doteq \alpha_n\}$. We say that Θ is regular if it has an orientation for which each variable occurs at most once across all LHSs and at most once across all RHSs.

We can easily adapt the algorithm from Section 3 to work more generally for systems of word equations, and with careful application, still make use of Theorem 8.11 in order to obtain (non-deterministic) polynomial running time. To do this, we need to extend the rewriting transformations (Nielsen transformations) underpinning the relation $\Rightarrow_{NT}$ which we have thus far defined for single equations only. Note that each possible rewriting of a single equation can be achieved by firstly applying a morphism to both sides of the equation then followed, if applicable, by cancelling the longest identical prefixes of the new LHS and RHS. For example, the rewriting $x\mathtt{a}yz\mathtt{ba} \doteq y\mathtt{b}w\mathtt{b}z\mathtt{a} \Rightarrow_{NT} \mathtt{a}xyz\mathtt{ba} \doteq y\mathtt{b}w\mathtt{b}z\mathtt{a}$ consists of applying the morphism $\psi_{y>x}$ (cf. Section 3) to both sides of the first equation in order to get $x\mathtt{a}xyz\mathtt{ba} \doteq xy\mathtt{b}w\mathtt{b}z\mathtt{a}$ and then cancelling the resulting leftmost occurrences of x.

The generalisation of the Nielsen Transformations to systems of equations is straightforward: we select one of the word equations E from the system, and apply any of the possible transformations to it as before. Then we simply need to apply the associated morphism to both sides of all the other equations in the system, followed by any further resulting cancellations. We shall say that such a transformation is rooted on the chosen equation E, and we shall write $\Theta \Rightarrow^{E}_{NT} \Theta'$ if Θ, Θ' are systems of word equations such that Θ' is the result of applying a transformation rooted on E to Θ. So if, for example, we have the system $\{x\mathtt{a}yz\mathtt{ba} \doteq y\mathtt{b}w\mathtt{b}z\mathtt{a}, w\mathtt{ba} \doteq \mathtt{ab}x\}$, then one possible transformation of the first equation is $x\mathtt{a}yz\mathtt{ba} \doteq y\mathtt{b}w\mathtt{b}z\mathtt{a} \Rightarrow_{NT} x\mathtt{a}yz\mathtt{ba} \doteq \mathtt{b}w\mathtt{b}z\mathtt{a}$ obtained by applying the morphism $\psi_{x>y}$ and cancelling the resulting leftmost occurrences of y. To extend this transformation to the whole system, we just need to apply $\psi_{x>y}$ to the other equation (no further cancellation is required in this case) so we have $\{x\mathtt{a}yz\mathtt{ba} \doteq y\mathtt{b}w\mathtt{b}z\mathtt{a}, w\mathtt{ba} \doteq \mathtt{ab}x\} \Rightarrow^{E}_{NT} \{x\mathtt{a}yz\mathtt{ba} \doteq \mathtt{b}w\mathtt{b}z\mathtt{a}, w\mathtt{ba} \doteq \mathtt{ab}yx\}$ where E is the equation $x\mathtt{a}yz\mathtt{ba} \doteq y\mathtt{b}w\mathtt{b}z\mathtt{a}$.

Taking the length $|\Theta|$ of a system Θ of word equations to be the sum of the lengths of all the individual word equations, it is easily seen that the important properties of this rewriting carry over to the case of systems. Specifically, it is easily verified that for any regular system Θ of word equations each of the following holds:

1. If $E \in \Theta$ and $\Theta \Rightarrow^{E}_{NT} \Theta'$, then Θ' is also regular,

2. If $E \in \Theta$ and $\Theta \Rightarrow_{NT}^{E} \Theta'$, then $|\Theta'| \leq |\Theta|$,
3. for any solution h to Θ, and for any $E \in \Theta$ with $|E| > 0$ there exists a system Θ' with a solution h' such that $\Theta \Rightarrow_{NT}^{E} \Theta'$ and either h' is smaller than h or $|\Theta'| < |\Theta|$.

With this in mind, we are now able to extend our main result that solving regular word equations is in NP to include regular systems of equations.

Theorem 11.2 *The satisfiability problem for regular systems of equations is NP-complete. Moreover, whether a system of word equations is regular can be decided in polynomial time.*

Proof Since the satisfiability problem is NP-hard for regular word equations, it is also NP-hard for regular systems of word equations. Next we shall show inclusion in NP. Let $\Theta = \{E_1, E_2, \ldots, E_n\}$ be a regular system of equations. From Observations 1-3 above, there is a solution to Θ if and only if there exists a finite sequence of transformations

$$\Theta_0 \Rightarrow_{NT}^{\hat{E}_1} \Theta_1 \Rightarrow_{NT}^{\hat{E}_2} \ldots \Rightarrow_{NT}^{\hat{E}_m} \Theta_m$$

satisfying $\Theta = \Theta_0$, $\Theta_m = \{\varepsilon \doteq \varepsilon\}$ and $\hat{E}_i \in \Theta_{i-1}$ for $1 \leq i \leq m$. In fact, by Observation 3, we may freely choose each $\hat{E}_i$ to be any equation from Θ_{i-1}, and such a finite sequence must still exist whenever there is a solution. Consequently, we may decide whether or not a solution exists with the following procedure (Algorithm 1) which searches for such a sequence by applying firstly transformations rooted on the first equation, followed transformations rooted on the second equation, then the third, etc. For convenience, we shall represent Θ as an ordered list $[E_1, E_2, \ldots, E_n]$ rather than a set.

Algorithm 1 Deciding if regular system of word equations has a solution.

Input: A regular system of word equations given as an (ordered) list $[E_1, E_2, \ldots, E_n]$

Output: "Yes" if the system $\{E_1, E_2, \ldots, E_n\}$ has a solution and "No" otherwise

```
1: Θ ← [E_1, E_2, ..., E_n]
2: for i, 1 ≤ i ≤ n do
3:     counter ← 0
4:     while Θ[i] ≠ ε ≐ ε ∧ counter ≤ C_Θ do
5:         Choose Θ' such that Θ ⇒_NT^{Θ[i]} Θ'
6:         Θ ← Θ'
7:     if Θ[i] ≠ ε ≐ ε then
8:         Return "No"
9: Return "Yes"
```

We begin by non-deterministically applying a sequence of Nielsen transformations (generalised for systems of word equations) rooted on the first equation in the

list until we reach a system of the form $[\varepsilon \doteq \varepsilon, E'_2, \ldots, E'_n]$. If we are not able to transform E_1 into $\varepsilon \doteq \varepsilon$, then no solution to E_1 exists and the system has no solution.

Otherwise, once we have transformed E_1 into the $\varepsilon \doteq \varepsilon$, we repeat the process of applying the generalised Nielsen transformations to the (new) second equation E'_2 until it has also been transformed into $\varepsilon \doteq \varepsilon$ (note that none of the transformations will change the trivial equation $\varepsilon \doteq \varepsilon$). Continue to repeat this process for each equation, in increasing order, until either an equation is reached which cannot be transformed into $\varepsilon \doteq \varepsilon$, or until we have transformed all equations into this form. In the former case, there is no solution, while in the latter case, a solution exists.

It remains to be seen that we can implement the procedure just described such that it runs in non-deterministic polynomial time. For this, we need a few further observations. The first is that when applying transformations rooted on the i^{th} equation, we are essentially traversing the same graph $\mathscr{G}^{\Rightarrow NT}_{[\tilde{E}_i]}$ as if we were to consider in isolation the equation $\tilde{E}_i$ obtained after transforming the first $i-1$ equations into $\varepsilon \doteq \varepsilon$. The only difference is that we are potentially changing the other equations as we go. The second important observation is that any transformation rooted on the i^{th} equation which changes any of the other (non-root) equations must necessarily decrease the length of the i^{th} equation. Finally, the equation on which a transformation is rooted never increases in length as a result of that transformation. Thus, by applying the transformations in the order specified, we never increase the length of i^{th} equation once it becomes the current root.

Consequently, when applying transformations which preserve the length of the i^{th} equation, we may, without affecting the outcome, take the shortest path through the graph. Moreover, since we can only decrease the length of an equation a linear number of times, the maximum number of transformations rooted on the i^{th} equation needed in order to find a solution when one exists is bounded above by

$$C_i = |\tilde{E}_i| \max\{diam(\mathscr{G}^{\Rightarrow}_{[E]}) \mid \tilde{E}_i \Rightarrow^*_{NT} E\}.$$

By Theorem 8.11, we can easily compute an upper bound $C_\Theta \geq \max\{C_i \mid 1 \leq i \leq n\}$ on the number of transformations needed which allows us to restrict the above procedure such that it works in non-deterministic polynomial time without affecting the correctness.

Finally, we describe the following procedure (Algorithm 2) for determining if a system $\Theta = \{E_1, E_2, \ldots, E_n\}$ is regular. First we check that each individual equation is regular and that no variable occurs more than twice across the whole system. We then initialise two sets L and R to the empty set. The sets L and R will keep track of variables occurring across the LHS's and RHS's of an orientation of Θ. We remove equations $\alpha \doteq \beta$ from Θ one-by-one, deciding each time whether $\alpha \doteq \beta$ or $\beta \doteq \alpha$ should be included in the orientation and updating L and R accordingly.

While there are still equations left in the system, there are two cases to consider. The first is that there exists an equation $\alpha \doteq \beta \in \Theta$ which contains at least one variable x which is already in L or R. In this case, we can rule out at least one choice of $\alpha \doteq \beta$ or $\beta \doteq \alpha$ when constructing an orientation satisfying the definition for regular systems. In particular, if $x \in L$, then whichever of α, β contains x should be the RHS in the orientation (so, if x occurs in α, we include $\beta \doteq \alpha$ in the orientation

instead of $\alpha \doteq \beta$). Likewise if $x \in R$ then whichever of α, β contains x should be the LHS. Once we have decided which of $\alpha \doteq \beta$ and $\beta \doteq \alpha$ is a bad choice (in that it would lead to two occurrences of x in either the LHS's or RHS's), we need to check that the remaining "oriented" equation does not lead to a similar conflict (possibly for one of the other variables). To do this, we simply need to check that the LHS does not share any variables with L and likewise that the RHS does not share any variables with R. If this test is failed then our system is not regular and we can stop and return "No". Otherwise we add all the variables from the LHS of the oriented equation to L and all the variables from the RHS to R. Then we remove the equation $\alpha \doteq \beta$ from the system Θ and continue.

Algorithm 2 Deciding if a system of word equations is regular.

Input: A system of word equations $\{E_1, E_2, \ldots, E_n\}$
Output: "Yes" if the system $\{E_1, E_2, \ldots, E_n\}$ is regular and "No" otherwise

$L, R \leftarrow \emptyset$
$S \leftarrow \{E_1, E_2, \ldots, E_n\}$
if Any variable occurs more than twice across all the E_is, or any variable occurs twice on the same side of a single equation E_i **then**
    Return "No"
while $S \neq \emptyset$ **do**
    if $\exists \alpha \doteq \beta \in S.var(\alpha\beta) \cap (L \cup R) \neq \emptyset$ **then**
        $S \leftarrow S \backslash \{\alpha \doteq \beta\}$
        if $var(\alpha) \cap L \neq \emptyset \vee var(\beta) \cap R \neq \emptyset$ **then**
            $LHS_{Oriented} \leftarrow \beta$
            $RHS_{Oriented} \leftarrow \alpha$
        else
            $LHS_{Oriented} \leftarrow \alpha$
            $RHS_{Oriented} \leftarrow \beta$
        if $var(LHS_{Oriented}) \cap L = var(RHS_{Oriented}) \cap R = \emptyset$ **then**
            $L \leftarrow L \cup var(LHS_{Oriented})$
            $R \leftarrow R \cup var(RHS_{Oriented})$
        else
            Return "No"
    else
        Choose any $\alpha \doteq \beta$ from S
        $S \leftarrow S \backslash \{\alpha \doteq \beta\}$
        $L \leftarrow L \cup var(\alpha)$
        $R \leftarrow R \cup var(\beta)$
Return "Yes"

The second case is when none of the variables occurring in the remaining equations are contained in either L or R. In this case, how we construct the rest of the orientation is not dependant on the previous choices. Moreover, for any orientation satisfying the definition, we can find another by simply swapping the LHS's and RHS's of all equations. Thus by symmetry, we may include any single one of the

remaining equations in the orientation without exchanging the LHS and RHS, and without affecting the possibility of constructing a valid orientation in the end. Thus, we then pick any of the remaining equations $\alpha \doteq \beta$ at random and add the variables from α to L and all the variables from β to R, before removing $\alpha \doteq \beta$ from Θ and continuing. If we are able to iterate through and discard all equations in the system like this without returning "No", then the system is regular and we may return "Yes". The correctness, along with the fact that the procedure runs in polynomial time are easily verified. □

12 Conclusions

A famous algorithm for solving quadratic word equations can be used to produce a (directed) graph containing all solutions to the equation. In the case of regular equations, we have described some underlying structures of these graphs with the intention of better understanding their solution sets. We give bounds on their diameter and number of vertices, as well as provide classes with bounded (resp. unbounded) DAG-width. Probably the most significant result arising from our analysis is that the satisfiability problem for regular word equations is in NP (and thus NP-complete), which we also extend to regular systems of equations.

We leave open many interesting problems, the most obvious of which is to generalise our results to the (full) quadratic case. We also believe that our analysis and techniques open up the possibility to investigate in far more detail the graphs $\mathscr{G}^{\Rightarrow}_{[E]}$, both in the case of regular equations and more generally. For example, in light of our results, it seems reasonable to suggest that determining whether $E_1 \Rightarrow^* E_2$ for two regular equations E_1 and E_2 may be done in polynomial time. A particularly nice characterisation of E_1 and E_2 such that $E_1 \Rightarrow^* E_2$ might yield a much quicker algorithm than the one resulting from our bound on the diameter of $\mathscr{G}^{\Rightarrow_{NT}}_{[E]}$ by significantly reducing the degree of the polynomial. We also expect that a detailed analysis of the length-reducing transformations and symmetries which may be found there would be particularly helpful in understanding further the structure of solution sets and the performance of algorithms solving regular equations in practice.

Finally, we mention the task of investigating the decidability of the satisfiability problem for regular equations with additional constraints, in particular length constraints, with the hope that having identified cases where the DAG-width is particularly high/low, along with improved means to describe precisely the structure of the solution-graphs, might provide some useful hints with how to proceed in this direction.

Acknowledgements We thank the anonymous referees for their detailed and thoughtful comments.

References

1. Abdulla, P.A., Atig, M.F., Chen, Y., Holík, L., Rezine, A., Rümmer, P., Stenman, J.: Norn: An SMT Solver for String Constraints. In: Proc. Computer Aided Verification (CAV), Lecture Notes in Computer Science (LNCS), vol. 9206, pp. 462–469 (2015)
2. Alkhalaf, M., Bultan, T., Yu, F.: STRANGER: an Automata-Based String Analysis Tool for PHP. In: Proc. Tools and Algorithms for the Construction and Analysis of Systems (TACAS), Lecture Notes in Computer Science (LNCS), vol. 6015 (2010)
3. Angluin, D.: Finding patterns common to a set of strings. J. Comput. Syst. Sci. **21**, 46–62 (1980)
4. Barrett, C., Conway, C.L., Deters, M., Hadarean, L., Jovanović, D., King, T., Reynolds, A., Tinelli, C.: CVC4. In: Proc. Computer Aided Verification (CAV), Lecture Notes In Computer Science (LNCS), vol. 6806, pp. 171–177 (2011)
5. Berwanger, D., Dawar, A., Hunter, P., Kreutzer, S., Obdrzálek, J.: The DAG-width of directed graphs. J Combin Theory Series B **102**(4), 900–923 (2012)
6. Berzish, M., Ganesh, V., Zheng, Y.: Z3str3: a String Solver with Theory-Aware Heuristics. In: Proc. Formal Methods in Computer-Aided Design (FMCAD), pp. 55–59. IEEE (2017)
7. Day, J.D., Ganesh, V., He, P., Manea, F., Nowotka, D.: The Satisfiability of Word Equations: Decidable and Undecidable Theories. In: Potapov, I., Reynier, P. (eds.) Proc. 12th International Conference on Reachability Problems, RP 2018, Lecture Notes in Computer Science (LNCS), vol. 11123, pp. 15–29 (2018)
8. Day, J.D., Manea, F., Nowotka, D.: The Hardness of Solving Simple Word Equations. In: Proc. Mathematical Foundations of Computer Science (MFCS), LIPIcs, vol. 83, pp. 18:1–18:14 (2017)
9. Day, J.D., Manea, F., Nowotka, D.: Upper Bounds on the Length of Minimal Solutions to Certain Quadratic Word Equations. In: Proc. Mathematical Foundations of Computer Science (MFCS), LIPIcs, vol. 138, pp. 44:1–44:15 (2019)
10. Diekert, V., Jeż, A., Plandowski, W.: Finding all solutions of equations in free groups and monoids with involution. Inf. Comput. **251**, 263–286 (2016)
11. Diekert, V., Robson, J.M.: On Quadratic Word Equations. In: Proc. 16Th Annual Symposium on Theoretical Aspects of Computer Science, STACS, Lecture Notes in Computer Science (LNCS), vol. 1563, pp. 217–226 (1999)
12. Ehrenfeucht, A., Rozenberg, G.: Finding a homomorphism between two words is NP-complete. Inf. Process. Lett. **9**, 86–88 (1979)
13. Freydenberger, D.D.: A logic for document spanners. Theory of Computing Systems **63**(7), 1679–1754 (2019)
14. Freydenberger, D.D., Holldack, M.: Document spanners: From expressive power to decision problems. Theory of Computing Systems **62**(4), 854–898 (2018)
15. Jeż, A.: Recompression: a simple and powerful technique for word equations. J. ACM 63 (2016)
16. Jeż, A.: Word Equations in Nondeterministic Linear Space. In: Proc. International Colloquium on Automata, Languages and Programming (ICALP), LIPIcs, vol. 80, pp. 95:1–95:13 (2017)
17. Karhumäki, J., Mignosi, F., Plandowski, W.: The expressibility of languages and relations by word equations. J. ACM **47**, 483–505 (2000)
18. Kiezun, A., Ganesh, V., Guo, P.J., Hooimeijer, P., Ernst, M.D.: HAMPI: a Solver for String Constraints. In: Proc. ACM SIGSOFT International Symposium on Software Testing and Analysis (ISSTA), pp. 105–116. ACM (2009)
19. Lin, A.W., Barceló, P.: String Solving with Word Equations and Transducers: Towards a Logic for Analysing Mutation Xss. In: ACM SIGPLAN Notices, vol. 51, pp. 123–136. ACM (2016)
20. Lin, A.W., Majumdar, R.: Quadratic Word Equations with Length Constraints, Counter Systems, and Presburger Arithmetic with Divisibility. In: Lahiri, S.K., Wang, C. (eds.) Proc. 16th International Symposium on Automated Technology for Verification and Analysis (ATVA), Lecture Notes in Computer Science (LNCS), vol. 11138, pp. 352–369. Springer (2018)
21. Lothaire, M.: Algebraic Combinatorics on Words. Cambridge University Press, New York (2002)
22. Makanin, G.S.: The problem of solvability of equations in a free semigroup. Sbornik: Mathematics **32**(2), 129–198 (1977)
23. Manea, F., Nowotka, D., Schmid, M.L.: On the complexity of solving restricted word equations. Int. J. Found. Comput. Sci. **29**(5), 893–909 (2018)
24. Petre, E.: An Elementary Proof for the Non-Parametrizability of the Equation $Xyz = Zvx$. In: Proc. 29Th International Symposium on Mathematical Foundations of Computer Science (MFCS), Lecture Notes in Computer Science (LNCS), vol. 3153, pp. 807–817 (2004)

25. Plandowski, W.: Satisfiability of Word Equations with Constants is in PSPACE. In: Proc. Foundations of Computer Science (FOCS), pp. 495–500. IEEE (1999)
26. Plandowski, W., Rytter, W.: Application of Lempel-Ziv Encodings to the Solution of Words Equations. In: Proc. International Colloquium on Automata, Languages and Programming (ICALP), Lecture Notes in Computer Science (LNCS), vol. 1443, pp. 731–742 (1998)
27. Schulz, K.U.: Makanin's Algorithm for Word Equations-Two Improvements and a Generalization. In: International Workshop on Word Equations and Related Topics, pp. 85–150. Springer (1990)