Similarity and bisimilarity notions appropriate for characterizing indistinguishability in fragments of the calculus of relations

GEORGE H.L. FLETCHER, *Department of Mathematics and Computer Science, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands.*
E-mail: g.h.l.fletcher@tue.nl

MARC GYSSENS and DIRK LEINDERS, *Databases and Theoretical Computer Science, Hasselt University and transnational University of Limburg, Martelarenlaan 42, 3500 Hasselt, Belgium.*
E-mail: marc.gyssens@uhasselt.be; dirk.leinders@uhasselt.be

JAN VAN DEN BUSSCHE, *Hasselt University and transnational University of Limburg, Martelarenlaan 42, 3500 Hasselt, Belgium.*
E-mail: jan.vandenbussche@uhasselt.be.

DIRK VAN GUCHT, *School of Informatics and Computing, Indiana University, Lindley Hall, Bloomington, IN 47405, USA.*
E-mail: vgucht@cs.indiana.edu

STIJN VANSUMMEREN, *Université Libre de Bruxelles (ULB), 50, Av. F.Roosevelt, CP 165/15, 1050 Brussels, Belgium.*
E-mail: stijn.vansummeren@ulb.ac.be

Abstract

Motivated by applications in databases, this article considers various fragments of the calculus of binary relations. The fragments are obtained by leaving out, or keeping in, some of the standard operators, along with some derived operators such as set difference, projection, coprojection and residuation. For each considered fragment, a characterization is obtained for when two given binary relational structures are indistinguishable by expressions in that fragment. The characterizations are based on appropriately adapted notions of simulation and bisimulation.

Keywords: Calculus of relations, indistinguishability, bisimulation, simulation, coprojection, residuation.

1 Introduction

The calculus of relations [16, 25, 33, 36] consists of five natural operations on binary relations: union, intersection, complementation, composition and converse. These operators can be applied

to given binary relations, combined with the four standard constant relations: empty, all, identity and diversity. The calculus of relations is a very natural formalism and occurs within logics for reasoning about binary relations, notably dynamic and description logics [5, 20]. The calculus also has historically motivated the development of the theory of relation algebras [22, 26]. In the present study, however, we are not looking at abstract relation algebras, but rather at the question of *indistinguishability* of two given finite binary relational structures within the calculus of relations.

This article has been inspired by the authors' ongoing research program to understand in a precise way the expressive power of the calculus of relations as a database query language for binary relation structures [11, 12, 18, 19, 35, 39]. Indistinguishability of structures in various logics is one of the most basic tools in the study of the expressive power of database query languages as well as in finite model theory [2, 9, 24]. Indeed the calculus of relations, as a core relational algebra query language on binary relations, is very relevant to the field of databases. Binary relations (directed graphs) show up naturally in data on the Web [1, 14], dataspaces [15], Linked Data [7], and RDF data [34]. Moreover, in restriction to directed graphs that are trees, the relational calculus is closely tied to the XML query language XPath, and the expressive power of XPath and various fragments has been intensively investigated [6, 18, 28, 29].

Here, working with general finite binary relation structures rather than trees, we consider, in addition to the five binary relation operations and four constant binary relations mentioned above, also four derived operations that are well known in the literature: set difference; projection; coprojection; and residuation. These derived operations can be expressed in terms of the other operations and constants, but can still be interesting on their own when considering fragments where some other operations or constants have been left out. We consider set difference because it is the standard domain-independent alternative to complementation in database query languages [2]. We consider projection and coprojection (existential and universal quantification) because they are standard logical operations, and have been shown important in the XPath setting [29], so it is natural to study their behaviour when generalizing from trees to general graphs. Finally, we consider residuation because it is similar to the standard relational division operation in databases, and corresponds to the set containment join [27]. Obviously, one could keep on inventing additional operations on binary relations and study their interdependencies, but our chosen set of operations is not too large and well motivated from the point of view of query languages.

Our goal now is to understand the relative importance of the various operations and the effect of their presence on indistinguishability. Thereto we consider all possible fragments of the calculus of relations that can be constructed as follows. The most basic fragment we consider has the empty and identity relations as constants, and the operations union, composition, and intersection. Then all other fragments arise by adding any choice of the remaining operations and constants. For each fragment, we provide a characterization of when two finite binary relation structures are indistinguishable by expressions in the fragment. Our approach follows the one outlined by Goranko and Otto [17]: we provide new notions of finite-round (degree-bounded) bisimulations, appropriate for fragments of the calculus of relations, and give characteristic expressions for them. For finite structures, such an approach immediately leads to a Hennessy–Milner-type theorem [8, 21].

One may ask why intersection is present in all the fragments we consider. Intersection is the most basic query language primitive [2]. Our results rely heavily on its presence, in the same way as the classical Hennessy–Milner theorem relies heavily on the presence of conjunction in the modal logic. Intersection is known not to be 'safe for bisimulation' [8, 38]. Nevertheless, the question whether or not all operations in a certain query language are safe for bisimulation appears to be quite different

from the problem of the present study: that of determining whether or not two given structures can be distinguished in a certain query language.

This is not to say, however, that indistinguishability of structures in fragments lacking intersection is uninteresting. But it changes the nature of the problem so drastically that we leave omitting intersection outside the scope of the present study. For example, consider the fragment consisting only of composition and nothing else. Then indistinguishability of finite structures amounts to the equivalence problem for finite automata, which is PSPACE-complete [4]. In contrast, for all fragments considered in this article, we will see that indistinguishability is decidable in polynomial time.

As mentioned above, bisimilarity-like characterizations of indistinguishability are common in modal logics. For non-modal logics, such as first-order logic, indistinguishability is typically captured by Ehrenfeucht-Fraïssé games. There also exist intermediate fragments of first-order logic, such as the guarded and the packed fragments [31], where indistinguishability can still be captured by appropriate notions of bisimulation. Note, however, that most of the fragments considered in this article are not subsumed by the packed fragment. For example, the expressions $(R/S) \circ T$, or $((R \circ S) - (R \circ T)) \circ S$, or $\bar{\pi}_1(R) \circ S \circ \bar{\pi}_2(T)$, are not expressible in the packed fragment.[1] Note als that projection such as as occurring in, e.g. $\pi_1(R/S)$, is in general not equivalent to guarded existential quantification. Only the 'positive' fragments, that include none of the residuals, set difference, complementation and coprojection, fall in the packed fragment.

To conclude this Introduction, we note another motivation to understand indistinguishability in database query language fragments, apart from the relevance to expressive power and the intrinsic foundational motivation. This is the new approach of *structural* indexing to database query processing, proposed by some of us and others [13, 32, 39], whereby a given query expression is processed by accessing blocks of data indistinguishable by the operations used in the given expression. By the results of our work, these blocks can be computed using similarity or bisimilarity checks.

Summary. The further contents of this article may be summarized as follows. In Section 2 we define the language fragments formally, and define the notion of indistinguishability. In Section 3 we discuss different ways how indistinguishability can be characterized; in particular we discuss the connection with multidimensional modal logics, and the 3-variable fragment of first-order logic. In Section 4, we define finite-round bisimulations appropriate for the fragments with the set difference operation. In Section 5, we define finite-round simulations appropriate for the fragments without set difference. In Section 6 we given Hennessy–Milner-type theorems for indistinguishability of finite structures. We conclude in Section 7.

2 Language fragments and indistinguishability

We assume an infinite universe of atomic data elements, denoted by U. A *binary relation* on U is a subset of $U^2 = U \times U$. We further fix an arbitrary finite set Λ of *relation names*, called the *vocabulary*. In the calculus of relations, a *structure* is a pair $\mathcal{G} = (V, (R^{\mathcal{G}})_{R \in \Lambda})$ where V is a subset of U and each $R^{\mathcal{G}}$ is a binary relation on V. The set V is called the set of nodes of $\mathcal{G}$; the vocabulary Λ can be thought of as a set of edge labels whereby $\mathcal{G}$ can be thought of as an edge-labeled directed graph. When V is finite, the structure is said to be a *finite* structure.

Expressions in the calculus of relations are built recursively from the relation names $R \in \Lambda$, and the constant symbols empty (0), all (1), diversity ($0'$), and identity ($1'$), using the following

---

[1]In the example, R/S stands for the left residual, $-$ stands for set difference and $\bar{\pi}$ stands for coprojection; these operators will be defined in the next Section.

standard and/or derived operations. The standard operations are union ($e_1 \cup e_2$), intersection ($e_1 \cap e_2$), complementation (e^c), composition ($e_1 \circ e_2$), and converse (e^{-1}); the derived operations we consider are set difference ($e_1 - e_2$), projection ($\pi_1 e$ or $\pi_2 e$), co-projection ($\bar{\pi}_1 e$ or $\bar{\pi}_2 e$), left residual (e_1 / e_2) and right residual ($e_1 \backslash e_2$).[2]

Semantically, on any structure $\mathcal{G}$ as above, an expression e defines a binary relation, denoted by $e(\mathcal{G})$. For convenience, we recall the semantics of the constants and the standard operations.

$$R(\mathcal{G}) = R^{\mathcal{G}};$$
$$0(\mathcal{G}) = \emptyset;$$
$$1(\mathcal{G}) = V^2;$$
$$0'(\mathcal{G}) = \{(s,t) \mid s,t \in V \ \& \ s \neq t\};$$
$$1'(\mathcal{G}) = \{(s,s) \mid s \in V\};$$
$$(e_1 \cup e_2)(\mathcal{G}) = e_1(\mathcal{G}) \cup e_2(\mathcal{G});$$
$$(e_1 \cap e_2)(\mathcal{G}) = e_1(\mathcal{G}) \cap e_2(\mathcal{G});$$
$$e^c(\mathcal{G}) = \{(s,t) \mid s,t \in V \ \& \ (s,t) \notin e(\mathcal{G})\};$$
$$(e_1 \circ e_2)(\mathcal{G}) = \{(s,t) \mid (\exists v)((s,v) \in e_1(\mathcal{G}) \ \& \ (v,t) \in e_2(\mathcal{G}))\};$$
$$(e^{-1})(\mathcal{G}) = \{(s,t) \mid (t,s) \in e(\mathcal{G})\}.$$

The semantics of the derived operations is as follows:

$$(e_1 - e_2)(\mathcal{G}) = \{(s,t) \mid (s,t) \in e_1(\mathcal{G}) \ \& \ (s,t) \notin e_2(\mathcal{G})\}$$
$$\pi_1(e)(\mathcal{G}) = \{(s,s) \mid (\exists t)(s,t) \in e(\mathcal{G})\}$$
$$\pi_2(e)(\mathcal{G}) = \{(s,s) \mid (\exists t)(t,s) \in e(\mathcal{G})\}$$
$$\bar{\pi}_1(e)(\mathcal{G}) = \{(s,s) \mid s \in V \ \& \ \neg(\exists t)(s,t) \in e(\mathcal{G})\}$$
$$\bar{\pi}_2(e)(\mathcal{G}) = \{(s,s) \mid s \in V \ \& \ \neg(\exists t)(t,s) \in e(\mathcal{G})\}$$
$$(e_1 / e_2)(\mathcal{G}) = \{(s,t) \mid (\forall v)((t,v) \in e_2(\mathcal{G}) \rightarrow (s,v) \in e_1(\mathcal{G}))\}$$
$$(e_1 \backslash e_2)(\mathcal{G}) = \{(s,t) \mid (\forall v)((v,s) \in e_1(\mathcal{G}) \rightarrow (v,t) \in e_2(\mathcal{G}))\}$$

EXAMPLE 2.1
Figure 1 shows a finite structure $\mathcal{G}$. The set of nodes equals {migraine, flu, sue, umi, saori, sriram, st jude's, inco}, and the vocabulary Λ equals {*knows, worksAt, patientOf, hasDisease*}.

- The doctors (i.e. persons having patients), can be retrieved from $\mathcal{G}$ by the expression

$$e_1 = \pi_2(patientOf)$$

resulting in $e_1(\mathcal{G}) = \{(\text{saori,saori})\}$.
- The people and the doctors they know can be obtained by the expression

$$e_2 = knows \circ e_1$$

resulting in $e_2(\mathcal{G}) = \{(\text{kotaro,saori})\}$.

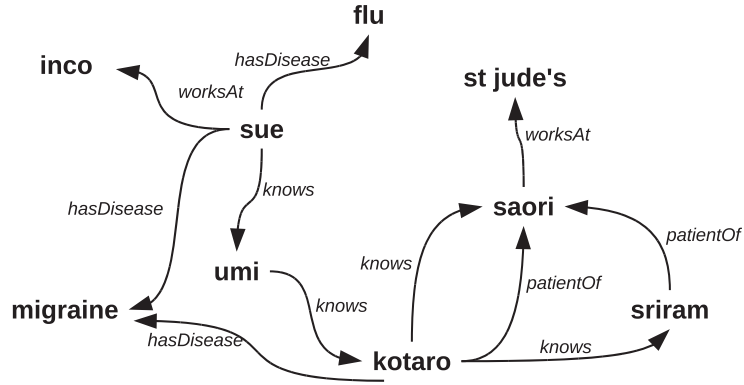[2]To distinguish between set difference and the right residual, we use the minus sign ($-$) for set difference.

FIGURE 1. Example structure from Example 2.1

- The doctors and the hospitals where they practice:

$$(e_1 \circ worksAt)(\mathcal{G}) = \{(\text{saori,st jude's})\}.$$

- Ill people without medical care:

$$(\pi_1(hasDisease) - \pi_1(patientOf))(\mathcal{G}) = \{(\text{sue,sue})\}.$$

- Healthy doctors:
$$(e_1 \cap \bar{\pi}_1(hasDisease))(\mathcal{G}) = \{(\text{saori,saori})\}.$$

- Finally, the doctors who know all the patients of some other doctor can be retrieved by the expression

$$e_1 \cap \pi_1((knows/patientOf) \cap 0'),$$

which on our example graph yields the empty relation, since the graph contains only one doctor.

2.1 Queries and equivalence

Expressions in the calculus of relations express queries. Formally, a *query* is a mapping Q from the set of all structures to the set of all binary relations on U, such that for each structure $\mathcal{G}$, if V is the node set of $\mathcal{G}$, then $Q(\mathcal{G})$ is a binary relation on V. Obviously the query Q expressed by an expression e is simply defined by $Q(\mathcal{G}) := e(\mathcal{G})$.

Two expressions e_1 and e_2 are now called *equivalent*, denoted by $e_1 \equiv e_2$, if they express the same query, i.e. if $e_1(\mathcal{G}) = e_2(\mathcal{G})$ for all possible structures $\mathcal{G}$. The following equivalences demonstrate that the derived operations are indeed derived, and also present some additional interdependencies among the constants and operations considered in this article:

$$1 \equiv 0^c \equiv 1' \cup 0' \equiv 0/0 \equiv 0 \backslash 0$$
$$0' \equiv 1'^c$$
$$e_1 - e_2 \equiv e_1 \cap e_2^c$$

$$e^c \equiv 1 - e$$
$$\pi_1(e) \equiv (e \circ e^{-1}) \cap 1' \equiv (e \circ 1) \cap 1' \equiv \bar{\pi}_1(\bar{\pi}_1(e))$$
$$\pi_2(e) \equiv (e^{-1} \circ e) \cap 1' \equiv (1 \circ e) \cap 1' \equiv \bar{\pi}_2(\bar{\pi}_2(e))$$
$$\bar{\pi}_i(e) \equiv 1' - \pi_i(e)$$
$$e_1 / e_2 \equiv (e_1^c \circ e_2^{-1})^c$$
$$e_1 \backslash e_2 \equiv (e_1^{-1} \circ e_2^c)^c$$

Of course the above list of equivalences is by no means complete. For example, another well-known equivalence is $(e_1 \circ e_2)^{-1} \equiv e_2^{-1} \circ e_1^{-1}$. It will be useful to have the following generalization of this equivalence:

PROPOSITION 2.2
Every expression e is equivalent to an expression e' that uses the same operations as e, and in which converse is only applied to relation names.

PROOF. We actually show the claim not for e but for e^{-1}; the claim for e then follows by applying it to each topmost application of converse within e. We go by induction on the structure of e. The case where e is a relation name is trivial. The constants are all equivalent to their converse. When e is of the form e_1^{-1}, we have $e^{-1} \equiv e_1$, which can be put in the required form by the induction hypothesis. When e is $e_1 \cup e_2$, we have $e^{-1} \equiv e_1^{-1} \cup e_2^{-1}$, and similarly when e is $e_1 \cap e_2$ or $e_1 - e_2$. When e is $e_1 \circ e_2$, we have $e^{-1} \equiv e_2^{-1} \circ e_1^{-1}$. When e is $\pi_1(e_1)$, we have $e^{-1} \equiv e$, and similarly for $\pi_2(e_1)$, $\bar{\pi}_1(e_1)$, and $\bar{\pi}_2(e_2)$. Finally, we have $(e_1 / e_2)^{-1} \equiv e_2^{-1} \backslash e_1^{-1}$ and $(e_1 \backslash e_2)^{-1} \equiv e_2^{-1} / e_1^{-1}$. ∎

2.2   Language fragments

We will consider various fragments of the calculus of relations. The most basic fragment we consider is denoted by $\mathcal{C}$: it has the constants 0 and $1'$ and the operators composition, union and intersection. All other fragments are defined by adding to $\mathcal{C}$ some additional constants and operators.

Formally, for any subset[3] $\mathcal{F}$ of $\{0', 1, ^{-1}, ^c, \pi, \bar{\pi}, -, /, \backslash\}$, we define the fragment $\mathcal{C}(\mathcal{F})$ consisting of the expressions built up from the relation names, 0, $1'$, and the constants from $\mathcal{F}$, using the operations composition, union, intersection and the operations from $\mathcal{F}$.

The fragment $\mathcal{C}(^{-1}, ^c)$ already amounts to the full calculus, since, by the equivalences listed above, all other operations can be derived in it. More precisely, from the listed equivalences, we can note the following:

- Any fragment containing complement also includes 1, $0'$, difference, projection and coprojection.
- Any fragment containing $0'$ also includes 1.
- Any fragment containing converse, $0'$, or coprojection, also includes projection.
- Any fragment containing projection and difference also contains coprojection.
- Any fragment containing both converse and complement also contains both residuals.

---

[3]In order to simplify our presentation somewhat, we only consider fragments containing both the first and second projection (π_1 and π_2) or none of them, and similarly for coprojection. This simplification is not essential to our results, however.

Accordingly, we say about a fragment $\mathcal{C}(\mathcal{F})$ that

- *1 is present in $\mathcal{C}(\mathcal{F})$ at degree 0* if $\mathcal{F}$ contains 1, $0'$, or complement;
- *1 is present in $\mathcal{C}(\mathcal{F})$ at degree 1* if 1 is not present at degree 0, and $\mathcal{F}$ contains the left or right residual.
- *1 is absent from $\mathcal{C}(\mathcal{F})$* if 1 is neither present at degree 0 nor at degree 1.

The idea behind these notions is that 1 is expressible as $0/0$ or as $0\backslash 0$. We will see later that these two expressions have degree one, as opposed to the expressions for 1 that have degree zero, viz., 1 itself, $0' \cup 1'$, or 0^c. The distinction between presence at degree 0, presence at degree 1 and absence of 1 in a fragment will manifest itself in Definition 2.4.

That 1 is expressible in terms of 0 and the residuals shows that our choice in this paper to include 0 by default in all fragments is not totally innocent, at least not in the presence of the residuals. We do not anticipate, however, that adapting the results of this article to a setting where 0 is absent should require new techniques. On a related note, it is also interesting to point out[4] that using 0 and the left residual, the first coprojection becomes expressible as $\bar{\pi}_1(e) = (0/e) \cap 1'$.

2.3 Degrees and paths

It is customary in finite model theory [9] to parameterize characterizations of indistinguishability by the quantifier rank of formulas. In our setting, the role of quantifier rank will be played by what we call the degree.

For an expression e, we define the *degree* $\deg(e)$ of e as follows. Every relation name and constant symbol has degree zero. Then,

$$\deg(e_1 \cup e_2) = \deg(e_1 \cap e_2) = \deg(e_1 - e_2) = \max(\deg(e_1), \deg(e_2));$$
$$\deg(e^c) = \deg(e^{-1}) = \deg(e);$$
$$\deg(e_1 \circ e_2) = \deg(e_1/e_2) = \deg(e_1 \backslash e_2) = 1 + \max(\deg(e_1), \deg(e_2));$$
$$\deg(\pi_1(e)) = \deg(\pi_2(e)) = \deg(\bar{\pi}_1(e)) = \deg(\bar{\pi}_2(e)) = 1 + \deg(e).$$

The degree of an expression is the maximum depth of nested applications of the composition, projection, co-projection and the left and right residual operation. Intuitively, the degree corresponds to the quantifier rank of e translated into first-order logic.

For a fragment $\mathcal{F}$ of the calculus of relations and a natural number k, we denote the set of expressions in $\mathcal{C}(\mathcal{F})$ of degree at most k by $\mathcal{C}(\mathcal{F})_k$.

Before introducing the crucial Definition 2.4, we need to agree on a natural way to view structures as directed, or as undirected graphs.

DEFINITION 2.3
Let $\mathcal{G} = (V, (R^{\mathcal{G}})_{R \in \Lambda})$ be a structure. Then *graph*$(\mathcal{G})$ is defined as the directed graph (V, E) where E equals the set of all pairs (x, y) in V^2 such that $(x, y) \in R^{\mathcal{G}}$ for some $R \in \Lambda$. Moreover, *ugraph*(G) is defined as the undirected version of *graph*(G), i.e. as the undirected graph (V, E') where E' is the set of all unordered pairs $\{x, y\}$ such that (x, y) or (y, x) belongs to E.

For any natural number k, we further define paths$_k(\mathcal{G})$ as the set of all pairs (x, y) in V^2 such that there is a path from x to y in *graph*$(\mathcal{G})$ of length at most 2^k. (The length of a path equals its number

---

[4]We thank the anonymous referee for this remark.

of edges, and we agree that there is a path of length 0 from x to x for any $x \in V$, i.e. (x,x) is always in $\text{paths}_k(\mathcal{G})$ for any k.) We define $\text{upaths}_k(\mathcal{G})$ similarly, but considering paths in the undirected graph $ugraph(\mathcal{G})$.

We now give:

DEFINITION 2.4 ($\mathcal{F}$-k-path)
Let $\mathcal{C}(\mathcal{F})$ be a fragment of the calculus of relations and let $\mathcal{G}$ be a structure. For any natural number k, we define $\text{paths}_k^{\mathcal{F}}(\mathcal{G})$ as follows.

- First, consider the case where 1 is absent in $\mathcal{C}(\mathcal{F})$. If $\mathcal{F}$ does not contain converse, then $\text{paths}_k^{\mathcal{F}}(\mathcal{G})$ is defined to be $\text{paths}_k(\mathcal{G})$; if $\mathcal{F}$ does contain converse, then $\text{paths}_k^{\mathcal{F}}(\mathcal{G})$ is defined to be $\text{upaths}_k(\mathcal{G})$.
- Next, assume 1 is present at degree 1. Then $\text{paths}_0^{\mathcal{F}}(\mathcal{G})$ is defined exactly as above for $k=0$, but $\text{paths}_k^{\mathcal{F}}(\mathcal{G})$ for $k>0$ is simply $1(\mathcal{G})$.
- Finally, 1 is present at degree 0, then $\text{paths}_k^{\mathcal{F}}(\mathcal{G})$ is again simply $1(\mathcal{G})$ for all k including zero.

We immediately note the following

LEMMA 2.5
For each $\mathcal{F}$ and each k, the query $\text{paths}_k^{\mathcal{F}}$ is expressible in $\mathcal{C}(\mathcal{F})_k$.

In the (easy) proof of the lemma, we will use the notion of *atomic expression* which will also be used throughout the article, so we define it separately here:

DEFINITION 2.6 (Atomic expressions aexp($\mathcal{F}$))
The *atomic* expressions are those from the finite set $\text{Atom} = (\{1',0'\} \cup \{R,R^{-1} \mid R \in \Lambda\})$. The set of atomic expressions belonging to $\mathcal{C}(\mathcal{F})$ is denoted by aexp($\mathcal{F}$).

Note that by Proposition 2.2, we can indeed assume for any expression that the leaves of its syntax tree are labeled by atomic expressions.

Now to the

PROOF OF LEMMA 2.5   When 1 is absent, the lemma follows from the following equivalences:

$$\text{paths}_0^{\mathcal{F}} \equiv \bigcup_{e \in \text{aexp}(\mathcal{F})} e$$
$$\text{paths}_{k+1}^{\mathcal{F}} \equiv \text{paths}_k^{\mathcal{F}} \cup (\text{paths}_k^{\mathcal{F}} \circ \text{paths}_k^{\mathcal{F}}).$$

When 1 is present at degree 1, $\text{paths}_0^{\mathcal{F}}$ is expressed as above and $\text{paths}_k^{\mathcal{F}}$ being equivalent to 1, can be expressed by $0/0$ or $0\backslash 0$. When 1 is present at degree 0, paths_k^F is again equivalent to 1 and expressible by 1 itself, $0' \cup 1'$, or 0^c. ∎

The next proposition shows the relevance of $\text{paths}_k^{\mathcal{F}}$.

PROPOSITION 2.7
Let $\mathcal{F}$ be a fragment of the calculus of relations and let $\mathcal{G}$ be a structure. For any natural number k and any expression $e \in \mathcal{C}(\mathcal{F})_k$, we have $e(\mathcal{G}) \subseteq \text{paths}_k^{\mathcal{F}}(\mathcal{G})$.

PROOF.   By the definition of $\text{paths}_k^{\mathcal{F}}$, the statement of the lemma is trivial when 1 is present in $\mathcal{C}(\mathcal{F})$ at degree 0. Also when 1 is present at degree 1, the lemma is trivial, except for the case $k=0$, but

then $\text{paths}_0^{\mathcal{F}} = \text{paths}_0^{\mathcal{F}'}$ where $\mathcal{F}'$ is obtained from $\mathcal{F}$ by removing the residuals. Note that 1 is absent in $\mathcal{F}'$ and that $\mathcal{C}(\mathcal{F}')_0 = \mathcal{C}(\mathcal{F})_0$.

Hence, it suffices to prove the lemma for the case that 1 is absent from $\mathcal{C}(\mathcal{F})$. This means that $\mathcal{F}$ does *not* contain 1, $0'$, complement, and the residuals.

We now proceed by structural induction on e. If $e \in \text{aexp}(\mathcal{F})$, then $e(\mathcal{G}) \subseteq \text{paths}_0^{\mathcal{F}}(\mathcal{G})$ by definition of $\text{paths}_0^{\mathcal{F}}(\mathcal{G})$.

If e is $e_1 \cup e_2$, $e_1 \cap e_2$, or $e_1 - e_2$, the result follows immediately from the induction hypothesis.

If e is $\pi_1(e_1)$, $\pi_2(e_1)$, $\bar{\pi}_1(e_1)$, or $\bar{\pi}_2(e_1)$, the result is immediate because $\pi_1(e_1)(\mathcal{G}) \subseteq 1'(\mathcal{G}) \subseteq \text{paths}_k^{\mathcal{F}}(\mathcal{G})$. (Similarly for $\pi_2(e_1)$, $\bar{\pi}_1(e_1)$, and $\bar{\pi}_2(e_1)$.)

Finally, if e is $e_1 \circ e_2$, let $k_1 = \deg(e_1)$, $k_2 = \deg(e_2)$, and $\ell = \max(k_1, k_2)$. Note that $k = \ell + 1$. Now assume $(a, b) \in e_1 \circ e_2(\mathcal{G})$. Then, for some $c \in V$, we have $(a, c) \in e_1(\mathcal{G})$ and $(c, b) \in e_2(\mathcal{G})$. By induction, we have $(a, c) \in \text{paths}_{k_1}^{\mathcal{F}}(\mathcal{G}) \subseteq \text{paths}_\ell^{\mathcal{F}}$ and $(c, b) \in \text{paths}_{k_2}^{\mathcal{F}}(\mathcal{G}) \subseteq \text{paths}_\ell^{\mathcal{F}}$, whence $(a, b) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G})$ as desired. ∎

2.4 Indistinguishability

A *marked structure* $\overline{\mathcal{G}}$ is a pair $(\mathcal{G}, a, b)$ where $\mathcal{G}$ is a relational structure, and (a, b) is an ordered pair of nodes from $\mathcal{G}$. Let $\mathcal{C}(\mathcal{F})$ be a fragment of the calculus of relations, and let k be a natural number. The $\mathcal{C}(\mathcal{F})_k$-*type* of $\overline{\mathcal{G}}$, denoted by $\text{tp}_k^{\mathcal{F}}(\overline{\mathcal{G}})$, is defined as the set of all expressions $e \in \mathcal{C}(\mathcal{F})_k$ such that $(a, b) \in e(\mathcal{G})$. For two marked structures $\overline{\mathcal{G}}_1 = (\mathcal{G}_1, a_1, b_1)$ and $\overline{\mathcal{G}}_2 = (\mathcal{G}_2, a_2, b_2)$, we write $\overline{\mathcal{G}}_1 \Rightarrow_k^{\mathcal{F}} \overline{\mathcal{G}}_2$ if $\text{tp}_k^{\mathcal{F}}(\mathcal{G}_1, a_1, b_1) \subseteq \text{tp}_k^{\mathcal{F}}(\mathcal{G}_2, a_2, b_2)$, i.e. for every expression $e \in \mathcal{C}(\mathcal{F})_k$ such that $(a_1, b_1) \in e(\mathcal{G}_1)$, also $(a_2, b_2) \in e(\mathcal{G}_2)$. We then say that $\overline{\mathcal{G}}_2$ is *one-sided indistinguishable* from $\overline{\mathcal{G}}_1$ in $\mathcal{C}(\mathcal{F})_k$. When both $\overline{\mathcal{G}}_1 \Rightarrow_k^{\mathcal{F}} \overline{\mathcal{G}}_2$ and $\overline{\mathcal{G}}_2 \Rightarrow_k^{\mathcal{F}} \overline{\mathcal{G}}_1$, we say that $\overline{\mathcal{G}}_1$ and $\overline{\mathcal{G}}_2$ are *indistinguishable in* $\mathcal{C}(\mathcal{F})_k$ and denote this by $\overline{\mathcal{G}}_1 \equiv_k^{\mathcal{F}} \overline{\mathcal{G}}_2$.

Recalling Definition 2.6, we also define the *atomic* $\mathcal{F}$-*type* of $\overline{\mathcal{G}}$, denoted by $\text{atp}^{\mathcal{F}}(\overline{\mathcal{G}})$, as $\text{Atom} \cap \text{tp}_0^{\mathcal{F}}(\overline{\mathcal{G}})$. Note that $\text{atp}^{\mathcal{F}}(\overline{\mathcal{G}})$ is always a subset of $\text{aexp}(\mathcal{F})$.

Since indistinguishability is the same as one-sided indistinguishability in both directions, it is more general to look for a characterization of one-sided indistinguishability, and that is what we will do. On the other hand, when the fragment contains complement or difference, one-sided indistinguishability actually coincides with indistinguishability, except in a trivial case:

PROPOSITION 2.8
Let $\mathcal{C}(\mathcal{F})$ be a fragment of the calculus of relations so that $\mathcal{F}$ contains complement or difference. Let $\overline{\mathcal{G}}_1 = (\mathcal{G}_1, a_1, b_1)$ and $\overline{\mathcal{G}}_2 = (\mathcal{G}_2, a_2, b_2)$ be two marked structures, and let k be a natural number. Then

$$\overline{\mathcal{G}}_1 \Rightarrow_k^{\mathcal{F}} \overline{\mathcal{G}}_2 \quad \Leftrightarrow \quad \overline{\mathcal{G}}_1 \equiv_k^{\mathcal{F}} \overline{\mathcal{G}}_2$$

whenever $(a_1, b_1) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G}_1)$. When $(a_1, b_1) \notin \text{paths}_k^{\mathcal{F}}(\mathcal{G}_1)$, the one-sided indistinguishability $\overline{\mathcal{G}}_1 \Rightarrow_k^{\mathcal{F}} \overline{\mathcal{G}}_2$ holds trivially, and $\overline{\mathcal{G}}_1 \equiv_k^{\mathcal{F}} \overline{\mathcal{G}}_2$ holds if and only if $(a_2, b_2) \notin \text{paths}_k^{\mathcal{F}}(\mathcal{G}_2)$.

PROOF. We first show that, when $(a_1, b_1) \in \text{paths}_k^{\mathcal{F}}(\mathcal{G}_1)$, then $\overline{\mathcal{G}}_1 \Rightarrow_k^{\mathcal{F}} \overline{\mathcal{G}}_2$ implies $\overline{\mathcal{G}}_2 \Rightarrow_k^{\mathcal{F}} \overline{\mathcal{G}}_1$. Thereto, let $e \in \mathcal{C}(\mathcal{F})_k$ such that $(a_2, b_2) \in e(\mathcal{G}_2)$; we must show that (a_1, b_1) belongs to $e(\mathcal{G}_1)$. Assume, for the sake of contradition, that it does not. Then $(a_1, b_1) \in (\text{paths}_k^{\mathcal{F}} - e)(\mathcal{G}_1)$. Note that when $\mathcal{F}$ would not contain difference, it would contain complement, and then the expression $\text{paths}_k^{\mathcal{F}} - e$ can be equivalently written in $\mathcal{C}(\mathcal{F})$ as $e^c(\mathcal{G}_1)$. In either case, the expression has degree k, so, since $\overline{\mathcal{G}}_1 \Rightarrow_k^{\mathcal{F}} \overline{\mathcal{G}}_2$,

we have $(a_2, b_2) \in (\text{paths}_k^{\mathcal{F}} - e)(\mathcal{G}_2)$. In particular it follows $(a_2, b_2) \notin e(\mathcal{G}_2)$ which yields the desired contradiction.

When $(a_1, b_1) \notin \text{paths}_k^{\mathcal{F}}(\mathcal{G}_1)$, then $\overline{\mathcal{G}}_1 \Rrightarrow_k^{\mathcal{F}} \overline{\mathcal{G}}_2$ is indeed voidly satisfied, since in that case, by Proposition 2.7, the $\text{tp}_k^{\mathcal{F}}(\overline{\mathcal{G}}_1)$ is empty. Moreover then, clearly $\overline{\mathcal{G}}_1 \equiv_k^{\mathcal{F}} \overline{\mathcal{G}}_2$ iff (a_2, b_2) does not belong to $e(\mathcal{G}_1)$, for any $e \in \mathcal{C}(\mathcal{F})_k$, either. We now note that the latter holds iff $(a_2, b_2) \notin \text{paths}_k^{\mathcal{F}}(\mathcal{G}_2)$. Indeed, the only-if is clear since $\text{paths}_k^{\mathcal{F}}$ belongs to $\mathcal{C}(\mathcal{F})_k$; the if-direction is again given by Proposition 2.7.   ∎

Similarly to indistinguishability in $\mathcal{C}(\mathcal{F})_k$, i.e. for a fixed degree k, we are also interested in indistinguishability in an entire fragment $\mathcal{C}(\mathcal{F})$. Thus define the $\mathcal{F}$-*type* of a marked structure $\overline{\mathcal{G}}$, denoted by $\text{tp}^{\mathcal{F}}(\overline{\mathcal{G}})$, as the set of all expressions e from $\mathcal{C}(\mathcal{F})$ such that $(a, b) \in e(\mathcal{G})$. Using this notion of type we can now define the indistinguishability notions $\overline{\mathcal{G}}_1 \Rrightarrow^{\mathcal{F}} \overline{\mathcal{G}}_2$ and $\overline{\mathcal{G}}_1 \equiv^{\mathcal{F}} \overline{\mathcal{G}}_2$ similarly to the fixed-degree case.

3   Approaches to bisimilarity

Before discussing indistinguishability for fragments of the calculus of relations, let us first look at the full calculus $\mathcal{C}(\mathcal{F}_{\text{full}})$, with $\mathcal{F}_{\text{full}}$ consisting of complement and converse. Tarski and Givant showed that the calculus has equal expressive power as $\text{FO}^3(2)$: the formulas with two free variables in the three-variable fragment FO^3 of first-order logic [37]. For FO^3, we have the three-pebble Ehrenfeucht-Fraïssé game as a characterization [9, 24]. Marx and Venema, however, showed that $\text{FO}^3(2)$ has also the same expressive power as arrow logic [30], a branch of multidimensional modal logic devised to provide a formalization for simple reasoning about objects that are thought of as arrows. By this correspondence, bisimulations in terms of back-and-forth conditions that are well known from modal logic can be used to characterize fragments of FO^3, and, hence, of the calculus of relations.

Concretely, the language of arrow logic is a modal language with the dyadic operator $\circ$, the monadic operator $\otimes$, and the modal constant id. Formulas in arrow logic are built up from a set of propositional variables and the modal constant id, using the operators $\circ$ and $\otimes$, and the boolean connectives $\wedge$, $\vee$, $\neg$. Using propositional variables to denote edge labels; by interpreting the modal constant id as being true for pairs (a, a) of identical nodes; by interpreting the monadic operator $\otimes$ as being true for pairs $((b, a), (a, b))$ of 'arrows' such that the first arrow is the converse of the second arrow; and finally, by interpreting the dyadic operator $\circ$ as being true for triples $((a, b), (a, c), (c, b))$ of arrows such that the first one is obtained by composing the second and the third arrow, we can apply the characterization theorem of modal logic to immediately obtain a characterization for the full calculus of relations. We will next make this more precise.

The notion of bisimulation for multidimensional modal logic, specialized to the above interpretation of arrow logic, becomes the following:

DEFINITION 3.1
Let $\mathcal{G}_1$ and $\mathcal{G}_2$ be two structures with node sets V_1 and V_2, respectively. A non-empty relation $Z \subseteq V_1^2 \times V_2^2$ is an *arrow-logic bisimulation* between $\mathcal{G}_1$ and $\mathcal{G}_2$ if it satisfies the following conditions:[5]

Atoms if (a_1, b_1, a_2, b_2) is in Z, then $(a_1, b_1) \in R(\mathcal{G}_1)$ if and only if $(a_2, b_2) \in R(\mathcal{G}_2)$, for all $R \in \Lambda$;
Forth if $(a_1, b_1, a_2, b_2) \in Z$, then

---

[5]The attentive reader will notice that the converse-forth condition and the converse-back condition are identical. This is a consequence of the symmetry of the converse operator. We could have simplified the definition by removing one of the identical conditions, but preferred to stay in line with the general format of bisimulation conditions for multidimensional modal logic.

> **composition**($\circ$) for each $c_1 \in V_1$ there exist $c_2 \in V_2$ such that both (a_1, c_1, a_2, c_2) and (c_1, b_1, c_2, b_2) are in Z;
> **identity**(id) if $a_1 = b_1$ then $a_2 = b_2$;
> **converse**($\otimes$) $(b_1, a_1, b_2, a_2) \in Z$;

Back if (a_1, b_1, a_2, b_2) is in Z, then

> **composition**($\circ$) for each $c_2 \in V_2$ there exist $c_1 \in V_1$ such that both (a_1, c_1, a_2, c_2) and (c_1, b_1, c_2, b_2) are in Z;
> **identity**(id) if $a_2 = b_2$ then $a_1 = b_1$;
> **converse**($\otimes$) $(b_1, a_1, b_2, a_2) \in Z$.

A marked structure $\overline{\mathcal{G}}_1 = (\mathcal{G}_1, a_1, b_1)$ is said to be *arrow-logic bisimilar* to a marked structure $\overline{\mathcal{G}}_2 = (\mathcal{G}_2, a_2, b_2)$ if there is an arrow-logic bisimulation Z between $\mathcal{G}_1$ and $\mathcal{G}_2$ containing (a_1, b_1, a_2, b_2).

The following characterization is now given by the Hennessy–Milner theorem [8, Theorem 2.24]:

PROPOSITION 3.2
Let $\overline{\mathcal{G}}_1 = (\mathcal{G}_1, a_1, b_1)$ and $\overline{\mathcal{G}}_2 = (\mathcal{G}_2, a_2, b_2)$ be *finite* marked structures. Then

$$\overline{\mathcal{G}}_1 \equiv^{\mathcal{F}_{\text{full}}} \overline{\mathcal{G}}_2 \quad \Leftrightarrow \quad \overline{\mathcal{G}}_1 \text{ is arrow-logic bisimilar to } \overline{\mathcal{G}}_2.$$

In the field of databases [2, Chapter 5], it is good practice to employ 'safe' query languages, meaning that 1 is not expressible, i.e. queries are domain-independent. This would mean replacing the complementation operator by the difference operator, and removing the diversity relation, leading to the fragment $\mathcal{C}(\mathcal{F}_{\text{safe}})$, with $\mathcal{F}_{\text{safe}}$ consisting of difference and converse. Furthermore, in database theory much attention is being paid to 'positive' query languages, i.e. without the difference operator. Since it is still important to understand the distinction between safe and unsafe query languages, one might add the diversity relation back in, which would lead one to fragments such as $\mathcal{C}(^{-1}, 0')$. (Note that adding diversity to $\mathcal{F}_{\text{safe}}$ would bring us back to the full calculus, since 1 is expressible as $1' \cup 0'$ and then complement e^c as $1 - e$.) Also, one may be interested in understanding the power of following relations backwards, and study fragments where converse is removed, such as $\mathcal{C}(-)$ or $\mathcal{C}(0')$. Then again one may add derived operations (projection, coprojection, residuals) that become primitive in specific fragments, leading to new fragments such as $C(-, /, \backslash)$ or $\mathcal{C}(^{-1}, \bar{\pi})$.

For some fragments discussed above, the Hennessy–Milner theorem adapts easily. As a case in point, consider the positive fragment with diversity, $\mathcal{C}(^{-1}, 0')$. To account for the absence of complementation, it suffices in the definition of bisimulation to remove the Back condition, thus obtaining a kind of *simulation* rather than bisimulation. To account for the diversity relation, it suffices to add it as a nullary modality by adding the following part to the forth-condition:

diversity(di) if $a_1 \neq b_1$, then $a_2 \neq b_2$.

We can then analogously show that $(\mathcal{G}_1, a_1, b_1) \Rightarrow^{-1, 0'} (\mathcal{G}_2, a_2, b_2)$ if and only if there exists such a diversity-simulation from $\mathcal{G}_1$ to $\mathcal{G}_2$ containing (a_1, b_1, a_2, b_2).

Many other fragments, however, require much less obvious adaptations to the notion of bisimulation for arrow logic. For instance, the coprojection and residual operations cannot simply be considered to be extra modalities in arrow logic. Another difficulty arises when we remove the converse operator or the diversity relation. Expressions in such fragments always return paths in the graph formed by the atomic steps (Proposition 2.7). It does not suffice now to remove the
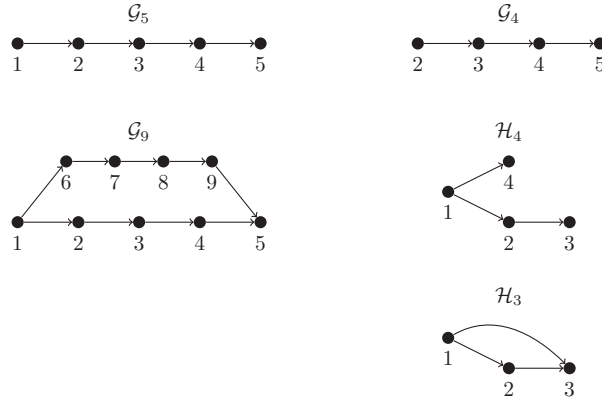
FIGURE 2. Five structures shown as directed graphs

converse-forth or the diversity-forth parts in the definition of bisimulation; we also need to adapt the composition-forth part.

In the remainder of this article, we will show how appropriate notions of simulation and bisimulation can be defined for all fragments $\mathcal{C}(\mathcal{F})$ of the calculus of relations considered in this article.

3.1   Some examples

We conclude the present section with a few examples of (in)distinguishability. Thereto we introduce five example structures in the vocabulary of a single relation name R, shown in Figure 2.

Let us begin by comparing the marked structures $(\mathcal{G}_5, 2, 5)$ and $(\mathcal{G}_4, 2, 5)$. They are distinguishable in the full calculus, for example, by the expression[6] $e_1 = \pi_2(R) \circ R^4$ which belongs to $\mathrm{tp}^{\mathcal{F}_{\text{full}}}(\mathcal{G}_5, 2, 5)$ but not to $\mathrm{tp}^{\mathcal{F}_{\text{full}}}(\mathcal{G}_4, 2, 5)$. In contrast, we have $(\mathcal{G}_5, 2, 5) \equiv^{-} (\mathcal{G}_4, 2, 5)$, i.e. the two marked structures are indistinguishable in $\mathcal{C}(-)$. The intuition, following Proposition 2.7, is that expressions in this fragment that return a pair (a, b) on a structure are confined to the part of the structure formed by all directed paths from a to b; on $\mathcal{G}_5$ and $\mathcal{G}_4$, the parts between 2 and 5 are identical. One way to make such an indistinguishability claim formal will be our main result, which includes a notion of bisimulation appropriate for the fragment $\mathcal{C}(-)$.

Consider now a positive fragment such as $\mathcal{C}(\pi)$. We obviously have $(\mathcal{G}_5, 2, 5) \not\Rightarrow^{\pi} (\mathcal{G}_4, 2, 5)$ since the expression e_1 above belongs to $\mathcal{C}(\pi)$. Nevertheless, in the other direction we do have $(\mathcal{G}_4, 2, 5) \Rightarrow^{\pi} (\mathcal{G}_5, 2, 5)$, as again will follow from our main result. On the other hand, in the fragment $\mathcal{C}(\bar{\pi})$ we have $(\mathcal{G}_4, 2, 5) \not\Rightarrow^{\bar{\pi}} (\mathcal{G}_5, 2, 5)$ by the expression $\bar{\pi}_2(R) \circ R^4$.

We can illustrate degree-bounded indistinguishability by $(\mathcal{G}_5, 1, 5) \equiv_2^{\mathcal{F}_{\text{safe}}} (\mathcal{G}_9, 1, 5)$. Indeed, degree-two expressions in $\mathcal{C}(\mathcal{F}_{\text{safe}})$ are limited to paths of length at most four, and both structures are identical inasfar as such paths between 1 and 5 are concerned. Of course, the two marked structures are distinguishable in degree three; by the expression[7] $R^4 \cap R^5$ we already have distinguishability $(\mathcal{G}_9, 1, 5) \not\Rightarrow_3 (\mathcal{G}_5, 1, 5)$ in the most basic fragment $\mathcal{C}$.

---

[6]Recall that projection is expressible in the full calculus.

[7]R^5 is expressible in degree three as $((R \circ R) \circ R) \circ (R \circ R)$.

Finally let us compare the marked structures $(\mathcal{H}_4,1,1)$ and $(\mathcal{H}_3,1,1)$. Clearly, in $\mathcal{C}(\pi)$, we have $(\mathcal{H}_3,1,1)\not\Rrightarrow_1^{\pi}(\mathcal{H}_4,1,1)$ as witnessed by the expression $\pi_1(R^2\cap R)$. In the other direction, however, we even have $(\mathcal{H}_4,1,1)\Rrightarrow^{-1}(\mathcal{H}_3,1,1)$. On the other hand, adding diversity, we have $(\mathcal{H}_4,1,1)\not\Rrightarrow^{-1,0'}(\mathcal{H}_3,1,1)$ by the expression $R\circ((R\circ 0')\cap 0')\circ R^{-1}$.

4 Bisimilarity and indistinguishability

In this and the following section, we proceed as announced in the preceding sections and define, for any fragment $\mathcal{C}(\mathcal{F})$ and any natural number k, an appropriate notion of bisimulation or simulation between structures. We will then show the adequacy of the proposed notions in capturing indistinguishability in $\mathcal{C}(\mathcal{F})_k$.

In the present section, we deal with fragments containing complement or difference.

4.1 General definition of $(\mathcal{F},k)$-bisimulation

Let $\mathcal{G}_1$ and $\mathcal{G}_2$ be two structures with node sets V_1 and V_2 respectively. Let (a_1,b_1,a_2,b_2) be an arbitrary element of $V_1^2\times V_2^2$, and let Z be an arbitrary subset of the same set $V_1^2\times V_2^2$.

For any given calculus fragment $\mathcal{C}(\mathcal{F})$ that contains complement or difference, we define the following suite of conditions. The conditions are relative to $\mathcal{F}$ in that they refer to $\mathrm{atp}^{\mathcal{F}}$ and $\mathrm{paths}^{\mathcal{F}}$. All the conditions are also clearly relative to $\mathcal{G}_1$ and $\mathcal{G}_2$.

Atoms Forth We say that (a_1,b_1,a_2,b_2) has the *Atoms Forth property* if

$$\mathrm{atp}^{\mathcal{F}}(\mathcal{G}_1,a_1,b_1)\subseteq\mathrm{atp}^{\mathcal{F}}(\mathcal{G}_2,a_2,b_2).$$

Atoms Back We say that (a_1,b_1,a_2,b_2) has the *Atoms Back property* if

$$\mathrm{atp}^{\mathcal{F}}(\mathcal{G}_1,a_1,b_1)\supseteq\mathrm{atp}^{\mathcal{F}}(\mathcal{G}_2,a_2,b_2).$$

Furthermore, let $i>0$ be an arbitrary natural number.

Composition Forth We say that (a_1,b_1,a_2,b_2) has the *Composition Forth property at degree i with respect to Z* if for every c_1 in V_1 with (a_1,c_1) and (c_1,b_1) in $\mathrm{paths}^{\mathcal{F}}_{i-1}(\mathcal{G}_1)$, there exists c_2 in V_2 such that both $(a_1,c_1,a_2,c_2)\in Z$ and $(c_1,b_1,c_2,b_2)\in Z$.
Composition Back We say that (a_1,b_1,a_2,b_2) has the *Composition Back property at degree i with respect to Z* if for every c_2 in V_2 with (a_2,c_2) and (c_2,b_2) in $\mathrm{paths}^{\mathcal{F}}_{i-1}(\mathcal{G}_2)$, there exists c_1 in V_1 such that both $(a_1,c_1,a_2,c_2)\in Z$ and $(c_1,b_1,c_2,b_2)\in Z$.
Projection Forth We say that (a_1,b_1,a_2,b_2) has the *Projection Forth property at degree i with respect to Z* if either $a_1\neq b_1$, or $a_1=b_1$ and $a_2=b_2$ and for every c_1 in V_1 with (a_1,c_1) in $\mathrm{paths}^{\mathcal{F}}_{i-1}(\mathcal{G}_1)$, there exists c_2 in V_2 such that $(a_1,c_1,a_2,c_2)\in Z$. Moreover, if $a_1=b_1$ then also for every c_1 in V_1 with (c_1,a_1) in $\mathrm{paths}^{\mathcal{F}}_{i-1}(\mathcal{G}_1)$, there must exist c_2 in V_2 such that $(c_1,a_1,c_2,a_2)\in Z$.
Projection Back We say that (a_1,b_1,a_2,b_2) has the *Projection Back property at degree i with respect to Z* if either $a_2\neq b_2$, or $a_2=b_2$ and $a_1=b_1$ and for every c_2 in V_2 with (a_2,c_2) in $\mathrm{paths}^{\mathcal{F}}_{i-1}(\mathcal{G}_2)$, there exists c_1 in V_1 such that $(a_1,c_1,a_2,c_2)\in Z$. Moreover, if $a_2=b_2$, then also for every c_2 in V_2 with (c_2,a_2) in $\mathrm{paths}^{\mathcal{F}}_{i-1}(\mathcal{G}_2)$, there must exist c_1 in V_1 such that $(c_1,a_1,c_2,a_2)\in Z$.

Left Residual Forth We say that (a_1, b_1, a_2, b_2) has the *Left Residual Forth property at degree i with respect to Z* if for every c_2 in V_2 with (b_2, c_2) in $\mathrm{paths}_{i-1}^{\mathcal{F}}(\mathcal{G}_2)$, there exists c_1 in V_1 such that both $(b_1, c_1, b_2, c_2) \in Z$ and either $(a_1, c_1) \notin \mathrm{paths}_{i-1}^{\mathcal{F}}(\mathcal{G}_1)$ or $(a_1, c_1, a_2, c_2) \in Z$.

Left Residual Back We say that (a_1, b_1, a_2, b_2) has the *Left Residual Back property at degree i with respect to Z* if for every c_1 in V_1 with (b_1, c_1) in $\mathrm{paths}_{i-1}^{\mathcal{F}}(\mathcal{G}_1)$, there exists c_2 in V_2 such that both $(b_1, c_1, b_2, c_2) \in Z$ and either $(a_2, c_2) \notin \mathrm{paths}_{i-1}^{\mathcal{F}}(\mathcal{G}_2)$ or $(a_1, c_1, a_2, c_2) \in Z$.

Right Residual Forth We say that (a_1, b_1, a_2, b_2) has the *Right Residual Forth property at degree i with respect to Z* if for every c_2 in V_2 with (c_2, a_2) in $\mathrm{paths}_{i-1}^{\mathcal{F}}(\mathcal{G}_2)$, there exists c_1 in V_1 such that both $(c_1, a_1, c_2, a_2) \in Z$ and either $(c_1, b_1) \notin \mathrm{paths}_{i-1}^{\mathcal{F}}(\mathcal{G}_1)$ or $(c_1, b_1, c_2, b_2) \in Z$.

Right Residual Back We say that (a_1, b_1, a_2, b_2) has the *Right Residual Back property at degree i with respect to Z* if for every c_1 in V_1 with (c_1, a_1) in $\mathrm{paths}_{i-1}^{\mathcal{F}}(\mathcal{G}_1)$, there exists c_2 in V_2 such that both $(c_1, a_1, c_2, a_2) \in Z$ and either $(c_2, b_2) \notin \mathrm{paths}_{i-1}^{\mathcal{F}}(\mathcal{G}_2)$ or $(c_1, b_1, c_2, b_2) \in Z$.

Now let k be a natural number and let $\bar{Z} = (Z_0, Z_1, \ldots, Z_k)$ be a decreasing sequence of relations with $Z_0 \subseteq V_1^2 \times V_2^2$, decreasing in the sense that $Z_i \subseteq Z_{i-1}$ for each $i \in \{1, \ldots, k\}$.

We lift the above conditions to apply to such sequences $\bar{Z}$ as follows:

- We say that $\bar{Z}$ has the Atoms Forth property if every element of Z_0 has this property, and similarly for the Atoms Back property.
- For any of the other properties (from Composition Forth to Right Residual Back), we say that $\bar{Z}$ has a certain property if for every $i \in \{1, \ldots, k\}$, every element of Z_i has that property at degree i with respect to Z_{i-1}.

We are finally ready for our main definition. The following definition is the most natural and easy to state, but we will see later in Corollary 4.10 that in some cases, some of the conditions are actually redundant.

DEFINITION 4.1
We call $\bar{Z}$ an $(\mathcal{F}, k)$-bisimulation from $\mathcal{G}_1$ to $\mathcal{G}_2$ if $\bar{Z}$ has

- the Atoms Forth and Back properties;
- the Composition Forth and Back properties;
- the Projection Forth and Back properties when $\mathcal{F}$ contains projection; and
- the Left (Right) Residual Forth and Back properties when $\mathcal{F}$ contains left (right) residual.

Given two marked structures $\overline{\mathcal{G}}_1 = (\mathcal{G}_1, a_1, b_1)$ and $\overline{\mathcal{G}}_2 = (\mathcal{G}_2, a_2, b_2)$, when there exists an $(\mathcal{F}, k)$-bisimulation $\bar{Z}$ from $\mathcal{G}_1$ to $\mathcal{G}_2$ such that $(a_1, b_1, a_2, b_2) \in Z_k$, we say that $\overline{\mathcal{G}}_1$ and $\overline{\mathcal{G}}_2$ are $(\mathcal{F}, k)$-*bisimilar*, and denote this by $\overline{\mathcal{G}}_1 \simeq_k^{\mathcal{F}} \overline{\mathcal{G}}_2$.

Note that in any bisimulation $\bar{Z}$, since both the Forth and Back versions of the Atoms property must be satisfied, and since identity is always present in our fragments, for each $(a_1, b_1, a_2, b_2) \in Z_0$ we have $a_1 = b_1$ iff $a_2 = b_2$. From this it follows that $a_1 \neq b_1$ iff $a_2 \neq b_2$, so it does not matter whether or not $0'$ belongs to $\mathcal{F}$. Hence, there really are only two variants of the Atoms properties, depending on whether or not $\mathcal{F}$ contains converse. Both variants stipulate $a_1 = b_1$ iff $a_2 = b_2$ as just seen. When $\mathcal{F}$ does not contain converse, the Atoms conditions stipulate furthermore that exactly the same relations R from Λ must hold for (a_1, b_1) and (a_2, b_2); moreover, when $\mathcal{F}$ does contain converse, also the converse relations that hold must be the same.

One may wonder why there are no conditions corresponding to the coprojection operation. The reason is that in fragments containing difference, coprojection plays no additional role beyond

that of projection, because $\bar{\pi}(e) \equiv 1' - \pi(e)$. Formally, we will see in the Invariance Lemma that coprojections are also preserved by bisimulations.

4.2 Adequacy theorem

We establish:

THEOREM 4.2 (Adequacy Theorem)
For any fragment $\mathcal{C}(\mathcal{F})$ where $\mathcal{F}$ contains complement or difference, we have $\overline{\mathcal{G}}_1 \simeq_k^{\mathcal{F}} \overline{\mathcal{G}}_2$ if and only if $\overline{\mathcal{G}}_1 \equiv_k^{\mathcal{F}} \overline{\mathcal{G}}_2$.

We will prove the only-if direction in the Invariance Lemma, where we will need the following immediate property:

LEMMA 4.3
Let $k > 0$ and let $(Z_0, Z_1, \ldots, Z_k)$ be an $(\mathcal{F}, k)$-bisimulation from $\mathcal{G}_1$ to $\mathcal{G}_2$. Then $(Z_0, \ldots, Z_{k-1})$ is an $(\mathcal{F}, k-1)$-bisimulation from $\mathcal{G}_1$ to $\mathcal{G}_2$.

We now give:

LEMMA 4.4 (Invariance)
If $\overline{\mathcal{G}}_1 \simeq_k^{\mathcal{F}} \overline{\mathcal{G}}_2$ then $\overline{\mathcal{G}}_1 \equiv_k^{\mathcal{F}} \overline{\mathcal{G}}_2$.

PROOF. Let e be an expression in $\mathcal{C}(\mathcal{F})_k$; we may assume by Proposition 2.2 that converse is only applied to relation names. We prove by induction on the structure of e that for any marked structures $\overline{\mathcal{G}}_1 = (\mathcal{G}_1, a_1, b_1) \simeq_k^{\mathcal{F}} \overline{\mathcal{G}}_2 = (\mathcal{G}_2, a_2, b_2)$, we have $(a_1, b_1) \in e(\mathcal{G}_1)$ if and only if $(a_2, b_2) \in e(\mathcal{G}_2)$.

Let V_1 and V_2 be the node sets of the structures $\mathcal{G}_1$ and $\mathcal{G}_2$, respectively. Let $\bar{Z}$ be an $(\mathcal{F}, k)$-bisimulation from $\mathcal{G}_1$ to $\mathcal{G}_2$ such that $(a_1, b_1, a_2, b_2) \in Z_k$.

For the base case, where e is an atomic expression, the result follows immediately from the Atoms Forth and Back properties applied to $(a_1, b_1, a_2, b_2) \in Z_k$.

If e is $e_1 \cup e_2$, $e_1 \cap e_2$, $e_1 - e_2$, or e_1^c, the result follows immediately from the induction hypothesis.

For the case where e is $e_1 \circ e_2$, consider the only-if, i.e. assume that $(a_1, b_1) \in e(\mathcal{G}_1)$. By definition of composition, there exists c_1 in V_1 with $(a_1, c_1) \in e_1(\mathcal{G}_1)$ and $(c_1, b_1) \in e_2(\mathcal{G}_1)$. Since e_1 and e_2 have depth at most $k-1$, by Proposition 2.7, both (a_1, c_1) and (c_1, b_1) are in $\text{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_1)$. By the Composition Forth condition, there exists c_2 in V_2 such that both (a_1, c_1, a_2, c_2) and (c_1, b_1, c_2, b_2) belong to Z_{k-1}. Hence, by induction (and using Lemma 4.3), we have $(a_2, c_2) \in e_1(\mathcal{G}_2)$ and $(c_2, b_2) \in e_2(\mathcal{G}_2)$, whence $(a_2, b_2) \in e_1 \circ e_2(\mathcal{G}_2)$. The argument for the if-direction is similar, using Composition Back instead of Composition Forth.

For the case where e is $\pi_1(e_1)$, consider the only-if, i.e. assume $(a_1, b_1) \in \pi_1(e_1)(\mathcal{G}_1)$. By definition of projection, we have $a_1 = b_1$, and there exists c_1 in V_1 with $(a_1, c_1) \in e_1(\mathcal{G}_1)$. By Proposition 2.7, we have (a_1, c_1) in $\text{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_1)$. By the Projection Forth condition, there exists c_2 in V_2 such that $(a_1, c_1, a_2, c_2) \in Z_{k-1}$. Hence, by induction, we have $(a_2, c_2) \in e_1(\mathcal{G}_2)$, whence $(a_2, b_2) \in \pi_1(e_1)(\mathcal{G}_2)$. The argument for the if-direction is similar, using Projection Back instead of Projection Forth. The argument for the case where e is $\pi_2(e_1)$ is analogous.

The case where e is $\bar{\pi}_i(e_1)$ now follows readily from the preceding as $\bar{\pi}_i(e_1) \equiv 1' - \pi_i(e_1)$.

Finally, for the case where e is e_1/e_2, consider the only-if, i.e. assume $(a_1, b_1) \in e(\mathcal{G}_1)$. Suppose now that $(a_2, b_2) \notin e_1/e_2(\mathcal{G}_2)$. Then, by definition of the left residual, there exists c_2 in V_2 such that $(b_2, c_2) \in e_2(\mathcal{G}_2)$ and $(a_2, c_2) \notin e_1(\mathcal{G}_2)$. By Proposition 2.7, we have that (b_2, c_2) in $\text{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_2)$. By the Left Residual Forth condition, there exists c_1 in V_1 such that both (b_1, c_1, b_2, c_2) and (a_1, c_1, a_2, c_2)

belong to Z_{k-1}. Hence, by induction, we obtain $(b_1, c_1) \in e_2(\mathcal{G}_1)$ and $(a_1, c_1) \notin e_1(\mathcal{G}_1)$. Now, this c_1 contradicts that $(a_1, b_1) \in e_1/e_2(\mathcal{G}_1)$. The argument for the if-direction is similar, using Left Residual Back.

The case of a right residual is completely analogous to that of a left residual. ∎

The other direction of the adequacy theorem will be established by the Characteristic Expression Lemma. For the proof of that Lemma we need to introduce the construction of the maximal bisimulation. This construction by successive refinement is a classical technique [17, Section 3.5] and is used as well in algorithms for classical bisimilarity [3] and for colour refinement [23].

DEFINITION 4.5
Given a fragment $\mathcal{F}$ as above and structures $\mathcal{G}_1$ and $\mathcal{G}_2$ with node sets V_1 and V_2 respectively, we construct an infinite decreasing sequence $Z_0, Z_1, Z_2, \cdots$ by induction on k as follows.

(1) Z_0 is the set of all elements of $V_1^2 \times V_2^2$ that have the Atoms Forth and Back properties relative to $\mathcal{F}$, $\mathcal{G}_1$ and $\mathcal{G}_2$.
(2) Z_i, for $i > 0$, is the set of all elements in Z_{i-1} that have

- the Composition Forth and Back properties at degree i with respect to Z_{i-1} (still relative to $\mathcal{F}$, $\mathcal{G}_1$ and $\mathcal{G}_2$);
- the Projection Forth and Back properties at degree i with respect to Z_{i-1}, if $\mathcal{F}$ contains projection; and
- the Left (Right) Residual Forth and Back properties at degree i with respect to Z_{i-1}, if $\mathcal{F}$ contains left (right) residual.

We denote this constructed sequence by $\mathrm{BiSim}^{\mathcal{F}}(\mathcal{G}_1, \mathcal{G}_2)$.

The relevant property about $\mathrm{BiSim}^{\mathcal{F}}(\mathcal{G}_1, \mathcal{G}_2)$ is the following. It follows immediately from the definitions.

PROPOSITION 4.6
Let $\mathrm{BiSim}^{\mathcal{F}}(\mathcal{G}_1, \mathcal{G}_2) = Z_0, Z_1, \ldots$. Then for each natural number k, the sequence $Z_0, Z_1, \ldots, Z_k$ is an $(\mathcal{F}, k)$-bisimulation from $\mathcal{G}_1$ to $\mathcal{G}_2$. Furthermore, it is the maximal bisimulation in the sense that, for any other such $(\mathcal{F}, k)$-bisimulation $Z'_0, \ldots, Z'_k$, we have $Z'_i \subseteq Z_i$ for each $i = 0, \ldots, k$.

As an immediate corollary, we have:

COROLLARY 4.7
$(\mathcal{G}_1, a_1, b_1) \simeq^{\mathcal{F}}_k (\mathcal{G}_2, a_2, b_2) \Leftrightarrow (a_1, b_1, a_2, b_2) \in \mathrm{BiSim}^{\mathcal{F}}(\mathcal{G}_1, \mathcal{G}_2)_k$.

Another relevant property is the following.

PROPOSITION 4.8
Let $\mathrm{BiSim}^{\mathcal{F}}(\mathcal{G}_1, \mathcal{G}_2) = Z_0, Z_1, \ldots$. Then Z_i, for every i, is *path-preserving at degree i*, in the sense that if $(a_1, b_1, a_2, b_2) \in Z_i$ and $(a_1, b_1) \in \mathrm{paths}^{\mathcal{F}}_i(\mathcal{G}_1)$, then also $(a_2, b_2) \in \mathrm{paths}^{\mathcal{F}}_i(\mathcal{G}_2)$.

PROOF. By induction on i. The base case $i = 0$ is clear from the definition of Z_0. For the case $i > 0$ we may assume that 1 is absent from $\mathcal{F}$, since otherwise the claim is trivial. We can thus rely on the equivalence $\mathrm{paths}^{\mathcal{F}}_i \equiv \mathrm{paths}^{\mathcal{F}}_{i-1} \cup (\mathrm{paths}^{\mathcal{F}}_{i-1} \circ \mathrm{paths}^{\mathcal{F}}_{i-1})$. If $(a_1, b_1) \in \mathrm{paths}^{\mathcal{F}}_{i-1}(\mathcal{G}_1)$, then, since $Z_i \subseteq Z_{i-1}$, the claim follows directly by induction. Otherwise, there exists $c_1 \in V_1$ such that (a_1, c_1) and (c_1, b_1) are in $\mathrm{paths}^{\mathcal{F}}_{i-1}(\mathcal{G}_1)$. By the Composition Forth Property, there exists $c_2 \in V_2$

such that (a_1,c_1,a_2,c_2) and (c_1,b_1,c_2,b_2) belong to Z. By induction, we have (a_2,c_2) and (c_2,b_2) in $\mathrm{paths}_{i-1}^{\mathcal{F}}(\mathcal{G}_2)$, whence $(a_2,b_2)\in\mathrm{paths}_i^{\mathcal{F}}(\mathcal{G}_2)$ as desired. ∎

Note that, since the above proposition shows path-preservation for the maximal bisimulation, path-preservation also holds for any arbitrary bisimulation.

We are now ready for:

LEMMA 4.9 (Characteristic Expression)
Let k be a natural number and let $\overline{\mathcal{G}}_1=(\mathcal{G}_1,a_1,b_1)$ be a marked structure. Then there exists an expression $e_{\overline{\mathcal{G}}_1}^{\mathcal{F},k}$ in $\mathcal{C}(\mathcal{F})_k$ such that for every structure $\mathcal{G}_2$ we have

$$e_{\overline{\mathcal{G}}_1}^{\mathcal{F},k}(\mathcal{G}_2)=\{(a_2,b_2)\in\mathrm{paths}_k^{\mathcal{F}}(\mathcal{G}_2)\,|\,\overline{\mathcal{G}}_1\simeq_k^{\mathcal{F}}(\mathcal{G}_2,a_2,b_2)\}.$$

PROOF. The construction of the required expression is by induction on k. For the base of the construction we put $e_{\overline{\mathcal{G}}_1}^{\mathcal{F},0}:=\varphi_{\overline{\mathcal{G}}_1,\mathrm{posatoms}}^{\mathcal{F}}-\varphi_{\overline{\mathcal{G}}_1,\mathrm{negatoms}}^{\mathcal{F}}$, where

$$\varphi_{\overline{\mathcal{G}}_1,\mathrm{posatoms}}^{\mathcal{F}}:=\bigcap_{e\in\mathrm{atp}^{\mathcal{F}}(\overline{\mathcal{G}}_1)}e\qquad\text{and}\qquad\varphi_{\overline{\mathcal{G}}_1,\mathrm{negatoms}}^{\mathcal{F}}:=\bigcup_{e\in\mathrm{aexp}(\mathcal{F})-\mathrm{atp}^{\mathcal{F}}(\overline{\mathcal{G}}_1)}e.$$

It is clear that $(a_2,b_2)\in e_{\overline{\mathcal{G}}_1}^{\mathcal{F},0}(\mathcal{G}_2)$ iff $\mathrm{atp}^{\mathcal{F}}(\overline{\mathcal{G}}_1)=\mathrm{atp}^{\mathcal{F}}(\overline{\mathcal{G}}_2)$, which is necessary and sufficient for $\overline{\mathcal{G}}_1\simeq_0^{\mathcal{F}}(\mathcal{G}_2,a_2,b_2)$ to hold.

For the inductive step of the construction, let $k>0$. Our approach, based on Corollary 4.7, is to show that each of the properties P involved in Definition 4.5, P ranging from Composition Forth until Right Residual Back, is expressible in $\mathcal{C}(\mathcal{F})_k$, in the following way. Let V_1 be the node set of the fixed structure $\mathcal{G}_1$, and let V_2 be the node set of any structure $\mathcal{G}_2$ to which our expressions will be applied. We may represent a set $Z\subseteq V_1^2\times V_2^2$ by the family of binary relations consisting of, for each $(a,b)\in V_1^2$, the binary relation

$$Z_{a,b}=\{(a',b')\in\mathrm{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_2)\,|\,(a,b,a',b')\in Z\}.$$

Note that these relations make only the 'slice' of Z visible governed by the constraint $(a',b')\in\mathrm{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_2)$.

We can add all these relations to $\mathcal{G}_2$, yielding an expanded structure, denoted by $(\mathcal{G}_2,Z)$, over the expansion of the given vocabulary Λ with relation names $Z_{a,b}$ for each $(a,b)\in V_1^2$. (Here, we are abusing notation a bit by making no formal distinction between the relation name and its contents.) Furthermore, for our purpose, it will be sufficient to assume that Z is path-preserving at degree $k-1$ in the sense of Proposition 4.8. Now we are going to express each property P by an expression $\varphi_{\overline{\mathcal{G}}_1,P}^{\mathcal{F},k}$ over the expanded vocabulary, in the sense for any $\mathcal{G}_2$ and any Z as above, $\varphi_{\overline{\mathcal{G}}_1,P}^{\mathcal{F},k}$ applied to $(\mathcal{G}_2,Z)$ returns the set of pairs $(a_2,b_2)\in\mathrm{paths}_k^{\mathcal{F}}(\mathcal{G}_2)$ for which (a_1,b_1,a_2,b_2) satisfies the P property at degree k with respect to Z, relative to $\mathcal{G}_1$, $\mathcal{G}_2$, and $\mathcal{F}$.

There is a caveat: if $\mathcal{G}_1$ is infinite, there are infinitely many pairs $(a,b)\in V_1^2$, so also infinitely many relations $Z_{a,b}$. Accordingly, we will allow the expression $\varphi_{\overline{\mathcal{G}}_1,P}^{\mathcal{F},k}$ to be infinitary, in that it can use infinite unions and intersections. We will see later in the inductive argument that normal, finitary expressions are still obtained in the end.

For example, for the Composition Forth property, we have the following expression, whose correctness is evident:

$$\varphi^{\mathcal{F},k}_{\overline{\mathcal{G}}_1,\text{composition forth}} := \text{paths}^{\mathcal{F}}_k \cap \bigcap_{\substack{c_1 \in V_1 \\ (a_1,c_1) \in \text{paths}^{\mathcal{F}}_{k-1}(\mathcal{G}_1) \\ (c_1,b_1) \in \text{paths}^{\mathcal{F}}_{k-1}(\mathcal{G}_1)}} Z_{a_1,c_1} \circ Z_{c_1,b_1}.$$

Here and below, it is understood that an empty intersection vanishes from the expression. Empty unions, as usual, are replaced by the expression 0.

For the Projection Forth property, if $a_1 \neq b_1$, we can simply define

$$\varphi^{\mathcal{F},k}_{\overline{\mathcal{G}}_1,\text{projection forth}} := \text{paths}^{\mathcal{F}}_k.$$

Otherwise, if $a_1 = b_1$, we use

$$\varphi^{\mathcal{F},k}_{\overline{\mathcal{G}}_1,\text{projection forth}} := 1'$$
$$\cap \bigcap_{\substack{c_1 \in V_1 \\ (a_1,c_1) \in \text{paths}^{\mathcal{F}}_{k-1}(\mathcal{G}_1)}} \pi_1(Z_{a_1,c_1}) \cap \bigcap_{\substack{c_1 \in V_1 \\ (c_1,a_1) \in \text{paths}^{\mathcal{F}}_{k-1}(\mathcal{G}_1)}} \pi_2(Z_{c_1,a_1}).$$

Again the correctness is evident.

For the Projection Back property, if $a_1 \neq b_1$, we can simply define

$$\varphi^{\mathcal{F},k}_{\overline{\mathcal{G}}_1,\text{projection back}} := \text{paths}^{\mathcal{F}}_k - 1'.$$

Indeed, if $(a_2,b_2) \in (\text{paths}^{\mathcal{F}}_k - 1')(\mathcal{G}_2)$ then $a_2 \neq b_2$, in which case (a_1,b_1,a_2,b_2) voidly satisfies the Projection Back property. Conversely, if $(a_2,b_2) \in \text{paths}^{\mathcal{F}}_k(\mathcal{G}_2)$ and (a_1,b_1,a_2,b_2) satisfies the Projection Back property, but $a_1 \neq b_1$, then $a_2 \neq b_2$ must hold as well so $(a_2,b_2) \in (\text{paths}^{\mathcal{F}}_k - 1')(\mathcal{G}_2)$ as desired.

If $a_1 = b_1$, we put

$$\varphi^{\mathcal{F},k}_{\overline{\mathcal{G}}_1,\text{projection back}} := \text{paths}^{\mathcal{F}}_k$$
$$- \left(1' \cap (\pi_1(\text{paths}^{\mathcal{F}}_{k-1} - \bigcup_{c_1 \in V_1} Z_{a_1,c_1}) \cup \pi_2(\text{paths}^{\mathcal{F}}_{k-1} - \bigcup_{c_1 \in V_1} Z_{c_1,a_1}))\right).$$

Let us verify the correctness in this case. Let $(a_2,b_2) \in \varphi^{\mathcal{F},k}_{\overline{\mathcal{G}}_1,\text{projection back}}(\mathcal{G}_2,Z)$. Then clearly $(a_2,b_2) \in \text{paths}^{\mathcal{F}}_k(\mathcal{G}_2)$. We must show that (a_1,b_1,a_2,b_2) satisfies the Projection Back property at degree k with respect to Z. If $a_2 \neq b_2$ this is trivial so assume $a_2 = b_2$. Then $(a_2,b_2) \in 1'(\mathcal{G}_2)$, so, considering the expression $\varphi^{\mathcal{F},k}_{\overline{\mathcal{G}}_1,\text{projection back}}$, this means that (i) $(a_2,a_2) \notin \pi_1(\text{paths}^{\mathcal{F}}_{k-1} - \bigcup_{c_1 \in V_1} Z_{a_1,c_1})(\mathcal{G}_2,Z)$ and (ii) $(a_2,a_2) \notin \pi_2(\text{paths}^{\mathcal{F}}_{k-1} - \bigcup_{c_1 \in V_1} Z_{c_1,a_1})(\mathcal{G}_2,Z)$. Now let $c_2 \in V_2$ such that $(a_2,c_2) \in \text{paths}^{\mathcal{F}}_{k-1}(\mathcal{G}_2)$. We must show there exists $c_1 \in V_1$ such that $(a_1,c_1,a_2,c_2) \in Z$. For the sake of contradiction, assume the contrary; then $(a_2,c_2) \in (\text{paths}^{\mathcal{F}}_{k-1} - \bigcup_{c_1 \in V_1} Z_{a_1,c_1})(\mathcal{G}_2,Z)$. By (i), this is impossible. Similarly, using (ii), we obtain that for any $c_2 \in V_2$ such that $(c_2,a_2) \in \text{paths}^{\mathcal{F}}_{k-1}(\mathcal{G}_2)$ there exists $c_1 \in V_1$ such that $(c_1,a_1,c_2,a_2) \in Z$. Hence the Projection Back property holds as desired.

The converse direction, that any $(a_2, b_2) \in \mathrm{paths}_k^{\mathcal{F}}(\mathcal{G}_2)$ must belong to $\varphi_{\overline{\mathcal{G}}_1, \text{projection back}}^{\mathcal{F},k}(\mathcal{G}_2, Z)$ if (a_1, b_1, a_2, b_2) satisfies the Projection Back property at degree k with respect to Z, is argued similarly.

For the Composition Back property, the expression is a little bit less evident:

$$\varphi_{\overline{\mathcal{G}}_1, \text{composition back}}^{\mathcal{F},k} := \mathrm{paths}_k^{\mathcal{F}}$$
$$- \bigcup_{V \subseteq V_1} \left((\mathrm{paths}_{k-1}^{\mathcal{F}} - \bigcup_{c_1 \in V} Z_{a_1, c_1}) \circ (\mathrm{paths}_{k-1}^{\mathcal{F}} - \bigcup_{c_1 \in V_1 - V} Z_{c_1, b_1}) \right)$$

To see the correctness of the above expression, we must show for any $(a_2, b_2) \in \mathrm{paths}_k^{\mathcal{F}}(\mathcal{G}_2)$ that $(a_2, b_2) \in \varphi_{\overline{\mathcal{G}}_1, \text{composition back}}^{\mathcal{F},k}(\mathcal{G}_2, Z)$ if and only if (a_1, b_1, a_2, b_2) has the Composition Back property at degree k with respect to Z. For the only-if direction, assume $(a_2, b_2) \in \varphi_{\overline{\mathcal{G}}_1, \text{composition back}}^{\mathcal{F},k}(\mathcal{G}_2, Z)$. Let $c_2 \in V_2$ such that both (a_2, c_2) and (c_2, b_2) are in $\mathrm{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_2)$. We must show that there exists $c_1 \in V_1$ such that both (a_1, c_1, a_2, c_2) and (c_1, b_1, c_2, b_2) belong to Z. For the sake of contradiction, suppose such c_1 does not exist, i.e. for each $c_1 \in V_1$ either $(a_2, c_2) \notin Z_{a_1, c_1}$ or $(c_2, b_2) \notin Z_{c_1, b_1}$ (or both). Then, letting $V := \{ c_1 \in V_1 \mid (a_2, c_2) \notin Z_{a_1, c_1} \}$, we have for any $c_1 \in V_1 - V$ that $(c_2, b_2) \notin Z_{c_1, b_1}$. Hence, through c_2, we see that

$$(a_2, b_2) \in (\mathrm{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_2) - \bigcup_{c_1 \in V} Z_{a_1, c_1}) \circ (\mathrm{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_2) - \bigcup_{c_1 \in V_1 - V} Z_{c_1, b_1}),$$

which contradicts $(a_2, b_2) \in \varphi_{\overline{\mathcal{G}}_1, \text{composition back}}^{\mathcal{F},k}(\mathcal{G}_2, Z)$.

For the if-direction, assume that (a_1, b_1, a_2, b_2) has the Composition Back property at degree k with respect to Z, and we must show that (a_2, b_2) belongs to $\varphi_{\overline{\mathcal{G}}_1, \text{composition back}}^{\mathcal{F},k}(\mathcal{G}_2, Z)$. For the sake of contradiction, assume there exists $V \subseteq V_1$ and $c_2 \in V_2$ such that $(a_2, c_2) \in \mathrm{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_2) - Z_{a_1, c_1}$ for all $c_1 \in V$, and $(c_2, b_2) \in \mathrm{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_2) - Z_{b_1, c_1}$ for all $c_1 \in V_1 - V$. This simply means that there exists $c_2 \in V_2$ such that (a_2, c_2) and (c_2, b_2) both belong to $\mathrm{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_2)$, but for which there exists no $c_1 \in V_1$ such that both (a_1, c_1, a_2, c_2) and (c_1, b_1, c_2, b_2) belong to Z. Thus we obtain a direct contradiction with the Composition Back property.

The expressions for the Left and Right Residual Back properties are more straightforward again:

$$\varphi_{\overline{\mathcal{G}}_1, \text{leftres back}}^{\mathcal{F},k} := \mathrm{paths}_k^{\mathcal{F}} - \bigcup_{\substack{c_1 \in V_1 \\ (b_1, c_1) \in \mathrm{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_1)}} (\mathrm{paths}_{k-1}^{\mathcal{F}} - Z_{a_1, c_1}) / Z_{b_1, c_1}$$

$$\varphi_{\overline{\mathcal{G}}_1, \text{rightres back}}^{\mathcal{F},k} := \mathrm{paths}_k^{\mathcal{F}} - \bigcup_{\substack{c_1 \in V_1 \\ (c_1, a_1) \in \mathrm{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_1)}} Z_{c_1, a_1} \backslash (\mathrm{paths}_{k-1}^{\mathcal{F}} - Z_{c_1, b_1})$$

Let us show the correctness of the Left Residual Back expression; the argument for the Right Residual is completely analogous. Let $(a_2, b_2) \in \mathrm{paths}_k^{\mathcal{F}}(\mathcal{G}_2)$. We see that $(a_2, b_2) \in \varphi_{\overline{\mathcal{G}}_1, \text{leftres back}}^{\mathcal{F},k}$ if and only if there does *not* exist $c_1 \in V_1$ such that $(b_1, c_1) \in \mathrm{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_1)$, and such that for each $c_2 \in V_2$ with $(b_2, c_2) \in Z_{b_1, c_1}$ we have $(a_2, c_2) \in \mathrm{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_2) - Z_{a_1, c_1}$. Equivalently, $(a_2, b_2) \in \varphi_{\overline{\mathcal{G}}_1, \text{leftres back}}^{\mathcal{F},k}$ iff for all $c_1 \in V_1$ with $(b_1, c_1) \in \mathrm{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_1)$, there exists $c_2 \in V_2$ such that $(b_2, c_2) \in \mathrm{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_2)$ and $(b_1, c_1, b_2, c_2) \in Z$ and $(a_2, c_2) \notin \mathrm{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_2) - Z_{a_1, c_1}$. By the path-preserving property of Z, the

qualification $(b_2,c_2)\in\text{paths}^{\mathcal{F}}_{k-1}(\mathcal{G}_2)$ is redundant. Moreover, $(a_2,c_2)\notin\text{paths}^{\mathcal{F}}_{k-1}(\mathcal{G}_2)-Z_{a_1,c_1}$ means that either $(a_2,c_2)\notin\text{paths}^{\mathcal{F}}_{k-1}(\mathcal{G}_2)$ or $(a_1,c_1,b_1,c_2)\in Z$. Thus we get exactly the formulation of the Left Residual Back property.

Finally, the Left and Right Residual Forth properties are expressed using a similar approach as for the Composition Back property:

$$\varphi^{\mathcal{F},k}_{\overline{\mathcal{G}}_1,\text{leftres forth}}:=\text{paths}^{\mathcal{F}}_k$$
$$\cap\bigcap_{V\subseteq V_1}\left[(\bigcup_{c_1\in V_1-V}Z_{a_1,c_1})/(\text{paths}^{\mathcal{F}}_{k-1}-\bigcup_{\substack{c_1\in V_1\\(a_1,c_1)\notin\text{paths}^{\mathcal{F}}_{k-1}(\mathcal{G}_1)}}Z_{b_1,c_1}-\bigcup_{c_1\in V}Z_{b_1,c_1})\right]$$

$$\varphi^{\mathcal{F},k}_{\overline{\mathcal{G}}_1,\text{rightres forth}}:=\text{paths}^{\mathcal{F}}_k$$
$$\cap\bigcap_{V\subseteq V_1}\left[(\text{paths}^{\mathcal{F}}_{k-1}-\bigcup_{\substack{c_1\in V_1\\(c_1,b_1)\notin\text{paths}^{\mathcal{F}}_{k-1}(\mathcal{G}_1)}}Z_{c_1,a_1}-\bigcup_{c_1\in V}Z_{c_1,a_1})\backslash\bigcup_{c_1\in V_1-V}Z_{c_1,b_1}\right]$$

Let us show the correctness of the Left Residual Forth expression; again the argument for the Right Residual is completely analogous. We must show for any $(a_2,b_2)\in\text{paths}^{\mathcal{F}}_k(\mathcal{G}_2)$ that $(a_2,b_2)\in\varphi^{\mathcal{F},k}_{\overline{\mathcal{G}}_1,\text{leftres forth}}(\mathcal{G}_2,Z)$ if and only if (a_1,b_1,a_2,b_2) has the Left Residual Forth property at degree k with respect to Z. For the only-if direction, assume $(a_2,b_2)\in\varphi^{\mathcal{F},k}_{\overline{\mathcal{G}}_1,\text{leftres forth}}(\mathcal{G}_2,Z)$. Let $c_2\in V_2$ such that $(b_2,c_2)\in\text{paths}^{\mathcal{F}}_{k-1}(\mathcal{G}_2)$. We must show that there exists $c_1\in V_1$ such that $(b_1,c_1,b_2,c_2)\in Z$ and either $(a_1,c_1)\notin\text{paths}^{\mathcal{F}}_{k-1}(\mathcal{G}_1)$ or $(a_1,c_1,a_2,c_2)\in Z$. If there exists $c_1\in V_1$ with $(a_1,c_1)\notin\text{paths}^{\mathcal{F}}_{k-1}$ such that $(b_1,c_1,b_2,c_2)\in Z$, there is nothing to prove. So, suppose no such c_1 exists, and consider $V:=\{c_1\in V_1\,|\,(b_2,c_2)\notin Z_{b_1,c_1}\}$. Then

$$(b_2,c_2)\in\text{paths}^{\mathcal{F}}_{k-1}(\mathcal{G}_2)-\bigcup_{\substack{c_1\in V_1\\(a_1,c_1)\notin\text{paths}^{\mathcal{F}}_{k-1}(\mathcal{G}_1)}}Z_{b_1,c_1}-\bigcup_{c_1\in V}Z_{b_1,c_1}.$$

Hence, since $(a_2,b_2)\in\varphi^{\mathcal{F},k}_{\overline{\mathcal{G}}_1,\text{leftres forth}}(\mathcal{G}_2,Z)$, we have $(a_2,c_2)\in\bigcup_{c_1\in V_1-V}Z_{a_1,c_1}$, i.e. there exists $c_1\in V$ such that $(a_1,c_1,a_2,c_2)\in Z$ and $(b_1,c_1,b_2,c_2)\in Z$, as desired.

For the if-direction, assume that (a_1,b_1,a_2,b_2) has the Left Residual Forth property at degree k with respect to Z, and let $V\subseteq V_1$ be arbitrary. Let $c_2\in V_2$ such that

$$(b_2,c_2)\in\text{paths}^{\mathcal{F}}_{k-1}(\mathcal{G}_2)-\bigcup_{\substack{c_1\in V_1\\(a_1,c_1)\notin\text{paths}^{\mathcal{F}}_{k-1}(\mathcal{G}_1)}}Z_{b_1,c_1}-\bigcup_{c_1\in V}Z_{b_1,c_1}.$$

By the Residual Forth property, there exists $c_1\in V_1$ such that $(b_1,c_1,b_2,c_2)\in Z$ (in particular, $c_1\in V_1-V$) and either $(a_1,c_1)\notin\text{paths}^{\mathcal{F}}_{k-1}(\mathcal{G}_1)$ or $(a_1,c_1,a_2,c_2)\in Z$. The possibility $(a_1,c_1)\notin\text{paths}^{\mathcal{F}}_{k-1}(\mathcal{G}_1)$ cannot occur, however, because

$$(b_2,c_2)\in\text{paths}^{\mathcal{F}}_{k-1}(\mathcal{G}_2)-\bigcup_{\substack{c_1\in V_1\\(a_1,c_1)\notin\text{paths}^{\mathcal{F}}_{k-1}(\mathcal{G}_1)}}Z_{b_1,c_1}.$$

Hence, $(a_1,c_1)\in\text{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_1)$ and $(a_1,c_1,a_2,c_2)\in Z$, whence $(a_2,c_2)\in Z_{a_1,c_1}$, since Z was assumed to be path-preserving. We conclude that

$$(a_2,c_2)\in\bigcup_{c_1\in V_1-V}Z_{a_1,c_1}$$

as desired.

We are now ready to conclude the construction of the required expression $e_{\overline{\mathcal{G}}_1}^{\mathcal{F},k}$. This expression, applied to any $\mathcal{G}_2$, should return the set of all pairs (a_2,b_2) such that $(a_1,b_1,a_2,b_2)\in\text{BiSim}^{\mathcal{F}}(\mathcal{G}_1,\mathcal{G}_2)_k$. By definition, these are the pairs (a_2,b_2) for which $(a_1,b_1,a_2,b_2)\in\text{BiSim}^{\mathcal{F}}(\mathcal{G}_1,\mathcal{G}_2)_{k-1}$, and such that (a_1,b_1,a_2,b_2) has all the properties P, required by Definition 4.5, at degree k with respect to $\text{BiSim}^{\mathcal{F}}(\mathcal{G}_1,\mathcal{G}_2)_{k-1}$. We have just seen that each such property P is expressible by the infinitary expression $\varphi_{\overline{\mathcal{G}}_1,P}^{\mathcal{F},k}$. Hence, we can obtain $e_{\overline{\mathcal{G}}_1}^{\mathcal{F},k}$ simply as the intersection of $e_{\overline{\mathcal{G}}_1}^{\mathcal{F},k-1}$ (obtained by induction) and the expressions $\varphi_{\overline{\mathcal{G}}_1,P}^{\mathcal{F},k}$ for the different properties P required by $(\mathcal{F},k)$-bisimulation.

The only problem remaining is that each expression $\varphi_{\overline{\mathcal{G}}_1,P}^{\mathcal{F},k}$ is still infinitary, and referring to extra relation names of the form $Z_{a,b}$. For our purpose, such a relation name should hold the relation $\{(a',b')\in\text{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_2)\,|\,(a,b,a',b')\in\text{BiSim}^{\mathcal{F}}(\mathcal{G}_1,\mathcal{G}_2)_{k-1}\}$. By the induction hypothesis however, we can express this relation by the expression $e_{\mathcal{G}_1,a,b}^{\mathcal{F},k-1}$. So, in $\varphi_{\overline{\mathcal{G}}_1,P}^{\mathcal{F},k}$, we can replace each occurrence of $Z_{a,b}$ by $e_{\mathcal{G}_1,a,b}^{\mathcal{F},k-1}$ and obtain an expression of degree k over the original given vocabulary Λ.

The resulting expression still has infinite unions and intersections. These unions and intersections are over sets of expressions of degree k, however. Hence, since, up to equivalence, there are only finitely many expressions of degree k over the fixed finite vocabulary Λ, we can equivalently replace the infinite unions and intersections by finite ones. The reason why there are only a finite number of inequivalent expression of degree k is the same as why there are only a finite number of inequivalent first-order logic formulas of quantifier rank k [10]. ∎

We can conclude the proof of the Adequacy Theorem as follows:

PROOF OF THEOREM 4.2 The only-if direction is proven by Lemma 4.4. The if-direction for $k=0$ is clear. So now assume $(\mathcal{G}_1,a_1,b_1)\equiv_k^{\mathcal{F}}(\mathcal{G}_2,a_2,b_2)$ with $k>0$. We distinguish two possibilities. If $(a_2,b_2)\notin\text{paths}_k^{\mathcal{F}}(\mathcal{G}_2)$, then $\mathcal{F}$ cannot contain the residuals, for otherwise $\text{paths}_k^{\mathcal{F}}\equiv 1$. (The only exception is when V_2 is empty, but then V_1 must be empty as well and the theorem becomes trivial.) Moreover, since $\text{paths}_k^{\mathcal{F}}$ is expressible by an expression of degree k and $(\mathcal{G}_1,a_1,b_1)\equiv_k^{\mathcal{F}}(\mathcal{G}_2,a_2,b_2)$, also $(a_1,b_1)\notin\text{paths}_k^{\mathcal{F}}(\mathcal{G}_1)$. But in that case we can see that the Atoms Forth and Back, the Composition Forth and Back, as well as the Projection Forth and Back properties are void, so $(\mathcal{G}_1,a_1,b_1)\simeq_k^{\mathcal{F}}(\mathcal{G}_2,a_2,b_2)$ holds trivially.

Hence, the non-degenerate case is where $(a_2,b_2)\in\text{paths}_k^{\mathcal{F}}(\mathcal{G}_2)$ and $(a_1,b_1)\in\text{paths}_k^{\mathcal{F}}(\mathcal{G}_1)$, which allows us to invoke the Charasteristic Expression Lemma. We argue as follows. First, we note that $(\mathcal{G}_1,a_1,b_1)\simeq_k^{\mathcal{F}}(\mathcal{G}_1,a_1,b_1)$ trivially holds; indeed, we can take the bisimulation $\bar{Z}$ where $Z_i=\{(a,b,a,b)\,|\,(a,b)\in V_1^2\}$ for each i. Hence, by the Characteristic Expression Lemma, we have $(a_1,b_1)\in e_{\mathcal{G}_1,a_1,b_1}^{\mathcal{F},k}(\mathcal{G}_1)$. Since $(\mathcal{G}_1,a_1,b_1)\equiv_k^{\mathcal{F}}(\mathcal{G}_2,a_2,b_2)$, this implies $(a_2,b_2)\in e_{\mathcal{G}_1,a_1,b_1}^{\mathcal{F},k}(\mathcal{G}_2)$. Again by the Characteristic Expression Lemma this implies $(\mathcal{G}_1,a_1,b_1)\simeq_k^{\mathcal{F}}(\mathcal{G}_1,a_1,b_1)$ and we are done. ∎

To conclude this section we note as an immediate corollary that the Projection properties, or the Left (Right) Residual properties, can be omitted from the definition of bisimulation in those cases

where the corresponding operation is not primitive:

COROLLARY 4.10
- Let $\mathcal{C}(\mathcal{F})$ be a calculus fragment where $\mathcal{F}$ contains projection, and either converse is present as well or 1 is present at degree 0. Let $\mathcal{F}'$ be $\mathcal{F}$ without projection. Then $(\mathcal{F}, k)$-bisimilarity is the same as $(\mathcal{F}', k)$-bisimilarity.
- Similarly, when $\mathcal{F}$ contains the left (right) residual and also converse, and 1 is present at degree 0 as well, then $(\mathcal{F}, k)$-bisimilarity is the same as $(\mathcal{F}', k)$-bisimilarity where $\mathcal{F}'$ is $\mathcal{F}$ without left (right) residual.

5 Similarity and one-sided indistinguishability

In the present section, we deal with fragments not containing difference, for which we will capture one-sided indistinguishability by appropriate notions of simulation between structures. The treatment will largely parallel that for fragments with difference from the previous section. Nevertheless, simulations differ from bisimulations in that they consist of two separate sequences $\bar{Z}$ and $\bar{W}$ of relations, one for each direction. The two separate directions are needed to be able to deal with the non-monotonic operations of coprojection and left and right residual, in the absence of complement and difference.

When the fragment contains neither difference, nor coprojection, nor residuals, it will be evident that the definition of simulation boils down to a simpler situation where only the sequence $\bar{Z}$ is needed.

5.1 *General definition of $(\mathcal{F}, k)$-simulation*

Let $\mathcal{G}_1$ and $\mathcal{G}_2$ be two structures with node sets V_1 and V_2, respectively. Let (a_1, b_1, a_2, b_2) be an arbitrary element of $V_1^2 \times V_2^2$, and let Z and W be arbitrary subset of $V_1^2 \times V_2^2$.

In parallel to Section 4.1, we define a suite of conditions, but now appropriate for calculus fragments $\mathcal{C}(\mathcal{F})$ containing neither complement nor difference. The Atoms Forth and Back are unmodified. Also the Composition Forth and Back, and Projection Forth and Back are unmodified with respect to their definitions in Section 4.1, with the exception that they are now defined with respect to two sets Z and W. The Left and Right Residual Forth and Back properties, however, are modified in that they 'cross over' between Z and W. In the same spirit we also define Coprojection Forth and Back properties.

Composition and Projection Forth We say that (a_1, b_1, a_2, b_2) has the *Composition Forth, or Projection Forth, property at degree i with respect to Z, W* if (a_1, b_1, a_2, b_2) has that property at degree i with respect to the set Z, as defined in Section 4.1.

Composition and Projection Back We say that (a_1, b_1, a_2, b_2) has the *Composition Back, or Projection Back, property at degree i with respect to Z, W* if (a_1, b_1, a_2, b_2) has that property at degree i with respect to the set W, as defined in Section 4.1.

Left Residual Forth We say that (a_1, b_1, a_2, b_2) has the *Left Residual Forth property at degree i with respect to Z, W* if for every c_2 in V_2 with (b_2, c_2) in $\mathrm{paths}_{i-1}^{\mathcal{F}}(\mathcal{G}_2)$, there exists c_1 in V_1 such that $(b_1, c_1, b_2, c_2) \notin W$ and either $(a_1, c_1) \notin \mathrm{paths}_{i-1}^{\mathcal{F}}(\mathcal{G}_1)$ or $(a_1, c_1, a_2, c_2) \in Z$.

Left Residual Back We say that (a_1, b_1, a_2, b_2) has the *Left Residual Back property at degree i with respect to Z, W* if for every c_1 in V_1 with (b_1, c_1) in $\mathrm{paths}_{i-1}^{\mathcal{F}}(\mathcal{G}_1)$, there exists c_2 in V_2 such that both $(b_1, c_1, b_2, c_2) \in Z$ and either $(a_2, c_2) \notin \mathrm{paths}_{i-1}^{\mathcal{F}}(\mathcal{G}_2)$ or $(a_1, c_1, a_2, c_2) \in W$.

Right Residual Forth We say that (a_1, b_1, a_2, b_2) has the *Right Residual Forth property at degree i with respect to Z, W* if for every c_2 in V_2 with (c_2, a_2) in $\text{paths}_{i-1}^{\mathcal{F}}(\mathcal{G}_2)$, there exists c_1 in V_1 such that both $(c_1, a_1, c_2, a_2) \in W$ and either $(c_1, b_1) \notin \text{paths}_{i-1}^{\mathcal{F}}(\mathcal{G}_1)$ or $(c_1, b_1, c_2, b_2) \in Z$.

Right Residual Back We say that (a_1, b_1, a_2, b_2) has the *Right Residual Back property at degree i with respect to Z, W* if for every c_1 in V_1 with (c_1, a_1) in $\text{paths}_{i-1}^{\mathcal{F}}(\mathcal{G}_1)$, there exists c_2 in V_2 such that both $(c_1, a_1, c_2, a_2) \in Z$ and either $(c_2, b_2) \notin \text{paths}_{i-1}^{\mathcal{F}}(\mathcal{G}_2)$ or $(c_1, b_1, c_2, b_2) \in W$.

Coprojection Forth We say that (a_1, b_1, a_2, b_2) has the *Projection Forth property at degree i with respect to Z, W* if either $a_1 \neq b_1$, or $a_1 = b_1$ and $a_2 = b_2$ and for every c_2 in V_2 with (a_2, c_2) in $\text{paths}_{i-1}^{\mathcal{F}}(\mathcal{G}_2)$, there exists c_1 in V_1 such that $(a_1, c_1, a_2, c_2) \in W$. Moreover, if $a_1 = b_1$, then also for every c_2 in V_2 with (c_2, a_2) in $\text{paths}_{i-1}^{\mathcal{F}}(\mathcal{G}_2)$, there must exist c_1 in V_1 such that $(c_1, a_1, c_2, a_2) \in W$.

Coprojection Back We say that (a_1, b_1, a_2, b_2) has the *Coprojection Back property at degree i with respect to Z, W* if either $a_2 \neq b_2$, or $a_2 = b_2$ and $a_1 = b_1$ and for every c_1 in V_1 with (a_1, c_1) in $\text{paths}_{i-1}^{\mathcal{F}}(\mathcal{G}_1)$, there exists c_2 in V_2 such that $(a_1, c_1, a_2, c_2) \in Z$. Moreover, if $a_2 = b_2$ then also for every c_1 in V_1 with (c_1, a_1) in $\text{paths}_{i-1}^{\mathcal{F}}(\mathcal{G}_1)$, there must exist c_2 in V_2 such that $(c_1, a_1, c_2, a_2) \in Z$.

Now let k be a natural number and let $\bar{Z} = (Z_0, Z_1, \ldots, Z_k)$ and $\bar{W} = (W_0, W_1, \ldots, W_k)$ be decreasing sequences of relations with Z_0 and W_0 subsets of $V_1^2 \times V_2^2$.

We now lift the above conditions to apply to such pairs $(\bar{Z}, \bar{W})$ of sequences. Plainly, the Forth properties apply to $\bar{Z}$ and the Back properties to $\bar{W}$.

- We say that $(\bar{Z}, \bar{W})$ has the Atoms Forth property if every element of Z_0 has this property.
- We say that $(\bar{Z}, \bar{W})$ has the Atoms Back property if every element of W_0 has this property.
- We say that $(\bar{Z}, \bar{W})$ has the Composition Forth, or the Projection Forth , or the Coprojection Forth, or the Left or Right Residual Forth property, if for every $i \in \{1, \ldots, k\}$, every element of Z_i has that property at degree i with respect to (Z_{i-1}, W_{i-1}).
- We say that $(\bar{Z}, \bar{W})$ has the Composition Back, or the Projection Back, or the Coprojection Back, or the Left or Right Residual Back property, if for every $i \in \{1, \ldots, k\}$, every element of W_i has that property at degree i with respect to (Z_{i-1}, W_{i-1}).

We then naturally have the following:

DEFINITION 5.1
We call $(\bar{Z}, \bar{W})$ an $(\mathcal{F}, k)$-simulation from $\mathcal{G}_1$ to $\mathcal{G}_2$ if $(\bar{Z}, \bar{W})$ has

- the Atoms Forth and Back properties;
- the Composition Forth and Back properties;
- the Projection Forth and Back properties when $\mathcal{F}$ contains projection;
- the Coprojection Forth and Back properties when $\mathcal{F}$ contains coprojection; and
- the Left (Right) Residual Forth and Back properties when $\mathcal{F}$ contains left (right) residual.

Given two marked structures $\overline{\mathcal{G}}_1 = (\mathcal{G}_1, a_1, b_1)$ and $\overline{\mathcal{G}}_2 = (\mathcal{G}_2, a_2, b_2)$, when there exists an $(\mathcal{F}, k)$-simulation $(\bar{Z}, \bar{W})$ from $\mathcal{G}_1$ to $\mathcal{G}_2$ such that $(a_1, b_1, a_2, b_2) \in Z_k$, we say that $\overline{\mathcal{G}}_1$ is $(\mathcal{F}, k)$-*similar* to $\overline{\mathcal{G}}_2$, and denote this by $\overline{\mathcal{G}}_1 \preceq_k^{\mathcal{F}} \overline{\mathcal{G}}_2$.

It is instructive to remark that, in the case where $\mathcal{F}$ does not have coprojection or residual, only $\bar{Z}$ matters; the component $\bar{W}$ is then entirely redundant, in the sense that the sequence $\emptyset, \ldots, \emptyset$ $(k+1$ times) would do fine, as it trivially satisfies the Composition and Projection Back properties. Only when coprojection or residual are present, there is a significant interplay between $\bar{Z}$ and $\bar{W}$.

After Definition 4.1 of bisimilarity, we observed that there it does not matter whether or not $0'$ belongs to the fragment. In contrast, here this matters, since the Atoms Forth property applies only to $\bar{Z}$ and the Atoms Back property applies only to $\bar{W}$.

The following important property follows immediately from the symmetries in the definition of simulation. For any $Z \subseteq V_1^2 \times V_2^2$, we define $\tilde{Z} := \{(a',b',a,b) \mid (a,b,a',b') \in Z\}$. We then have:

PROPOSITION 5.2
If $(Z_0,\ldots,Z_k; W_0,\ldots,W_k)$ is an $(\mathcal{F},k)$-simulation from $\mathcal{G}_1$ to $\mathcal{G}_2$, then $(\tilde{W}_0,\ldots,\tilde{W}_k; \tilde{Z}_0,\ldots,\tilde{Z}_k)$ is an $(\mathcal{F},k)$-simulation from $\mathcal{G}_2$ to $\mathcal{G}_1$.

We also note the following analogue of Lemma 4.3:

LEMMA 5.3
Let $k > 0$ and let $(Z_0,Z_1,\ldots,Z_k; W_0,W_1,\ldots,W_k)$ be an $(\mathcal{F},k)$-simulation from $\mathcal{G}_1$ to $\mathcal{G}_2$. Then $(Z_0,\ldots,Z_{k-1}; W_0,\ldots,W_{k-1})$ is an $(\mathcal{F},k-1)$-simulation from $\mathcal{G}_1$ to $\mathcal{G}_2$.

5.2 *Adequacy theorem*

We establish:

THEOREM 5.4 (Adequacy Theorem)
For any fragment $\mathcal{C}(\mathcal{F})$ where $\mathcal{F}$ contains neither complement nor difference, we have $\overline{\mathcal{G}}_1 \preceq_k^{\mathcal{F}} \overline{\mathcal{G}}_2$ if and only if $\overline{\mathcal{G}}_1 \Rrightarrow_k^{\mathcal{F}} \overline{\mathcal{G}}_2$.

We have the following analogue of Lemma 4.4:

LEMMA 5.5 (Invariance)
If $\overline{\mathcal{G}}_1 \preceq_k^{\mathcal{F}} \overline{\mathcal{G}}_2$ then $\overline{\mathcal{G}}_1 \Rrightarrow_k^{\mathcal{F}} \overline{\mathcal{G}}_2$.

PROOF. Let e be an expression in $\mathcal{C}(\mathcal{F})_k$. We prove by induction on the structure of e that for marked structures $\overline{\mathcal{G}}_1 = (\mathcal{G}_1,a_1,b_1) \preceq_k^{\mathcal{F}} \overline{\mathcal{G}}_2 = (\mathcal{G}_2,a_2,b_2)$ and any $(a_1,b_1) \in e(\mathcal{G}_1)$, we also have $(a_2,b_2) \in e(\mathcal{G}_2)$.

Let V_1 and V_2 be the node sets of the structures $\mathcal{G}_1$ and $\mathcal{G}_2$, respectively. Let $(\bar{Z},\bar{W})$ be an $(\mathcal{F},k)$-simulation from $\mathcal{G}_1$ to $\mathcal{G}_2$ such that $(a_1,b_1,a_2,b_2) \in Z_k$.

For the cases where e is an atomic expression, a union, an intersection, a composition, or a projection, the reasoning is identical to the corresponding only-if cases in the proof of Lemma 4.4.

Consider the case where e is $\bar{\pi}_1(e_1)$. By definition of coprojection, we have $a_1 = b_1$, whence $a_2 = b_2$ by the Atoms Forth condition. We have to show that there does not exist $c_2 \in V_2$ with $(a_2,c_2) \in e_1(\mathcal{G}_2)$. For the sake of contradiction, suppose there exists such c_2. By Lemma 2.7, we have $(a_2,c_2) \in \text{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_2)$. Then by the Coprojection Forth property, there exists $c_1 \in V_1$ such that $(a_1,c_1,a_2,c_2) \in W_{k-1}$. Hence, by Proposition 5.2, Lemma 5.3 and the induction hypothesis, we obtain $(a_1,c_1) \in e_1(\mathcal{G}_1)$ which is in contradiction with $(a_1,b_1) \in \bar{\pi}_1(e_1)(\mathcal{G}_1)$.

Finally, consider the case where e is e_1/e_2. So we have to show that $(a_2,b_2) \in e_1/e_2(\mathcal{G}_2)$. Thereto, let $c_2 \in V_2$ such that $(b_2,c_2) \in e_2(\mathcal{G}_2)$. By Proposition 2.7, we have that (b_2,c_2) in $\text{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_2)$. Then by the Left Residual Forth condition, there exists c_1 in V_1 such that $(b_1,c_1,b_2,c_2) \in W_{k-1}$ and either $(a_1,c_1) \notin \text{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_1)$ or $(a_1,c_1,a_2,c_2) \in Z_{k-1}$. By induction, we have $(b_1,c_1) \in e_2(\mathcal{G}_1)$. Hence, since $(a_1,b_1) \in e_1/e_2(\mathcal{G}_1)$, we have $(a_1,c_1) \in e_1(\mathcal{G}_1)$ so $(a_1,c_1) \in \text{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_1)$. Thus, the above qualification $(a_1,c_1) \notin \text{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_1)$ is redundant, and $(a_1,c_1,a_2,c_2) \in Z_{k-1}$. Again applying the induction hypothesis we obtain $(a_2,c_2) \in e_1(\mathcal{G}_2)$ as desired.

The case of a right residual is completely analogous to that of a left residual. ∎

In order to prove the simulation-analogue of the Characteristic Expression Lemma, we now present the maximal simulation in analogy to Definition 4.5, and state its properties.

DEFINITION 5.6
Given a fragment $\mathcal{F}$ as above and structures $\mathcal{G}_1$ and $\mathcal{G}_2$ with node sets V_1 and V_2 respectively, we construct two infinite decreasing sequences $Z_0, Z_1, Z_2, \cdots$ and $W_0, W_1, W_2, \cdots$ by induction on k as follows.

1. Z_0 is the set of all elements of $V_1^2 \times V_2^2$ that have the Atoms Forth property relative to $\mathcal{F}$, $\mathcal{G}_1$ and $\mathcal{G}_2$.
2. W_0 is the set of all elements of $V_1^2 \times V_2^2$ that have the Atoms Back property (still relative to $\mathcal{F}$, $\mathcal{G}_1$ and $\mathcal{G}_2$).
3. Z_i, for $i > 0$, is the set of all elements in Z_{i-1} that have

    - the Composition Forth property at degree i with respect to Z_{i-1};
    - the Projection Forth property at degree i with respect to Z_{i-1}, if $\mathcal{F}$ contains projection;
    - the Coprojection Forth property at degree i with respect to Z_{i-1}, W_{i-1}, if $\mathcal{F}$ contains coprojection;
    - the Left (Right) Residual Forth property at degree i with respect to Z_{i-1}, W_{i-1}, if $\mathcal{F}$ contains left (right) residual.

4. Similarly, W_i, for $i > 0$, is the set of all elements in W_{i-1} that have

    - the Composition Back property at degree i with respect to W_{i-1};
    - the Projection Back property at degree i with respect to W_{i-1}, if $\mathcal{F}$ contains projection;
    - the Coprojection Back property at degree i with respect to Z_{i-1}, W_{i-1}, if $\mathcal{F}$ contains coprojection;
    - the Left (Right) Residual Back property at degree i with respect to Z_{i-1}, W_{i-1}, if $\mathcal{F}$ contains left (right) residual.

We denote the constructed sequence $Z_0, Z_1, \ldots$ by $\mathrm{Sim}_{\mathrm{forth}}^{\mathcal{F}}(\mathcal{G}_1, \mathcal{G}_2)$, and $W_0, W_1, \ldots$ by $\mathrm{Sim}_{\mathrm{back}}^{\mathcal{F}}(\mathcal{G}_1, \mathcal{G}_2)$.

PROPOSITION 5.7
Let $\mathrm{Sim}_{\mathrm{forth}}^{\mathcal{F}}(\mathcal{G}_1, \mathcal{G}_2) = Z_0, Z_1, \ldots$ and $\mathrm{Sim}_{\mathrm{back}}^{\mathcal{F}}(\mathcal{G}_1, \mathcal{G}_2) = W_0, W_1, \ldots$. Then for each natural number k, the pair of sequences $(Z_0, Z_1, \ldots, Z_k; W_0, W_1, \ldots, W_k)$ is an $(\mathcal{F}, k)$-simulation from $\mathcal{G}_1$ to $\mathcal{G}_2$. Furthermore, it is the maximal simulation in the sense that, for any other such $(\mathcal{F}, k)$-bisimulation $(Z_0', \ldots, Z_k'; W_0', \ldots, W_k')$, we have $Z_i' \subseteq Z_i$ and $W_i' \subseteq W_i$ for each $i = 0, \ldots, k$.

COROLLARY 5.8
$(\mathcal{G}_1, a_1, b_1) \preceq_k^{\mathcal{F}} (\mathcal{G}_2, a_2, b_2) \Leftrightarrow (a_1, b_1, a_2, b_2) \in \mathrm{Sim}_{\mathrm{forth}}^{\mathcal{F}}(\mathcal{G}_1, \mathcal{G}_2)_k$.

PROPOSITION 5.9 (Path Preservation)
Let $\mathrm{Sim}_{\mathrm{forth}}^{\mathcal{F}}(\mathcal{G}_1, \mathcal{G}_2) = Z_0, Z_1, \ldots$ and $\mathrm{Sim}_{\mathrm{back}}^{\mathcal{F}}(\mathcal{G}_1, \mathcal{G}_2) = W_0, W_1, \ldots$, and let i be a natural number. If $(a_1, b_1, a_2, b_2) \in Z_i$ and $(a_1, b_1) \in \mathrm{paths}_i^{\mathcal{F}}(\mathcal{G}_1)$, then also $(a_2, b_2) \in \mathrm{paths}_i^{\mathcal{F}}(\mathcal{G}_2)$. Similarly, if $(a_1, b_1, a_2, b_2) \in W_i$ and $(a_2, b_2) \in \mathrm{paths}_i^{\mathcal{F}}(\mathcal{G}_2)$, then also $(a_1, b_1) \in \mathrm{paths}_i^{\mathcal{F}}(\mathcal{G}_1)$.

We are now ready for:

LEMMA 5.10 (Characteristic Expression)
Let k be a natural number and let $\overline{\mathcal{G}}_1 = (\mathcal{G}_1, a_1, b_1)$ be a marked structure. Then there exists an expression $e_{\overline{\mathcal{G}}_1}^{\mathcal{F},k}$ in $\mathcal{C}(\mathcal{F})_k$ such that for every structure $\mathcal{G}_2$ we have

$$e_{\overline{\mathcal{G}}_1}^{\mathcal{F},k}(\mathcal{G}_2) = \{(a_2, b_2) \in \mathrm{paths}_k^{\mathcal{F}}(\mathcal{G}_2) \mid \overline{\mathcal{G}}_1 \preceq_k^{\mathcal{F}} (\mathcal{G}_2, a_2, b_2)\}.$$

PROOF. Analogous to the proof of Lemma 4.9, our approach is now based on Corollary 5.8 and will show that each of the properties P involved in Definition 5.6 is expressible in $\mathcal{C}(\mathcal{F})_k$. Since these properties are now with respect to two sets Z and W, and since the expressions cannot use complement or difference, we must adapt the approach from the proof of Lemma 4.9 as follows. Let V_1 be the node set of the fixed structure $\mathcal{G}_1$, and let V_2 be the node set of any structure $\mathcal{G}_2$ to which our expressions will be applied. As in the proof of Lemma 4.9, we represent the first set $Z \subseteq V_1^2 \times V_2^2$ by the family of binary relations consisting of, for each $(a,b) \in V_1^2$, the binary relation

$$Z_{a,b} = \{(a', b') \in \mathrm{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_2) \mid (a,b,a',b') \in Z\}.$$

The second set $W \subseteq V_1^2 \times V_2^2$, however, is represented in a complementary manner. Specifically, we represent W by the family of binary relations consisting of, for each $(a,b) \in V_1^2$, the binary relation

$$\underline{W}_{a,b} = \{(a',b') \in \mathrm{paths}_{k-1}^{\mathcal{F}}(\mathcal{G}_2) \mid (a,b,a',b') \notin W\}.$$

The underscore is used to remind us that the relations $\underline{W}$ give us the complement of W.

As before we can add all these relations to $\mathcal{G}_2$, yielding an expanded structure, denoted by $(\mathcal{G}_2, Z, W)$, over the expansion of the given vocabulary Λ with relation names $Z_{a,b}$ and $\underline{W}_{a,b}$ for each $(a,b) \in V_1^2$. Furthermore, for our purpose, it will be sufficient to assume that Z and W are path-preserving in the sense of Proposition 5.9. Now we are going to express each property P by an expression $\psi_{\overline{\mathcal{G}}_1, P}^{\mathcal{F},k}$ over the expanded vocabulary, in the following sense.

- If P is a Forth property, then for any $\mathcal{G}_2$ and any Z and W as above, $\psi_{\overline{\mathcal{G}}_1, P}^{\mathcal{F},k}$ applied to $(\mathcal{G}_2, Z, W)$ returns the set of pairs $(a_2, b_2) \in \mathrm{paths}_k^{\mathcal{F}}(\mathcal{G}_2)$ for which (a_1, b_1, a_2, b_2) satisfies the P property at degree k with respect to Z, relative to $\mathcal{G}_1$, $\mathcal{G}_2$, and $\mathcal{F}$.
- Complementarily, if P is a Back property, then $\psi_{\overline{\mathcal{G}}_1, P}^{\mathcal{F},k}$ applied to $(\mathcal{G}_2, Z, W)$ returns the set of pairs $(a_2, b_2) \in \mathrm{paths}_k^{\mathcal{F}}(\mathcal{G}_2)$ for which (a_1, b_1, a_2, b_2) does *not* satisfy the P property.

It turns out that the required expressions $\psi_{\overline{\mathcal{G}}_1, P}^{\mathcal{F},k}$ can be deduced easily from the expressions $\varphi_{\overline{\mathcal{G}}_1, P}^{\mathcal{F},k}$ given in the proof of Lemma 4.9. Indeed, in all these expressions, it turns out that whenever in some property a Z must be changed to a W due to the new modified properties that cross over between Z and W, we already used subexpressions of the form $\mathrm{paths}_{k-1}^{\mathcal{F}} - Z_{a,b}$ in the right places. Hence, it suffices to replace these subexpressions by $\underline{W}_{a,b}$ to obtain $\psi_{\overline{\mathcal{G}}_1, P}^{\mathcal{F},k}$ from $\varphi_{\overline{\mathcal{G}}_1, P}^{\mathcal{F},k}$. Moreover, the expressions for the Back properties already were expressed as complements relative to $\mathrm{paths}_k^{\mathcal{F}}$; there we can simply

keep the parts on the right-hande side of the set difference operator. The correctness proofs are then completely analogous. Specifically, the required expressions are as follows.

$$\psi^{\mathcal{F},k}_{\overline{\mathcal{G}}_1,\text{composition forth}} := \text{paths}^{\mathcal{F}}_k \cap \bigcap_{\substack{c_1 \in V_1 \\ (a_1,c_1) \in \text{paths}^{\mathcal{F}}_{k-1}(\mathcal{G}_1) \\ (c_1,b_1) \in \text{paths}^{\mathcal{F}}_{k-1}(\mathcal{G}_1)}} Z_{a_1,c_1} \circ Z_{c_1,b_1}$$

$$\psi^{\mathcal{F},k}_{\overline{\mathcal{G}}_1,\text{composition back}} := \text{paths}^{\mathcal{F}}_k \cap \bigcup_{V \subseteq V_1} \left(\left(\bigcap_{c_1 \in V} \underline{W}_{a_1,c_1} \right) \circ \left(\bigcap_{c_1 \in V_1 - V} \underline{W}_{c_1,b_1} \right) \right)$$

$$\psi^{\mathcal{F},k}_{\overline{\mathcal{G}}_1,\text{projection forth}} := 1'$$
$$\cap \bigcap_{\substack{c_1 \in V_1 \\ (a_1,c_1) \in \text{paths}^{\mathcal{F}}_{k-1}(\mathcal{G}_1)}} \pi_1(Z_{a_1,c_1}) \cap \bigcap_{\substack{c_1 \in V_1 \\ (c_1,a_1) \in \text{paths}^{\mathcal{F}}_{k-1}(\mathcal{G}_1)}} \pi_2(Z_{c_1,a_1})$$

$$\psi^{\mathcal{F},k}_{\overline{\mathcal{G}}_1,\text{projection back}} := 1' \cap \left(\pi_1 \left(\bigcap_{c_1 \in V_1} \underline{W}_{a_1,c_1} \right) \cup \pi_2 \left(\bigcap_{c_1 \in V_1} \underline{W}_{c_1,a_1} \right) \right)$$

$$\psi^{\mathcal{F},k}_{\overline{\mathcal{G}}_1,\text{coproj forth}} := \bar{\pi}_1 \left(\text{paths}^{\mathcal{F}}_{k-1} \cap \bigcap_{c_1 \in V_1} \underline{W}_{a_1,c_1} \right) \cap \bar{\pi}_2 \left(\text{paths}^{\mathcal{F}}_{k-1} \cap \bigcap_{c_1 \in V_1} \underline{W}_{c_1,a_1} \right)$$

$$\psi^{\mathcal{F},k}_{\overline{\mathcal{G}}_1,\text{coproj back}} := \bar{\pi}_1 \left(\bigcup_{c_1 \in V_1} Z_{a_1,c_1} \right) \cup \bar{\pi}_2 \left(\bigcup_{c_1 \in V_1} Z_{c_1,a_1} \right)$$

$$\psi^{\mathcal{F},k}_{\overline{\mathcal{G}}_1,\text{leftres forth}} := \text{paths}^{\mathcal{F}}_k$$
$$\cap \bigcap_{V \subseteq V_1} \left[\left(\bigcup_{c_1 \in V_1 - V} Z_{a_1,c_1} \right) / \left(\text{paths}^{\mathcal{F}}_{k-1} \cap \bigcap_{\substack{c_1 \in V_1 \\ (a_1,c_1) \notin \text{paths}^{\mathcal{F}}_{k-1}(\mathcal{G}_1)}} \underline{W}_{b_1,c_1} \cap \bigcup_{c_1 \in V} \underline{W}_{b_1,c_1} \right) \right]$$

$$\psi^{\mathcal{F},k}_{\overline{\mathcal{G}}_1,\text{rightres forth}} := \text{paths}^{\mathcal{F}}_k$$
$$\cap \bigcap_{V \subseteq V_1} \left[\left(\text{paths}^{\mathcal{F}}_{k-1} \cap \bigcap_{\substack{c_1 \in V_1 \\ (c_1,b_1) \notin \text{paths}^{\mathcal{F}}_{k-1}(\mathcal{G}_1)}} \underline{W}_{c_1,a_1} \cap \bigcup_{c_1 \in V} \underline{W}_{c_1,a_1} \right) \setminus \bigcup_{c_1 \in V_1 - V} Z_{c_1,b_1} \right]$$

$$\psi^{\mathcal{F},k}_{\overline{\mathcal{G}}_1,\text{leftres back}} := \text{paths}^{\mathcal{F}}_k \cap \bigcup_{\substack{c_1 \in V_1 \\ (b_1,c_1) \in \text{paths}^{\mathcal{F}}_{k-1}(\mathcal{G}_1)}} \underline{W}_{a_1,c_1} / Z_{b_1,c_1}$$

$$\psi^{\mathcal{F},k}_{\overline{\mathcal{G}}_1,\text{rightres back}} := \text{paths}^{\mathcal{F}}_k \cap \bigcup_{\substack{c_1 \in V_1 \\ (c_1,a_1) \in \text{paths}^{\mathcal{F}}_{k-1}(\mathcal{G}_1)}} Z_{c_1,a_1} \setminus \underline{W}_{c_1,b_1}$$

We are now ready to present the construction of the required expression $e^{\mathcal{F},k}_{\overline{\mathcal{G}}_1}$, by induction on k. Actually, we will simultaneously construct an expression $e'^{\mathcal{F},k}_{\overline{\mathcal{G}}_1}$ with the property that for every $\mathcal{G}_2$

we have

$$e'^{\mathcal{F},k}_{\overline{\mathcal{G}}_1}(\mathcal{G}_2) = \{(a_2,b_2) \in \mathrm{paths}^{\mathcal{F}}_k(\mathcal{G}_2) \mid (a_1,b_1,a_2,b_2) \notin \mathrm{Sim}^{\mathcal{F}}_{\mathrm{back}}(\mathcal{G}_1,\mathcal{G}_2)_k\},$$

thus complementing expression $e^{\mathcal{F},k}_{\overline{\mathcal{G}}_1}$ which must satisfy

$$e^{\mathcal{F},k}_{\overline{\mathcal{G}}_1}(\mathcal{G}_2) = \{(a_2,b_2) \in \mathrm{paths}^{\mathcal{F}}_k(\mathcal{G}_2) \mid (a_1,b_1,a_2,b_2) \in \mathrm{Sim}^{\mathcal{F}}_{\mathrm{forth}}(\mathcal{G}_1,\mathcal{G}_2)_k\}.$$

For the base of the construction, we put

$$e^{\mathcal{F},0}_{\overline{\mathcal{G}}_1} := \mathrm{paths}^{\mathcal{F}}_0 \cap \bigcap_{e \in \mathrm{atp}^{\mathcal{F}}(\overline{\mathcal{G}}_1)} e;$$

$$e'^{\mathcal{F},0}_{\overline{\mathcal{G}}_1} := \mathrm{paths}^{\mathcal{F}}_0 \cap \bigcup_{e \notin \mathrm{atp}^{\mathcal{F}}(\overline{\mathcal{G}}_1)} e.$$

For $k > 0$, expression $e^{\mathcal{F},k}_{\overline{\mathcal{G}}_1}$, applied to any $\mathcal{G}_2$, should return the set of all pairs $(a_2,b_2) \in \mathrm{paths}^{\mathcal{F}}_k(\mathcal{G}_2)$ such that $(a_1,b_1,a_2,b_2) \in \mathrm{Sim}^{\mathcal{F}}_{\mathrm{forth}}(\mathcal{G}_1,\mathcal{G}_2)_k$. By definition, these are the pairs (a_2,b_2) for which $(a_1,b_1,a_2,b_2) \in \mathrm{Sim}^{\mathcal{F}}_{\mathrm{forth}}(\mathcal{G}_1,\mathcal{G}_2)_{k-1}$, and such that (a_1,b_1,a_2,b_2) has all the Forth properties required by Definition 5.6, at degree k with respect to $\mathrm{Sim}^{\mathcal{F}}_{\mathrm{forth}}(\mathcal{G}_1,\mathcal{G}_2)_{k-1}, \mathrm{Sim}^{\mathcal{F}}_{\mathrm{back}}(\mathcal{G}_1,\mathcal{G}_2)_{k-1}$. We have just seen that each Forth property P is expressible by $\psi^{\mathcal{F},k}_{\overline{\mathcal{G}}_1,P}$. Hence, we can obtain $e^{\mathcal{F},k}_{\overline{\mathcal{G}}_1}$ simply as the intersection of $e^{\mathcal{F},k-1}_{\overline{\mathcal{G}}_1}$ (obtained by induction) and the expressions $\psi^{\mathcal{F},k}_{\overline{\mathcal{G}}_1,P}$ for the different Forth properties P required by Definition 5.6.

Complementarily, expression $e'^{\mathcal{F},k}_{\overline{\mathcal{G}}_1}$ should return the set of all pairs $(a_2,b_2) \in \mathrm{paths}^{\mathcal{F}}_k(\mathcal{G}_2)$ such that $(a_1,b_1,a_2,b_2) \notin \mathrm{Sim}^{\mathcal{F}}_{\mathrm{back}}(\mathcal{G}_1,\mathcal{G}_2)_k$. This means that (a_1,b_1,a_2,b_2) must *not* satisfy at least one of the Back properties required by Definition 5.6, at degree k with respect to $\mathrm{Sim}^{\mathcal{F}}_{\mathrm{forth}}(\mathcal{G}_1,\mathcal{G}_2)_{k-1}, \mathrm{Sim}^{\mathcal{F}}_{\mathrm{back}}(\mathcal{G}_1,\mathcal{G}_2)_{k-1}$. We have just seen that the complement of each Back property P is expressible by $\psi^{\mathcal{F},k}_{\overline{\mathcal{G}}_1,P}$. Hence, we can obtain $e^{\mathcal{F},k}_{\overline{\mathcal{G}}_1}$ simply as the intersection of $\mathrm{paths}^{\mathcal{F}}_k$ with the union of the expressions $\psi^{\mathcal{F},k}_{\overline{\mathcal{G}}_1,P}$ for the different Back properties P required by Definition 5.6.

In analogy to the proof of Lemma 4.9, in the resulting expressions, we replace each relation name $Z_{a,b}$ by $e^{\mathcal{F},k-1}_{\mathcal{G}_1,a,b}$; furthermore, we replace each relation name $\underline{W}_{a,b}$ by $e'^{\mathcal{F},k-1}_{\mathcal{G}_1,a,b}$. The reduction to a finitary expression, based on quantifier rank, is exactly as in the proof of Lemma 4.9, and we are done.

∎

6   Indistinguishability of finite structures

The bisimilarity characterizations we have given of when two structures are indistinguishable by expressions of $\mathcal{C}(\mathcal{F})_k$, for some fixed degree k and some fixed fragment $\mathcal{F}$, are valid for arbitrary structures. For finite structures, by classical arguments [8, 17, 21], our methods lead immediately to Hennessy–Milner-style theorems about indistinguishability in the full fragment $\mathcal{C}(\mathcal{F})$, without a degree restriction, as we will show in the present section. It also follows that indistinguishability in $\mathcal{C}(\mathcal{F})$ is decidable in polynomial time.

6.1 Bisimulation without degree restriction

Let $\mathcal{C}(\mathcal{F})$ be a fragment containing complement or difference. We want to define a natural notion of $\mathcal{F}$-bisimulation without a degree restriction k. Thereto we must make two small adaptations.

1. Recalling Definition 2.3, let us define $\mathrm{paths}^{\mathcal{F}}(\mathcal{G})$ as the set of all pairs (x,y) in V^2 such that

   - there is a directed path from x to y in $graph(\mathcal{G})$, if $\mathcal{F}$ does not contain converse; or
   - there is an undirected path from x to y in $ugraph(\mathcal{G})$, if $\mathcal{F}$ contains converse.

2. Recall the suite of Forth and Back properties introduced in Section 4.1. We naturally introduce variants of these properties that are no longer degree-restricted. It suffices to replace each reference to $\mathrm{paths}^{\mathcal{F}}_i$ or $\mathrm{paths}^{\mathcal{F}}_{i-1}$ by $\mathrm{paths}^{\mathcal{F}}$, so that the degree parameter i becomes irrelevant.

We now define an *$\mathcal{F}$-bisimulation from $\mathcal{G}_1$ to $\mathcal{G}_2$* as a relation $Z \subseteq V_1^2 \times V_2^2$ such that each of its elements has the Atoms Forth and Back properties, as well as the (degree-unrestricted versions of the) Composition Forth and Back properties with respect to Z, and also the Projection (Left Residual, Right Residual) Forth and Back properties with respect to Z depending on whether $\mathcal{F}$ contains projection (left residual, right residual), as usual. When there exists an $\mathcal{F}$-bisimulation from $\mathcal{G}_1$ to $\mathcal{G}_2$ containing (a_1, b_1, a_2, b_2), we say that $\overline{\mathcal{G}}_1 = (\mathcal{G}_1, a_1, b_1)$ and $\overline{\mathcal{G}}_2 = (\mathcal{G}_2, a_2, b_2)$ are *$\mathcal{F}$-bisimilar* and denote this by $\mathcal{G}_1 \simeq^{\mathcal{F}} \mathcal{G}_2$.

We show:

THEOREM 6.1
For finite marked structures $\overline{\mathcal{G}}_1$ and $\overline{\mathcal{G}}_2$, we have $\overline{\mathcal{G}}_1 \simeq^{\mathcal{F}} \overline{\mathcal{G}}_2$ if and only if $\overline{\mathcal{G}}_1 \equiv^{\mathcal{F}} \overline{\mathcal{G}}_2$. (The only-if implication actually holds for all structures, finite or infinite.)

PROOF. Let $\overline{\mathcal{G}}_1 = (\mathcal{G}_1, a_1, b_1)$ and $\overline{\mathcal{G}}_2 = (\mathcal{G}_2, a_2, b_2)$, and assume $\overline{\mathcal{G}}_1 \simeq^{\mathcal{F}} \overline{\mathcal{G}}_2$ by the bisimulation Z. In order to show $\overline{\mathcal{G}}_1 \equiv^{\mathcal{F}} \overline{\mathcal{G}}_2$, we must show that $\overline{\mathcal{G}}_1 \equiv^{\mathcal{F}}_k \overline{\mathcal{G}}_2$ holds for all degrees k. So, let k be arbitrary. Consider the sequence $\bar{Z} = Z, Z, \ldots, Z$ that simply consists of $k+1$ times Z. It is now readily verified that $\bar{Z}$ is an $(\mathcal{F}, k)$-bisimulation from $\mathcal{G}_1$ to $\mathcal{G}_2$. Since $(a_1, b_1, a_2, b_2) \in Z$, we conclude $\overline{\mathcal{G}}_1 \equiv^{\mathcal{F}}_k \overline{\mathcal{G}}_2$ by Lemma 4.4.

Conversely, assume $\overline{\mathcal{G}}_1 \equiv^{\mathcal{F}}_k \overline{\mathcal{G}}_2$ for every degree k. This means that $(a_1, b_1, a_2, b_2) \in$ $\mathrm{BiSim}^{\mathcal{F}}(\mathcal{G}_1, \mathcal{G}_2)_k$ for every k. Let $\mathrm{BiSim}^{\mathcal{F}}(\mathcal{G}_1, \mathcal{G}_2) = Z_0, Z_1, \ldots$. Recall that this is a decreasing sequence. Hence, since $\mathcal{G}_1$ and $\mathcal{G}_2$ are finite, there exists a degree p such that $Z_{q+1} = Z_q$ for all $q \geq p$. As a consequence, Z_p is an $\mathcal{F}$-bisimulation from $\mathcal{G}_1$ to $\mathcal{G}_2$ and we are done. ∎

6.2 Simulation without degree restriction

For any fragment $\mathcal{C}(\mathcal{F})$ containing neither complement nor difference, in an entirely analogous manner we can define degree-unrestricted versions of the properties for simulations listed in Section 5.1. Then an *$\mathcal{F}$-simulation from $\mathcal{G}_1$ to $\mathcal{G}_2$* is a pair (Z, W) of subsets of $V_1^2 \times V_2^2$ such that each element of Z has all Forth properties with respect to (Z, W) corresponding to the operators present in $\mathcal{F}$, and each element of W has all Back properties with respect to (Z, W). We can then again show that $\overline{\mathcal{G}}_1 \Rightarrow^{\mathcal{F}} \overline{\mathcal{G}}_2$ if and only if there exists an $\mathcal{F}$-simulation (Z, W) from $\mathcal{G}_1$ to $\mathcal{G}_2$ such that $(a_1, b_1, a_2, b_2) \in Z$.

6.3   Polynomial-time complexity

As a corollary of the above, we obtain:

COROLLARY 6.2
For any fixed fragment $\mathcal{C}(\mathcal{F})$, it can be decided in polynomial time whether or not two given finite marked structures are indistinguishable in $\mathcal{C}(\mathcal{F})$.

PROOF. We give the proof for fragments with complement or difference; the case for the other fragments is entirely analogous. Let $\overline{\mathcal{G}}_1 = (\mathcal{G}_1, a_1, b_1)$ and $\overline{\mathcal{G}}_2 = (\mathcal{G}_2, a_2, b_2)$ be finite marked structures. Let n_1 (n_2) be the number of nodes of $\mathcal{G}_1$ ($\mathcal{G}_2$). Let $\mathrm{BiSim}^{\mathcal{F}}(\mathcal{G}_1, \mathcal{G}_2) = Z_0, Z_1, \ldots$. Recall from the proof of Theorem 6.1 that there exists p such that $Z_{q+1} = Z_q$ for all $q \geq p$, and $\overline{\mathcal{G}}_1 \equiv^{\mathcal{F}} \overline{\mathcal{G}}_2$ if and only if $(a_1, b_1, a_2, b_2) \in Z_p$. Let p be the smallest such p. Then in the worst case, for each $i < p$, there is exactly one element less in Z_{i+1} compared to Z_i. Hence, $p \leq n_1^2 n_2^2$ which is polynomial in n_1 and n_2. Moreover, it is evident from Definition 4.5 that Z_{i+1} can be computed from Z_i in time polynomial in n_1, n_2, and the size of Z_i which is itself bounded by $n_1^2 n_2^2$. Hence, we can compute Z_p by performing a polynomial number of iterations where each iteration takes polynomial time, and we are done. ∎

7   Concluding remarks

We have always included the identity relation and the three operations union, intersection and composition in the logics that we consider. As already mentioned in the Introduction, it is an interesting topic for further research to see what happens if some of these operators are left out.

The results of the present paper provide the tools to continue the research on the relative expressive power of fragments of the calculus of relations. In earlier work [11] the precise relationships between all fragments were clarified, ignoring residuals however. Since in the present paper we have fully integrated the residuals, an interesting direction for further research is now to throw the residuals in the picture.

For example, it is an intriguing question how the fragments $\mathcal{C}(-)$ and $\mathcal{C}(/, \backslash)$ relate to each other in their power to express boolean queries (where 'true' is represented by any non-empty answer and 'false' is represented by the empty answer). Both fragments extend the basic fragment $\mathcal{C}$ with significant nonmonotonic operators, viz., difference on the one hand and the residuals on the other hand. We actually conjecture that the two expressive powers are incomparable. For example, following the standard approach [2, 9, 24], one may try to prove that $\mathcal{C}(/, \backslash)$ is not subsumed by $\mathcal{C}(-)$ for boolean queries by exhibiting an expression e in $\mathcal{C}(/, \backslash)$ and, for each degree k, two structures $\mathcal{A}_k$ and $\mathcal{B}_k$ such that for each k the following holds:

- $e(\mathcal{A}_k) \neq \emptyset = e(\mathcal{B}_k)$;
- for any pair (a, b) of nodes of $\mathcal{A}_k$ there exists a pair (a', b') of nodes of $\mathcal{B}_k$ such that $(\mathcal{A}_k, a, b) \preceq_k^{\{/, \backslash\}} (\mathcal{B}_k, a', b')$.

This indeed implies that no expression e' in $\mathcal{C}(/, \backslash)$ can correctly express the boolean query expressed by e. For, because $e(\mathcal{A}_k) \neq \emptyset$, any such expression should return at least one pair (a, b) on $\mathcal{A}_k$, but then by the similarity relationship, it will also return a pair (a', b') on $\mathcal{B}_k$, in contradiction with $e(\mathcal{B}_k) = \emptyset$.

For fragments with set difference or complementation, we have worked with bisimulations; for fragments without set difference, we have worked with simulations. In modal logic there are results [8, Theorem 2.78] that turn this around, showing that formulas invariant under simulations are in

fact equivalent to positive-existential formulas. It is an interesting direction to see how such a result could be formulated in the setting considered in the present article.

Acknowledgments

References

[1] S. Abiteboul, P. Buneman, and D. Suciu. *Data on the Web: From Relations to Semistructured Data and XML*. Morgan Kaufmann, 1999.

[2] S. Abiteboul, R. Hull, and V. Vianu. *Foundations of Databases*. Addison-Wesley, 1995.

[3] L. Aceto, A. Ingolfsdottir, and J. Srba. The algorithmics of bisimilarity. In *Advanced Topics in Bisimulation and Coinduction*, chapter 3. D. Sangiorgi and J. Rutten, eds, Cambridge University Press, 2011.

[4] A. V. Aho, J. E. Hopcroft, and J. D. Ullman. *The Design and Analysis of Computer Algorithms*. Addison-Wesley, 1974.

[5] F. Baader, D. Calvanese, D. McGuiness, D. Nardi, and P. Patel-Schneider, editors. *The Description Logic Handbook*. Cambridge University Press, 2003.

[6] M. Benedikt, W. Fan, and G. Kuper. Structural properties of XPath fragments. *Theoretical Computer Science*, **336**, 3–31, 2005.

[7] C. Bizer, T. Heath, and T. Berners-Lee. Linked data - the story so far. *International Journal on Semantic Web and Information Systems*, **5**, 1–22, 2009.

[8] P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Cambridge University Press, 2001.

[9] H.-D. Ebbinghaus and J. Flum. *Finite Model Theory*. Springer, 1999.

[10] H.-D. Ebbinghaus, J. Flum, and W. Thomas. *Mathematical Logic*. Undergraduate Texts in Mathematics. Springer-Verlag, 1984.

[11] G. H. L. Fletcher, M. Gyssens, D. Leinders, J. Van den Bussche, D. Van Gucht, S. Vansummeren, and Y. Wu. Relative expressive power of navigational querying on graphs. In *Proceedings 14th International Conference on Database Theory*, 2011.

[12] G. H. L. Fletcher, M. Gyssens, D. Leinders, J. Van den Bussche, D. Van Gucht, S. Vansummeren, and Y. Wu. The impact of transitive closure on the expressiveness of navigational query languages on unlabeled graphs. *Annals of Mathematics and Artificial Intelligence*, 2013. Published online, 2 April.

[13] G. H. L. Fletcher, D. Van Gucht, Y. Wu, M. Gyssens, S. Brenes, and J. Paredaens. A methodology for coupling fragments of XPath with structural indexes for XML documents. *Information Systems*, **34**, 657–670, 2009.

[14] D. Florescu, A. Y. Levy, and A. O. Mendelzon. Database techniques for the World-Wide Web: A survey. *SIGMOD Record*, **27**, 59–74, 1998.

[15] M. J. Franklin, A. Halevy, and D. Maier. From databases to dataspaces: A new abstraction for information management. *SIGMOD Record*, **34**, 27–33, 2005.

[16] S. Givant. The calculus of relations as a foundation for mathematics. *Journal Autom. Reasoning*, **37**, 277–322, 2006.

[17] V. Goranko and M. Otto. Model theory of modal logic. In *Handbook of Modal Logic*, chapter 5. P. Blackburn, J. van Benthem, and F. Wolter, eds, Elsevier, 2007.

[18] M. Gyssens, J. Paredaens, D. Van Gucht, et al. Structural characterizations of the semantics of XPath as navigation tool on a document. In *Proceedings 25th ACM Symposium on Principles of Database Systems*, pp. 318–327. ACM Press, 2006.

[19] M. Gyssens, L. V. Saxton, and D. Van Gucht. Tagging as an alternative to object creation. In *Query Processing For Advanced Database Systems*, chapter 8. J.C. Freytag, D. Maier, and G. Vossen, eds, Morgan Kaufmann, 1994.

[20] D. Harel, D. Kozen, and J. Tiuryn. *Dynamic Logic*. MIT Press, 2000.

[21] M. Hennessy and R. Milner. Algebraic laws for nondeterminism and concurrency. *Journal of the ACM*, **32**, 137–161, 1985.

[22] R. Hirsch and I. Hodkinson. *Relation Algebras by Games*. Elsevier, 2002.

[23] N. Immerman and E. Lander. Describing graphs: a first-order approach to graph canonization. In *Complexity Theory Retrospective*, A. Selman, ed. pp. 59–81. Springer-Verlag, 1990.

[24] L. Libkin. *Elements of Finite Model Theory*. Springer, 2004.

[25] R. D. Maddux. The origin of relation algebras in the development and axiomatization of the calculus of relations. *Studia Logica*, **50**, 421–455, 1991.

[26] R. D. Maddux. *Relation Algebras*. Elsevier, 2006.

[27] N. Mamoulis. Efficient processing of joins on set-valued attributes. In *Proceedings ACM SIGMOD International Conference on Management of Data*, pp. 157–168, 2003.

[28] M. Marx. Conditional XPath. *ACM Transactions on Database Systems*, **30**, 929–959, 2005.

[29] M. Marx and M. de Rijke. Semantic characterizations of navigational XPath. *SIGMOD Record*, **34**, 41–46, 2005.

[30] M. Marx and Y. Venema. *Multi-Dimensional Modal Logic*. Springer, 1997.

[31] M. Marx and Y. Venema. Local variations on a loose theme: modal logic and decidability. In *Finite Model Theory and Its Applications*, chapter 7. Springer, 2007.

[32] F. Picalausa, Y. Luo, G.H.L. Fletcher, J. Hidders, and S. Vansummeren. A structural approach to indexing triples. In *Proceedings 9th Extended Semantic Web Conference*. Vol. 7295 of *Lecture Notes in Computer Science*, E. Simperl et al., eds, pp. 406–521. Springer, 2012.

[33] V. R. Pratt. Origins of the calculus of binary relations. In *Proceedings of the 7th IEEE Symposium on Logic in Computer Science*, pp. 248–254, 1992.

[34] RDF primer. W3C Recommendation, February 2004.

[35] V. M. Sarathy, L. V. Saxton, and D. Van Gucht. Algebraic foundation and optimization for object based query languages. In *Proceedings 9th International Conference on Data Engineering*, pp. 81–90. IEEE Computer Society, 1993.

[36] A. Tarski. On the calculus of relations. *Journal of Symbolic Logic*, **6**, 73–89, 1941.

[37] A. Tarski and S. Givant. *A Formalization of Set Theory Without Variables*, volume 41 of *Colloquium Publications*. American Mathematical Society, 1987.

[38] J. van Benthem. Program constructions that are safe for bisimulation. *Studia Logica*, **60**, 311–330, 1998.

[39] Y. Wu, D. Van Gucht, M. Gyssens, and J. Paredaens. A study of a positive fragment of path queries: expressiveness, normal form and minimization. *The Computer Journal*, **54**, 1091–1118, 2011.