Algorithms for Algebraic Path Properties in Concurrent Systems of Constant Treewidth Components

Krishnendu Chatterjee Amir Kafshdar Goharshady Rasmus Ibsen-Jensen Andreas Pavlogiannis

IST Austria (Institute of Science and Technology Austria) Klosterneuburg, Austria
{krish.chat, goharshady, ribsen, pavlogiannis}@ist.ac.at

Abstract

We study algorithmic questions for concurrent systems where the transitions are labeled from a complete, closed semiring, and path properties are algebraic with semiring operations. The algebraic path properties can model dataflow analysis problems, the shortest path problem, and many other natural problems that arise in program analysis. We consider that each component of the concurrent system is a graph with constant treewidth, a property satisfied by the controlflow graphs of most programs. We allow for multiple possible queries, which arise naturally in demand driven dataflow analysis. The study of multiple queries allows us to consider the tradeoff between the resource usage of the one-time preprocessing and for each individual query. The traditional approach constructs the product graph of all components and applies the best-known graph algorithm on the product. In this approach, even the answer to a single query requires the transitive closure (i.e., the results of all possible queries), which provides no room for tradeoff between preprocessing and query time.

Our main contributions are algorithms that significantly improve the worst-case running time of the traditional approach, and provide various tradeoffs depending on the number of queries. For example, in a concurrent system of two components, the traditional approach requires hexic time in the worst case for answering one query as well as computing the transitive closure, whereas we show that with one-time preprocessing in almost cubic time, each subsequent query can be answered in at most linear time, and even the transitive closure can be computed in almost quartic time. Furthermore, we establish conditional optimality results showing that the worst-case running time of our algorithms cannot be improved without achieving major breakthroughs in graph algorithms (i.e., improving the worst-case bound for the shortest path problem in general graphs). Preliminary experimental results show that our algorithms perform favorably on several real-world benchmarks.

Categories and Subject Descriptors F.3.2 [Logics and Meanings of Programs]: Semantics of Programming Languages—Program Analysis

Keywords Concurrent systems, Constant-treewidth graphs, Algebraic path properties, Shortest path.

1. Introduction

In this work we consider concurrent finite-state systems where each component is a constant-treewidth graph, and the algorithmic question is to determine algebraic path properties between pairs of nodes in the system. Our main contributions are algorithms which significantly improve the worst-case running time of the existing algorithms. We establish conditional optimality results for some of our algorithms in the sense that they cannot be improved without achieving major breakthroughs in the algorithmic study of graph problems. Finally, we provide a prototype implementation of our algorithms which significantly outperforms the existing algorithmic methods on several benchmarks.

Concurrency and algorithmic approaches. The analysis of concurrent systems is one of the fundamental problems in computer science in general, and programming languages in particular. A finite-state concurrent system consists of several components, each of which is a finite-state graph, and the whole system is a composition of the components. Since errors in concurrent systems are hard to reproduce by simulations due to combinatorial explosion in the number of interleavings, formal methods are necessary to analyze such systems. In the heart of the formal approaches are graph algorithms, which provide the basic search procedures for the problem. The basic graph algorithmic approach is to construct the product graph (i.e., the product of the component systems) and then apply the best-known graph algorithms on the product graph. While there are many practical approaches for the analysis of concurrent systems, a fundamental theoretical question is whether special properties of graphs that arise in analysis of programs can be exploited to develop asymptotically faster algorithms as compared to the basic approach.

Special graph properties for programs. A very well-studied notion in graph theory is the concept of treewidth of a graph, which is a measure of how similar a graph is to a tree (a graph has treewidth 1 precisely if it is a tree) [67]. The treewidth of a graph is defined based on a tree decomposition of the graph [40], see Section 2 for a formal definition. On one hand the treewidth property provides a mathematically elegant way to study graphs, and on the other hand there are many classes of graphs which arise in practice and have constant treewidth. The most important example is that the controlflow graph for goto-free programs for many programming languages are of constant treewidth [70], and it was also shown in [39] that typically all Java programs have constant treewidth.

Algebraic path properties. To specify properties of traces of concurrent systems we consider a very general framework, where edges of the system are labeled from a complete, closed semiring (which subsumes bounded and finite distributive semirings), and we refer to the labels of the edges as weights. For a given path, the weight of the path is the semiring product of the weights on the edges of the path, and the weights of different paths are combined using the semiring plus operator. For example, (i) the Boolean semiring (with semiring product as AND, and semiring plus as OR) expresses the reachability property; (ii) the tropical semiring (with real numbers as edge weights, semiring product as standard sum, and semiring plus as minimum) expresses the shortest path property; and (iii) with letter labels on edges, semiring product as string concatenation and semiring plus as union we can express the regu-
lar expression of reaching from one node to another. The algebraic path properties subsume the dataflow analysis of the IFDS/IDE frameworks [66, 68] in the intraprocedural setting, which consider compositions of distributive dataflow functions, and meet-over-all-paths as the semiring plus operator. Since IFDS/IDE is a special case of our framework, a large and important class of dataflow analysis problems that can be expressed in IFDS/IDE can also be expressed in our framework. However, the IFDS/IDE framework works for sequential interprocedural analysis, whereas we focus on intraprocedural analysis, but in the concurrent setting.

Expressiveness of algebraic path properties. The algebraic path properties provide an expressive framework with rich modeling power. Here we elaborate on three important classes.

1. Weighted shortest path. The algebraic paths framework subsumes several problems on weighted graphs. The most well-known such problem is the shortest path problem [6, 36, 37, 45, 74], phrased on the tropical semiring. For example, the edge weights (positive and negative) can express energy consumptions, and the shortest path problem asks for the least energy consuming path. Another important quantitative property is the mean-payoff property, where each edge weight represents a reward or cost, and the problem asks for a path that minimizes the average of the weights along a path. Many quantitative properties of relevance for program analysis (e.g., to express performance or resource consumption) can be modeled as mean-payoff properties [18, 25]. The mean-payoff and other fundamental problems on weighted graphs (e.g., the most probable path and the minimum initial credit problem) can be reduced to the shortest-path problem [16, 18, 21, 23, 48, 57, 73, 75].

2. Dataflow problems. A wide range of dataflow problems has an algebraic paths formulation, expressed as a “meet-over-all-paths” analysis [49]. Perhaps the most well-known case is that of distributive flow functions considered in the IFDS framework [66, 68]. Given a finite domain $D$ and a universe $F$ of distributive dataflow functions $f : 2^D \rightarrow 2^D$, a weight function $w : E \rightarrow F$ associates each edge of the controlflow graph with a flow function. The weight of a path is then defined as the composition of the flow functions along its edges, and the dataflow distance between two nodes $u, v$ is the meet $\sqcap$ (union or intersection) of the weights of all $u \rightarrow v$ paths. The problem can be formulated on the meet-composition semiring $(F, \sqcap, \circ, \emptyset, I)$, where $I$ is the identity function. We note, however, that the IFDS/IDE framework considers interprocedural paths in sequential programs. In contrast, the current work focuses on intraprocedural analysis of concurrent programs. The dataflow analysis of concurrent programs has been a problem of intensive study (e.g., [27, 29, 33, 38, 46, 51]), where (part of) the underlying analysis is based on an algebraic, “meet-over-all-paths” approach.

3. Regular expressions. Consider the case that each edge is annotated with an observation or action. Then the regular expression to reach from one node to another represents all the sequences of observable actions that lead from the start node to the target. The regular languages of observable actions have provided useful formulations in the analysis and synthesis of concurrent systems [19, 31, 34]. Regular expressions have also been used as algebraic relaxations of interprocedurally valid paths in sequential and concurrent systems [14, 76].

For a detailed discussion, see Appendix A.

The algorithmic problem. In graph theoretic parlance, graph algorithms typically consider two types of queries: (i) a pair query given nodes $u$ and $v$ (called $(u, v)$-pair query) asks for the algebraic path property from $u$ to $v$; and (ii) a single-source query given a node $u$ asks for the answer of $(u, v)$-pair queries for all nodes $v$. In the context of concurrency, in addition to the classical pair and single-source queries, we also consider partial queries. Given a concurrent system with $k$ components, a node in the product graph is a tuple of $k$ component nodes. A partial node $\pi$ in the product only specifies nodes of a nonempty strict subset of all the components. Our work also considers partial pair and partial single-source queries, where the input nodes are partial nodes. Queries on partial nodes are very natural, as they capture properties between local locations in a component, that are shaped by global paths in the whole concurrent system. For example, constant propagation and dead code elimination are local properties in a program, but their analysis requires analyzing the concurrent system as a whole.

Preprocess vs query. A topic of widespread interest in the programming languages community is that of on-demand analysis [5, 24, 30, 44, 60, 64, 65, 68, 77, 78]. Such analysis has several advantages, such as (quoting from [44, 65]) (i) narrowing down the focus to specific points of interest, (ii) narrowing down the focus to specific dataflow facts of interest, (iii) reducing work in preliminary phases, (iv) sidestepping incremental updating problems, and (v) offering demand analysis as a user-level operation. For example, in alias analysis, the question is whether two pointers may point to the same object, which is by definition modeled as a question between a pair of nodes. Similarly, in constant propagation a relevant question is whether some variable remains constant between a pair of controlflow locations. The problem of on-demand analysis allows us to distinguish between a single preprocessing phase (one time computation), and a subsequent query phase, where queries are answered on demand. The two extremes of the preprocessing and query phase are: (i) complete preprocessing (aka transitive closure computation) where the result is precomputed for every possible query, and hence queries are answered by simple table lookup; and (ii) no preprocessing where every query requires a new computation. However, in general, there can be a tradeoff between the preprocessing and query computation. Most of the existing works for on-demand analysis do not make a formal distinction between preprocessing and query phases, as the provided complexities only guarantee the same worst-case complexity property, namely that the total time for handling any sequence of queries is no worse than the complete preprocessing. Hence most existing tradeoffs are practical, without any theoretical guarantees.

Previous results. In this work we consider finite-state concurrent systems, where each component graph has constant treewidth, and the trace properties are specified as algebraic path properties. Our framework can model a large class of problems: typically the controlflow graphs of programs have constant treewidth [17, 39, 70], and if there is a constant number of synchronization variables with constant-size domains, then each component graph has constant treewidth. Note that this imposes little practical restrictions, as typically synchronization variables, such as locks, mutexes and condition variables have small (even binary) domains (e.g. locked/unlocked state). The best-known graph algorithm for the algebraic path property problem is the classical Warshall-Floyd-Kleeene [36, 50, 58, 74] style dynamic programming, which requires cubic time. Two well-known special cases of the algebraic paths problem are (i) computing the shortest path from a source to a target node in a weighted graph, and (ii) computing the regular expression from a source to a target node in an automaton whose edges are labeled with letters from a finite alphabet. In the first case, the best-known algorithm is the Bellman-Ford algorithm with time complexity $O(n \cdot m)$. In the second case, the well-known construction of Kleeene’s [50] theorem requires cubic time. The only existing algorithmic approach for the problem we consider is to first construct the product graph (thus if each component graph has...
sures (time and space) are in the number of basic machine opera-

Hence the existing approach does not allow a tradeoff of prepro-

Hence for the important special case of two components we obtain

Table 1: The algorithmic complexity for computing algebraic path queries wrt a closed, complete semiring on a concurrent graph

<table>
<thead>
<tr>
<th>Preprocess</th>
<th>Query time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>Space</td>
</tr>
<tr>
<td>Previous results [36, 50, 58, 74]</td>
<td>(O(n^6))</td>
</tr>
<tr>
<td>Our result Corollary 2 ((ε &gt; 0))</td>
<td>(O(n^3))</td>
</tr>
<tr>
<td>Our result Theorem 3 ((ε &gt; 0))</td>
<td>(O(n^{3κ}))</td>
</tr>
<tr>
<td>Our result Corollary 3 ((ε &gt; 0))</td>
<td>(O(n^{3κ+1}))</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1. Improved upper bounds.</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General case.</strong> We show that for (k \geq 3) components with (n) nodes each, after (O(n^{3(k−1)})) preprocessing time, we can answer (i) single-source queries in (O(n^{2(k−1)})) time, (ii) pair queries in (O(n^{k−1})) time, (iii) partial single-source queries in (O(n^k)) time, and (iv) partial pair queries in (O(1)) time; while using at all times (O(n^{2k−1})) space. In contrast, the existing methods [36, 50, 58, 74] compute the transitive closure even for a single query, and thus require (O(n^{3k})) time and (O(n^{4k})) space.</td>
<td></td>
</tr>
<tr>
<td><strong>Two components.</strong> For the important case of two components, the existing methods require (O(n^3)) time and (O(n^5)) space even for one query. In contrast, we establish a variety of tradeoffs between preprocessing and query times, and the best choice depends on the number of expected queries. In particular, for any fixed (ε &gt; 0), we establish the following three results.</td>
<td></td>
</tr>
<tr>
<td>Three results. First, we show (Corollary 2) that with (O(n^3)) preprocessing time and using (O(n^{2+κ})) space, we can answer single-source queries in (O(n^{2+κ})) time, and pair and partial pair queries require (O(n^2)) time. Second, we show (Theorem 3) that with (O(n^{3κ})) preprocessing time and using (O(n^3)) space, we can answer pair and partial pair queries in time (O(n)) and (O(1)), respectively. Third, we show (Corollary 3) that the transitive closure can be computed using (O(n^{1+κ})) preprocessing time and (O(n^3)) space, after which single-source queries require (O(n^2)) time, and pair and partial pair queries require (O(1)) time (i.e., all queries require linear time in the size of the output).</td>
<td></td>
</tr>
<tr>
<td>Tradeoffs. Our results provide various tradeoffs: The first result is best for answering (O(n^{1+κ})) pair and partial pair queries; the second result is best for answering between (Ω(n^{1+κ})) and (O(n^{3κ+1})) pair queries, and (Ω(n^{1+κ})) partial</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1: Given a concurrent graph \(G\) of two constant-treewidth graphs of \(n\) nodes each, the figure illustrates the time required by the variants of our algorithms to preprocess \(G\), and then answer \(i\) pair queries and \(j\) partial pair queries. The different regions correspond to the best variant for handling different number of such queries. In contrast, the current best solution requires \(O(n^3 + i + j)\) time. For ease of presentation we omit the \(O(\cdot)\) notation.

Our contributions. Our main contributions are improved algorithmic upper bounds, proving several optimality results of our algorithms, and experimental results. Below all the complexity measures (time and space) are in the number of basic machine operations and number of semiring operations. We elaborate our contributions below.

2. Optimality of our results. Given our significant improvements for the case of two components, a very natural question is whether the algorithms can be improved further. While presenting matching bounds for polynomial-time graph algorithms to establish optimality is very rare in the whole of computer science, we present conditional lower bounds which show that our combined preprocessing and query time cannot be improved without achieving a major breakthrough in graph algorithms.

- Almost optimality. First, note that in the first result (obtained from Corollary 2) our space usage and single-source query time are arbitrarily close to optimal, as both the input and the output have size \(Θ(n^3)\). Moreover, the result is achieved with preprocessing time less than \(Ω(n^4)\), which is a lower bound for computing the transitive closure (which has \(n^4\) entries). Furthermore, in our third result (obtained from Corollary 3) the \(O(n^{4κ+1})\) preprocessing time is arbitrarily close to optimal, and the \(O(n^4)\) preprocessing space is indeed optimal, as the transitive closure computes the dis-
1. Technical contributions. The results of this paper rely on several novel technical contributions.


(a) The first key result is an algorithm for constructing a strongly balanced tree-decomposition $T$. A tree is called $(\beta, \gamma)$-balanced if for every node $u$ and descendant $v$ of $u$ that appears $\gamma$ levels below, the size of the subtree of $T$ rooted at $v$ is at most a $\beta$ fraction of the size of the subtree of $T$ rooted at $u$. For any fixed $\delta > 0$ and $\lambda \in \mathbb{N}$ with $\lambda \geq 2$, let $\beta = ((1 + \delta)/2)^{\lambda - 1}$ and $\gamma = \lambda$. We show that a $(\beta, \gamma)$-balanced tree decomposition of a constant-treewidth graph with $n$ nodes can be constructed in $O(n \cdot \log n)$ time and $O(n)$ space. To our knowledge, this is the first algorithm that constructs a tree decomposition with such a strong notion of balance. This property is crucial for achieving the resource bounds of our algorithms for algebraic paths. The construction is presented in Section 3.

(b) Given a concurrent graph $G$ obtained from $k$ constant-treewidth graphs $G_i$, we show how a tree-decomposition of $G$ can be constructed from the strongly balanced tree-decompositions $T_i$ of the components $G_i$, in time that is linear in the size of the output. We note that $G$ can have large treewidth, and thus determining the treewidth of $G$ can be computationally expensive. Instead, our construction avoids computing the treewidth of $G$, and directly constructs a tree-decomposition of $G$ from the strongly balanced tree decompositions $T_i$. The construction is presented in Section 4.

(c) Given the above tree-decomposition algorithm for concurrent graphs $G$, in Section 5 we present the algorithms for handling algebraic path queries. In particular, we introduce the partial expansion $C$ of $G$ for additionally handling partial queries, and describe the algorithms for preprocessing and querying $C$ in the claimed time and space bounds.

2. Lower bound. Given an arbitrary graph $G$ (not of constant treewidth) of $n$ nodes, we show how to construct a constant-treewidth graph $G'$ of $2 \cdot n$ nodes, and a graph $G''$ that is the product of $G'$ with itself, such that algebraic path queries in $G$ coincide with such queries in $G''$. This construction requires quadratic time on $n$. The conditional optimality of our algorithms follows, as improvement over our algorithms must achieve the same improvement for algebraic path properties on arbitrary graphs.

All our algorithms are simple to implement and provided as pseudocode in Appendix F. Several technical proofs are also relegated to the full version due to lack of space.

1.1 Related Works

Treewidth of graphs. The notion of treewidth of graphs as an elegant mathematical tool to analyze graphs was introduced in [67]. The significance of constant treewidth in graph theory is large mainly because several problems on graphs become complexity-wise easier. Given a tree decomposition of a graph with low treewidth $t$, many NP-complete problems for arbitrary graphs can be solved in time polynomial in the size of the graph, but exponential in $t$ [4, 7, 9–11]. Even for problems that can be solved in polynomial time, faster algorithms can be obtained for low treewidth graphs, e.g., for the distance problem [26]. The constant-treewidth property of graphs has also been used in the context of logic: Monadic Second Order (MSO) logic is a very expressive logic, and a celebrated result of [28] showed that for constant-treewidth graphs the decision questions for MSO can be solved in polynomial time; and the result of [32] shows that this can even be achieved in deterministic log-space. Various other models (such as probabilistic models of Markov decision processes and games played on graphs for synthesis) with the constant-treewidth restriction have also been considered [20, 61]. The problem of computing a balanced tree decomposition for a constant treewidth graph was considered in [63]. More importantly, in the context of programming languages, it was shown in [70] that the controlflow graphs of goto-free programs in many programming languages have constant treewidth. This theoretical result was subsequently followed up in several practical approaches, and although in the presence of gotos the treewidth is not guaranteed to be bounded, it has been shown that programs in several programming languages have typically low treewidth [17, 39].

The constant-treewidth property of graphs has been used to develop faster algorithms for sequential interprocedural analysis [24], and on the analysis of automata with auxiliary storage (e.g., stacks and queues) [59]. These results have been followed in practice, and some compilers (e.g., SDCC) implement tree-decomposition-based algorithms for performance optimizations [52].

Concurrent system analysis. The problem of concurrent system analysis has been considered in several works, both for intraprocedural as well context-bounded interprocedural analysis [3, 15, 34, 42, 47, 53–55, 62], and many practical tools have been developed as well [54, 56, 62, 69]. In this work we focus on the intraprocedural analysis with constant-treewidth graphs, and present algorithms with better asymptotic complexity. To our knowledge, none of the previous works consider the constant-treewidth property, nor do they improve the asymptotic complexity of the basic algorithm for the algebraic path property problem.
2. Definitions

In this section we present definitions related to semirings, graphs, concurrent graphs, and tree decompositions. We start with some basic notation on sets and sequences.

**Notation on sets and sequences.** Given a number \(|r| \in \mathbb{N}\), we denote by \([r] = \{1, 2, \ldots, r\}\) the natural numbers from 1 to \(r\). Given a set \(X\) and a \(k \in \mathbb{N}\), we denote by \(X^k = \prod_{i=1}^{k} X\), the \(k\) times Cartesian product of \(X\). A sequence \(x_1, \ldots, x_k\) is denoted for short by \((x_i)_{1 \leq i \leq k}\), or \((x_i)\) when \(k\) is implied from the context. Given a sequence \(Y\), we denote by \(y \in Y\) the fact that \(y\) appears in \(Y\).

### 2.1 Complete, closed semirings

**Definition 1** (Complete, closed semirings). We fix a complete semiring \(S = (\Sigma, \oplus, \otimes, 0, 1)\) where \(\Sigma\) is a countable set, \(\oplus\) and \(\otimes\) are binary operators on \(\Sigma\), and \(0, 1 \in \Sigma\), and the following properties hold:

1. \(\oplus\) is infinitely associative, commutative, and \(0\) is the neutral element.
2. \(\otimes\) is associative, and \(1\) is the neutral element.
3. \(\otimes\) infinitely distributes over \(\oplus\).
4. \(0\) absorbs in multiplication, i.e., \(\forall a \in \Sigma: a \otimes 0 = 0\).

Additionally, we consider that \(S\) is equipped with a closure operator \(\#\), such that \(\forall s \in \Sigma: s^\# = 1 \oplus (s \otimes s^\#) = 1 \oplus (s^\# \otimes s)\) (i.e., the semiring is closed).

### 2.2 Graphs and tree decompositions

**Graphs and weighted paths.** Let \(G = (V, E)\) be a weighted finite directed graph (henceforth called simply a graph) where \(V\) is a set of \(n\) nodes and \(E \subseteq V \times V\) is an edge relation, along with a weight function \(w: E \to \mathbb{R}\) that assigns to each edge of \(G\) an element from \(\Sigma\). Given a set of nodes \(X \subseteq V\), we denote by \(G[X] = (X, E \cap (X \times X))\) the subgraph of \(G\) induced by \(X\). A path \(P: u \prec \ldots \prec v\) is a sequence of nodes \((x_1, \ldots, x_k)\) such that \(x_1 = u, x_k = v\), and for all \(1 \leq i < k\) we have \((x_i, x_{i+1}) \in E\).

The length of \(P\) is \(|P| = k - 1\), and a single node is itself a 0-length path. A path is simple if no node repeats in the path (i.e., it does not contain a cycle). Given a path \(P = (x_1, \ldots, x_k)\), the weight of \(P\) is \(\text{wt}(P) = \prod_{i=1}^{k} w(x_{i+1}, x_i)\), if \(|P| \geq 1\) else \(\text{wt}(P) = 1\).

Given nodes \(u, v \in V\), the *seminizing distance* \(d(u, v)\) is defined as \(d(u, v) = \min\{d(u, v) + \text{wt}(P)\}\), and \(d(u, v) = 0\) if no such \(P\) exists.

**Trees.** A tree \(T = (V, E)\) is an undirected graph with a root node \(v_0\), such that between every two nodes there is a unique simple path. For a node \(v\) we denote by \(Lv(v)\) the *level* of \(u\) which is defined as the length of the simple path from \(v_0\) to \(u\). A *child of a node* \(u\) is a node \(v\) such that \(Lv(v) = Lv(u) + 1\) and \((u, v) \in E\), and then \(u\) is the *parent* of \(v\). For a node \(v\), any node (including \(v\) itself) that appears in the path from \(v_0\) to \(u\) is an ancestor of \(v\), and if \(v\) is an ancestor of \(u\), then \(u\) is a *descendant of \(v\)*. Given two nodes \(u, v\), the *lowest common ancestor* (LCA) is the common ancestor of \(u\) and \(v\) with the highest level. Given a tree \(T\), a *contiguous subtree* is the subgraph \((X, E')\) of \(T\) such that \(E' = E \cap (X \times X)\) and for every pair \((u, v) \in E\), every node that appears in the unique path from \(u\) to \(v\) belongs to \(X\). A tree is *k-ary* if every node has at most \(k\)-children (e.g., a binary tree has at most two children for every node). In a full \(k\)-ary tree, every node that 0 or \(k\)-children.

**Tree decompositions.** A tree-decomposition \(T = (V_T, E_T)\) of a graph \(G\) is a tree, where every node \(B_i \in T\) is a subset of the following conditions hold:

- \(V_T = \bigcup_{i} B_i\)
- \(X, E \subseteq V_T\)
- \(B_i \cap B_j = B_{i,j}\)
- \(\forall x \in V: \exists i: x \in B_i\)
- \(\forall i, j: \exists k: B_{i,j} \subseteq B_k\)

![Figure 2: A graph \(G\) with a tree-decomposition \(T\) and a corresponding tree-decomposition \(T'\) of \(G\) of 8 bags and width 2.](image)

The distinguished bag \(B_0\) is the root bag of node 6. We have \(V_T(B_0) = \{6, 7, 9, 4, 5\}\) and \(V_T'(B_0) = \{6, 7, 9, 8, 10, 2\}\). The subtree \(T(B_0)\) is shown in bold.

C1. \(V_p = \{B_0, \ldots, B_7\}\) with \(B_i \subseteq V\), and \(\bigcup_{B_i \in V} B_i = V\)
C2. For all \((u, v) \in E\) there exists \(B_i \in T\) such that \(u, v \in B_i\)
C3. For all \((u, v) \in E\) there exists \(B_i \in T\) such that \(u, v \in B_i\)

The sets \(B_i\) which are nodes in \(V_T\) are called bags. We denote by \(|T| = |V_T|\) the number of bags in \(T\). Conventionally, we call \(B_0\) the root of \(T\), and denote by \(Lv(B_i)\) the level of \(B_i\) in \(T\). For a bag \(B\), we denote by \(T(B)\) the subtree of \(T\) rooted at \(B\). A bag \(B\) is called the root bag of a node \(u\) if \(u \in B\) and every \(B'\) that contains \(u\) appears in \(T(B)\).

The sets \(B_i\) which are nodes in \(V_T\) are called bags. We denote by \(|T| = |V_T|\) the number of bags in \(T\). Conventionally, we call \(B_0\) the root of \(T\), and denote by \(Lv(B_i)\) the level of \(B_i\) in \(T\). For a bag \(B\), we denote by \(T(B)\) the subtree of \(T\) rooted at \(B\). A bag \(B\) is called the root bag of a node \(u\) if \(u \in B\) and every \(B'\) that contains \(u\) appears in \(T(B)\).

The width of the tree-decomposition \(T\) is the size of the largest bag minus 1. The tree width of \(G\) is the smallest width among the widths of all tree decompositions of \(G\). Note that if \(T\) achieves the treewidth of \(G\), we have \(|V_T(B_i)| \leq (t + 1) \cdot |T(B)|\).

**Product graphs.** A graph \(G = (V_p, E_p)\) is said to be the product graph of a set of graphs \((G_i = (V_i, E_i))_{1 \leq i \leq k}\) if \(V_p = \prod_{i} V_i\) and \(E_p\) is such that for all \(u, v \in V_p\) with \(u = (u_i)_{1 \leq i \leq k}\) and \(v = (v_i)_{1 \leq i \leq k}\), we have \((u, v) \in E_p\) if there exists a set \(I \subseteq [k]\) such that \(i \in I\) for all \(i \in I\), and \(u_i = v_i\) for all \(i \notin I\).

In words, an edge \((u, v) \in E_p\) is formed in the product graph by traversing a set of edges \((u_i, v_i) \in E_i\) in each component graph \((G_i)\) of \(G\), and traversing no edges in the remaining \((G_j)\) of \(G\).

We say that \(G_p\) is the *k-self-product* of a graph \(G^i\) if \(G^i = G^i\) for all \(1 \leq i \leq k\).

**Concurrent graphs.** A graph \(G = (V, E)\) is called a concurrent graph if \(G = (V, E)\) is the product graph of \((G_i)\) of some component graphs \((G_i)\) of \(G\). A concurrent graph \((G, E)\) of a node \(u\) in \(V\), we will denote by \(u_i\) the \(i\)th constituent of \(u\). We say that \(G\) is a k-self-concurrent of a graph \(G^i\) if \(G_p\) is the k-self-product of \(G^i\).
Various notions of composition. The framework we consider is quite general as it captures various different notions of concurrent composition. Indeed, the edge set of the concurrent graph is any possible subset of the edge set of the corresponding product graph. Then, two well-known composition notions can be modeled as follows. For any edge \((u, v) \in E\) of the concurrent graph \(G\), let \(Z_{u,v} = \{i \in [k] : (u, v) \in E_i\}\) denote the components that execute a transition \((u, v)\).

1. In synchronous composition at every step all components make one move each simultaneously. This is captured by \(Z_{u,v} = [k]\) for all \((u, v) \in E\).
2. In asynchronous composition at every step only one component makes a move. This is captured by \(|Z_{u,v}| = 1\) for all \((u, v) \in E\).

Thus the framework we consider is not specific to any particular notion of composition, and all our results apply to various different notions of concurrent composition that exist in the literature.

Partial nodes of concurrent graphs. A partial node \(\overline{\pi}\) of a concurrent graph \(G\) is an element of \([\bigcup_{i=1}^k (V_i \cup \{i\})]\), where \(\bot \notin \bigcup_{i=1}^k V_i\). Intuitively, \(\bot\) is a fresh symbol to denote that a component is unspecified. A partial node \(\overline{\pi}\) is said to refine a partial node \(\pi\), denoted by \(\overline{\pi} \subseteq \pi\) if for all \(1 \leq i \leq k\) either \(\pi_i = \emptyset\) or \(\pi_i = \overline{\pi}_i\). We say that the partial node \(\overline{\pi}\) strictly refines \(\pi\), denoted by \(\overline{\pi} \subset \pi\) if \(\overline{\pi} \subseteq \pi\) and \(\pi \neq \overline{\pi}\) (i.e., for at least one constituent \(i\) we have \(\pi_i = \emptyset\) but \(\overline{\pi}_i \neq \emptyset\)). A partial node \(\overline{\pi}\) is called strictly partial if it is strictly refined by some node \(u \in V\) (i.e., \(\overline{\pi}\) has at least one \(\bot\)). The notion of semiring distances is extended to partial nodes, and for partial nodes \(\overline{\pi}, \overline{\tau}\) of \(G\) we define the semiring distance from \(\overline{\pi}\) to \(\overline{\tau}\) as

\[
d(\overline{\pi}, \overline{\tau}) = \bigoplus_{\overline{u}, \overline{v} \in \overline{\pi} \times \overline{\tau}} d(u, v)
\]

where \(u, v \in V\). In the sequel, a partial node \(\overline{\pi}\) will be either (i) a node of \(V\), or (ii) a strictly partial node. We refer to nodes of the first case as actual nodes, and write \(u\) (i.e., without the bar). Distances where one endpoint is a strictly partial node \(\overline{\pi}\) succinctly quantify over all nodes of all the components for which the corresponding constituent of \(\overline{\pi}\) is \(\emptyset\). Observe that the distance still depends on the unspecified components.

The algebraic paths problem on concurrent graphs of constant-treewidth components. In this work we are interested in the following problem. Let \(G = (V, E)\) be a concurrent graph of \(k \geq 2\) constant-treewidth graphs \((G_i, (V_i, E_i))_{1 \leq i \leq k}\) and \(wt : E \rightarrow \Sigma\) be a weight function that assigns to every edge of \(G\) a weight from a set \(\Sigma\) that is turned to an \(\Sigma\)-tree-decomposition if every bag of \(G\) asks the \(\alpha, \beta, \gamma\)-tree-decomposition \(\text{Tree}(G)\) with \(O(n)\) bags can be constructed in \(O(n \log n)\) time and \(O(n)\) space.

Sketch of Theorem 1. The construction of Theorem 1 considers that a tree-decomposition \(\text{Tree}'(G)\) of width \(t\) and \(O(n)\) bags is given (which can be obtained using e.g. [12] in \(O(n)\) time). Given two parameters \(\delta > 0\) and \(\lambda \in \mathbb{N}\) with \(\lambda \geq 2\), let \(\alpha = 4 : \lambda / \delta\), \(\beta = ((1 + \delta) / 2)^{\lambda - 1}\), and \(\gamma = \lambda\). A \((\alpha, \beta, \gamma)\) tree-decomposition \(\text{Tree}(G)\) is turned to an \((\alpha, \beta, \gamma)\) tree-decomposition, for \(\alpha = 4 : \lambda / \delta\), \(\beta = ((1 + \delta) / 2)^{\lambda - 1}\), and \(\gamma = \lambda\), in two conceptual steps.

1. A tree of bags \(R_G\) is constructed, which is \((\beta, \gamma)\)-balanced.
2. \(R_G\) is turned to an \(\alpha\)-approximate tree decomposition of \(G\).

The first construction is obtained by a recursive algorithm Rank, which operates on inputs \((C, \ell)\), where \(C\) is a component of \(\text{Tree}'(G)\), and \(\ell \in [\lambda]\) specifies the type of operation the algorithm performs on \(C\). Given such a component \(C\), we denote by \(\text{Nh}(C)\) the neighborhood of \(C\), defined as the set of bags of \(\text{Tree}'(G)\) that are incident to \(C\). Informally, on input \((C, \ell)\), the algorithm partitions \(C\) into two sub-components \(C_1\) and \(C_2\) such that either (i) the size of each \(C_1\) is approximately half the size of \(C\), or (ii) the size of the neighborhood of each \(C_1\) is approximately half the size of the neighborhood of \(C\). More specifically,

1. If \(\ell > 0\), then \(C\) is partitioned into components \(C' = (C_1, \ldots, C_\ell)\), by removing a list of bags \(X = (B_1, \ldots, B_n)\), such that \(|C_i| \leq \frac{1}{2} \cdot |C|\). The union of \(X\) yields a new bag
**Method: DiningPhilosophers**

```plaintext
1 while True do
2   while fork not mine or knife not mine do
3     if fork is free then
4       lock(fork)
5       acquire(fork)
6       unlock(fork)
7     end
8     if knife is free then
9       lock(knife)
10      acquire(knife)
11      unlock(knife)
12    end
13   end
14   dine(fork, knife) // for some time
15   lock(fork)
16   release(fork)
17   release(knife)
18   unlock(fork)
19   discuss() // for some time
20 end
```

Figure 3: A concurrent program (left), its controlflow graph (middle), and a tree decomposition of the controlflow graph (right).

Figure 4: Illustration of one recursive step of Rank on a component $C$ (gray). $C$ is split into two sub-components $\overline{C}_1$ and $\overline{C}_2$ by removing a list of bags $X = \{B_i\}_i$. Once every $\lambda$ recursive calls, $X$ contains one bag, such that the neighborhood $\text{Nh}(\overline{C}_i)$ of each $\overline{C}_i$ is at most half the size of $\text{Nh}(C)$ (i.e., the red area is split in half). In the remaining $\lambda-1$ recursive calls, $X$ contains $m$ bags, such that the size of each $\overline{C}_i$, is at most $\frac{1}{1+\frac{m}{\lambda}}$ fraction the size of $C$. (i.e., the gray area is split in almost half).

1. $B$ in $R_i$. Then $Y$ is merged into two components $\overline{C}_1, \overline{C}_2$ with $|\overline{C}_1| \leq |\overline{C}_2| \leq \frac{1+\frac{m}{\lambda}}{2} \cdot |C|$. Finally, each $\overline{C}_i$ is passed on to the next recursive step with $\ell = (\ell + 1) \mod \lambda$.
2. If $\ell = 0$, then $C$ is partitioned into two components $\overline{C}_1, \overline{C}_2$ such that $|\text{Nh}(\overline{C}_1) \cap \text{Nh}(C)| \leq \frac{|\text{Nh}(C)|}{2}$ by removing a single bag $B$.

This bag becomes a new bag $B$ in $R_i$, and each $\overline{C}_i$ is passed on to the next recursive step with $\ell = (\ell + 1) \mod \lambda$.

Figure 4 provides an illustration. The second construction is obtained simply by inserting in each bag $B$ of $R_i$, the nodes contained in the neighborhood $\text{Nh}(C)$ of the component $C$ upon which $B$ was constructed.

**Remark.** The notion of balanced tree decompositions exists in the literature [13, 32], but balancing only requires that the height of the tree is logarithmic in its size. Here we develop a stronger notion of balancing, which is crucial for proving the complexity results of the algorithms presented in this work.

### 4. Concurrent Tree Decomposition

In this section we present the construction of a tree-decomposition $\text{Tree}(G)$ of a concurrent graph $G = (V, E)$ of $k$ constant-treewidth graphs. In general, $G$ can have treewidth which depends on the number of its nodes (e.g., $G$ can be a grid, which has treewidth $n$, obtained as the product of two lines, which have treewidth 1). While the treewidth computation for constant-treewidth graphs is linear time [12], it is NP-complete for general graphs [11]. Hence computing a tree decomposition that achieves the treewidth of $G$ can be computationally expensive (e.g., exponential in the size of $G$). Here we develop an algorithm $\text{ConcurTree}$ which constructs a tree-decomposition $\text{ConcurTree}(G)$ of $G$, given a $(\alpha, \beta, \gamma)$ tree-decomposition of the components, in $O(n^k)$ time and space (i.e., linear in the size of $G$), such that the following properties hold: (i) the width is $O(n^{k-1})$; and (ii) for every bag in level at least $i \cdot \gamma$, the size of the bag is $O(n^{k-1} \cdot \beta^i)$ (i.e., the size of the bags decreases geometrically along the levels).

**Algorithm $\text{ConcurTree}$ for concurrent tree decomposition.** Let $G$ be a concurrent graph of $k$ graphs $(G_i)_{1 \leq i \leq k}$. The input consists of a full binary tree-decomposition $T_i$ of constant width for every graph $G_i$. In the following, $B_i$ ranges over bags of $T_i$, and we denote by $B_{i,r}$, with $r \in [2]$, the $r$-th child of $B_i$. We construct the concurrent tree-decomposition $T = \text{ConcurTree}(G) = (V_T, E_T)$ of $G$ using the recursive procedure $\text{ConcurTree}$, which operates as follows. On input $(T_i(B_i))_{1 \leq i \leq k}$, return a tree decomposition where
Figure 5: The tree-decomposition ConcurTree(G) of a concurrent graph G of two constant-treewidth graphs G1 and G2.

1. The root bag B is

\[ B = \bigcup_{1 \leq i \leq k} \left( \prod_{j < i} V_{T_i}(B_j) \times B_i \times \prod_{j > i} V_{T_i}(B_j) \right) \]  

(1)

2. If every \( B_i \) is a non-leaf bag of \( T_i \), for every choice of \( \{ x_1, \ldots, x_k \} \subseteq \{ 1, 2, \ldots, n \} \), we establish that the solution to the above recurrence is such that for every constant \( \alpha, \beta, \gamma \)

\[ T(n_1, \ldots, n_k) \leq \sum_{1 \leq i \leq k} n_j \prod_{i \neq j} n_j + \sum_{(r_1, \ldots, r_k) \in [2]^k} T(n_{r_1}, \ldots, n_{r_k}) \]  

(2)

such that for every i we have that \( \sum_{r_i \in [2]} n_{i,r_i} \leq n_i \). In Appendix C we establish that the solution to the above recurrence is \( O(n^k) \), where \( n_i \leq n \) for all \( 1 \leq i \leq k \).

The following theorem establishes that ConcurTree(G) is a tree-decomposition of G constructed in \( O(n^k) \) time and space. Additionally, if every input tree-decomposition \( T_i \) is \( (\beta, \gamma) \)-balanced, then the size of each bag \( B \) of ConcurTree(G) decreases geometrically with its level \( \text{Lv}(B) \). See Appendix C for a formal proof.

**Theorem 2.** Let \( G = (V, E) \) be a concurrent graph of k constant-treewidth graphs \( (G_i)_{1 \leq i \leq k} \) of n nodes each. Let a binary \((\alpha, \beta, \gamma)\) tree-decomposition \( T \) for every graph \( G_i \) be given, for some constant \( \alpha \). ConcurTree constructs a \( 2^\alpha \)-ary tree-decomposition ConcurTree(G) of G in \( O(n^k) \) time and space, with the following properties. For every \( i \in \mathbb{N} \) and bag B at level \( \text{Lv}(B) \geq i \gamma \), we have \( |B| = O(n^{k-1} \cdot \beta^i) \).

5. Concurrent Algebraic Paths

We now turn our attention to the core algorithmic problem of this paper, namely answering semiring distance queries in a concurrent graph G of k constant-treewidth graphs \( (G_i)_{1 \leq i \leq k} \). To this direction, we develop a data-structure ConcurAP (for concurrent algebraic paths) which will preprocess G and afterwards support single-source, pair, and partial pair queries on G. As some of the proofs are technical, they are presented only in Appendix D.

Semiring distances on tree decompositions. The preprocessing and query of our data-structure exploits a key property of semiring distances on tree decompositions. This property is formally stated in Lemma 1, and concerns any two nodes \( u, v \) that appear in some distinct bags \( B_1, B_2 \) of Tree(G). Informally, the semiring distance \( d(u, v) \) can be written as the semiring multiplication of distances \( d(x_i, x_{i+1}) \), where \( x_i \) is a node that appears in the i-th and \((i-1)\)-th bags of the unique simple path \( B_1 \rightsquigarrow B_2 \) in Tree(G). Figure 6 provides an illustration.

**Lemma 1.** Consider a graph \( G = (V, E) \) with a weight function \( w : E \rightarrow \Sigma \) and a tree-decomposition Tree(G). Let \( u, v \in V \), and let \( P : B_1, B_2, \ldots, B_m \) be a simple path in T such that \( u \in B_1 \) and \( v \in B_m \). Let \( A = \{u\} \times \prod_{1 \leq i \leq m} (B_i - 1 \cap B_i) \times \{v\} \). Then

\[ d(u, v) = \bigoplus_{(x_1, x_2, \ldots, x_m) \in A} \bigotimes_{i=1}^m d(x_i, x_{i+1}). \]

**Informal description of the preprocessing.** The preprocessing phase of ConcurAP is handled by algorithm ConcurPreprocess, which performs the following steps.

1. First, the partial expansion \( \mathcal{T} \) of G is constructed by introducing a pair of strictly partial nodes \( \pi, \pi' \) for every strictly partial node \( \pi \) of G, and edges between strictly partial nodes and the corresponding nodes of G that refine them.

2. Second, the concurrent tree-decomposition \( T = \text{ConcurTree}(G) \) of G is constructed, and modified to a tree-decomposition \( \mathcal{T} \) of the partial expansion graph \( \mathcal{G} \).

3. Third, a standard, two-way pass of \( \mathcal{T} \) is performed to compute local distances. In this step, for every bag \( B \) in \( \mathcal{T} \) and all partial nodes \( \pi, \pi \in B \), the distance \( d(\pi, \pi) \) is computed (i.e., all-pair distances in B). Since we compute distances between nodes that are local in a bag, this step is called local distance computation. This information is used to handle (i) single-source queries and (ii) partial pair queries in which both nodes are strictly partial.

4. Finally, a top-down pass of \( \mathcal{T} \) is performed in which for every node \( u \) and partial node \( \pi \in \mathcal{T}(B_u) \) (i.e., it appears in some ancestor of \( B_u \)) the distances \( d(\pi, u) \) and \( d(\pi, v) \) are computed.
Algorithm ConcurPreprocess. We now formally describe algorithm ConcurPreprocess for preprocessing the concurrent graph $G = (V, E)$ for the purpose of answering algebraic path queries. For any desired $0 < \epsilon < 1$, we choose appropriate constants $\alpha, \beta, \gamma$, which will be defined later for the complexity analysis. On input $G = (V, E)$, where $G$ is a concurrent graph of $k$ constant-treewidth graphs $(G_i = (V_i, E_i))_{1 \leq i \leq k}$, and a weight function $w : E \to \Sigma$, ConcurPreprocess operates as follows:

1. Construct the partial expansion $G = (\tilde{V}, \tilde{E})$ of $G$ together with an extended weight function $\tilde{w} : \tilde{E} \to \Sigma$ as follows:
   (a) The node set is $\tilde{V} = V \cup \{\tilde{u}, \tilde{v} : \exists u \in V \text{ s.t. } u \subset \tilde{u}\}$; i.e., $\tilde{V}$ consists of nodes in $V$ and two copies for every partial node $\tilde{v}$ that is strictly refined by a node $u$ of $G$.
   (b) The edge set is $\tilde{E} = E \cup \{(\tilde{u}, u), (u, \tilde{v}) : \tilde{u}, \tilde{v} \in \tilde{V} \text{ and } u \in V \text{ s.t. } u \subset \tilde{v}\}$, i.e., along with the original edges $E$, the first (resp. second) copy of every strictly partial node has outgoing (resp. incoming) edges to (resp. from) the nodes of $G$ that refine it.
   (c) For the weight function we have $\tilde{w}(\tilde{u}, u) = w(u, \tilde{v})$ if $\tilde{v} \in V$, and $\tilde{w}(\tilde{u}, \tilde{v}) = \tilde{I}$ otherwise. This is, the original weight function is extended with value $\tilde{I}$ (which is neutral for semiring multiplication) to all new edges in $\tilde{G}$.

2. Construct the tree-decomposition $T = (V_T, E_T)$ of $\tilde{G}$ as follows.
   (a) Obtain an $(\alpha, \beta, \gamma)$-tree-decomposition $T_i = \text{Tree}(G_i)$ of every graph $G_i$ using Theorem 1.
   (b) Construct the concurrent tree-decomposition $T = \text{ConcurTree}(G)$ of $G$ using $(T_i)_{1 \leq i \leq k}$.
   (c) Let $\tilde{T}$ be identical to $T$, with the following exception: For every bag $B$ of $T$ and $\tilde{B}$ the corresponding bag in $\tilde{T}$, for every node $u \in B$, insert in $\tilde{B}$ all strictly partial nodes $\tilde{v}'$, $\tilde{v}''$ of $\tilde{V}$ that $u$ refines. Formally, set $\tilde{B} = B \cup \{\tilde{v}', \tilde{v}'' : \exists u \in B \text{ s.t. } u \subset \tilde{v}'\}$. Note that also $u \in \tilde{B}$.

Observe that the root bag of $\tilde{T}$ contains all strictly partial nodes.

3. Perform the local distance computation on $\tilde{T}$ as follows.
   For every partial node $\tilde{v}$, maintain two map-data-structures $\text{FWD}_{\tilde{v}}$, $\text{BWD}_{\tilde{v}} : \tilde{B}_{\tilde{v}} \to \Sigma$. Intuitively, $\text{FWD}_{\tilde{v}}$ (resp. $\text{BWD}_{\tilde{v}}$) aims to store the forward (resp., backward) distance, i.e., distance from (resp., to) $\tilde{v}$ to $\tilde{v'}$ (resp. from) vertices in $\tilde{B}_{\tilde{v}}$. Initially set $\text{FWD}_{\tilde{v}}(\tilde{v}) = \tilde{I}$ and $\text{BWD}_{\tilde{v}}(\tilde{v}) = \tilde{I}$ for all partial nodes $\tilde{v} \in \tilde{B}_{\tilde{v}}$. At any point in the computation, a given bag $\tilde{B}$ we denote by $\text{wrt}_{\tilde{B}} : \tilde{B} \times \tilde{B} \to \Sigma$ a map-data-structure such that for every pair of partial nodes $\tilde{v}$, $\tilde{v'}$ with $\tilde{w}(\tilde{v}) \leq \tilde{w}(\tilde{v'})$ we have $\text{wrt}_{\tilde{B}}(\tilde{v}, \tilde{v'}) = \text{FWD}_{\tilde{v'}}(\tilde{v})$ and $\text{wrt}_{\tilde{B}}(\tilde{v}, \tilde{v'}) = \text{BWD}_{\tilde{v'}}(\tilde{v})$.

(a) Traverse $\tilde{T}$ bottom-up, and for every bag $\tilde{B}$, execute an all-pairs algebraic path computation on $G[\tilde{B}]$ with weight function $\text{wrt}_{\tilde{B}}$. This is done using classical algorithms for the transitive closure, e.g., [36, 50, 58, 74]. For every pair of partial nodes $\tilde{v}$, $\tilde{v'}$ with $\tilde{w}(\tilde{v}) \leq \tilde{w}(\tilde{v'})$, set $\text{BWD}_{\tilde{v}}(\tilde{v}) = d'(\tilde{v}, \tilde{v'})$ and $\text{FWD}_{\tilde{v}}(\tilde{v}) = d'(\tilde{v}, \tilde{v'})$, where $d'(\tilde{v}, \tilde{v'})$ and $d'(\tilde{v}, \tilde{v'})$ are the computed distances in $G[\tilde{B}]$.

(b) Traverse $\tilde{T}$ top-down, and for every bag $\tilde{B}$ perform the computation of Item 3a.

4. Perform the ancestor distance computation on $\tilde{T}$ as follows.
   For every node $u$, maintain two map-data-structures $\text{FWD}_{\tilde{u}}$, $\text{BWD}_{\tilde{u}} : \tilde{B}_{\tilde{u}} \to \Sigma$ from partial nodes that appear in the ancestor bags of $\tilde{B}_{\tilde{u}}$ to $\Sigma$. These maps aim to capture distances between the node $u$ and nodes in the ancestor bags of $\tilde{B}_{\tilde{u}}$ (in contrast to FWD$_u$ and BWD$_u$ which store distances only between $u$ and nodes in $B_u$). Initially, set $\text{FWD}_{\tilde{u}}(\tilde{v}) = \text{FWD}_{\tilde{u}}(\tilde{v})$ and $\text{BWD}_{\tilde{u}}(\tilde{v}) = \text{BWD}_{\tilde{u}}(\tilde{v})$ for every partial node $\tilde{v} \in \tilde{B}_{\tilde{u}}$. Given a pair of partial nodes $\tilde{v}$, $\tilde{v'}$ with $\text{wrt}_{\tilde{B}}(\tilde{v'}) \leq \text{wrt}_{\tilde{B}}(\tilde{v})$ we denote by $\text{wrt}_{\tilde{B}}(\tilde{v}', \tilde{v}) = \text{FWD}_{\tilde{v}}(\tilde{v})$ and $\text{wrt}_{\tilde{B}}(\tilde{v}', \tilde{v}) = \text{BWD}_{\tilde{v}}(\tilde{v})$. Traverse $\tilde{T}$ via a DFS starting from the root, and for every encountered bag $\tilde{B}$ with parent $\tilde{B}'$, for every node $u$ such that $\tilde{B}$ is the root bag of $u$, for every partial node $\tilde{v} \in \text{wrt}_{\tilde{B}}(\tilde{B}_u)$, assign

$$\text{FWD}_{\tilde{u}}(\tilde{v}) = \bigoplus_{x \in \text{wrt}_{\tilde{B}}(\tilde{B}_u)} \text{FWD}_{\tilde{u}}(x) \otimes \text{wrt}_{\tilde{B}}(x, \tilde{v})$$

(b) Traverse $\tilde{T}$ top-down, and for every bag $\tilde{B}$ perform the computation of Item 3a.

We now turn our attention to the analysis of ConcurPreprocess:

**Lemma 2.** $\tilde{T}$ is a tree decomposition of the partial expansion $\tilde{G}$.

In Lemma 3 we establish that the forward and backward maps computed by ConcurPreprocess store the distances between nodes.

**Lemma 3.** At the end of ConcurPreprocess, the following assertions hold:

1. For all nodes $u, v \in V$ such that $\tilde{B}_{\tilde{u}}$ appears in $\tilde{T}$(\tilde{B}_v), we have $\text{FWD}_{\tilde{u}}(v) = \text{d}(u, v)$ and $\text{BWD}_{\tilde{u}}(v) = \text{d}(v, u)$.

2. For all strictly partial nodes $\tilde{v} \in \tilde{V}$ and nodes $u \in V$, we have $\text{FWD}_{\tilde{v}}(\tilde{v}) = \text{d}(u, \tilde{v})$ and $\text{BWD}_{\tilde{v}}(\tilde{v}) = \text{d}(\tilde{v}, u)$.

3. For all strictly partial nodes $\tilde{v}, \tilde{v}' \in \tilde{V}$ we have $\text{FWD}_{\tilde{v}}(\tilde{v}') = \text{d}(\tilde{v}, \tilde{v}')$ and $\text{BWD}_{\tilde{v}}(\tilde{v}') = \text{d}(\tilde{v}', \tilde{v})$.

**Proof.** We describe the key invariants that hold during the traversals of $\tilde{T}$ by ConcurPreprocess in Item 3a, Item 3b and Item 4 after the algorithm processes a bag $\tilde{B}$. All cases depend on Lemma 1.
Item 3a For every pair of partial nodes π, τ ∈ $\mathcal{B}$ such that $Lirement{\tau} ≤ Lirement{\tau}$ we have $\text{FWD}_\tau(\tau) = \tau \oplus \rho_1 \ominus \rho_1$ and $\text{BWD}_\tau(\tau) = \tau \oplus \rho_2 \ominus \rho_2$ where $\rho_1$ and $\rho_2$ are $\tau \rightarrow \tau \text{ and } \tau \rightarrow \tau$ paths respectively that only traverse nodes in $\nabla_\tau(\mathcal{B})$. The statement follows by a straightforward induction on the levels processed by the algorithm in the bottom-up pass. Note that if $\tau$ and $\tau$ are partial nodes in the root of $\mathcal{T}$, the statement yields $\text{FWD}_\tau(\tau) = d(u, v)$ and $\text{BWD}_\tau(\tau) = d(v, u)$.

Item 3b The invariant is similar to the previous, except that $\rho_1$ and $\rho_2$ range over all $\tau \rightarrow \tau$ and $\tau \rightarrow \tau$ paths in $\mathcal{G}$ respectively. Hence now $\text{FWD}_\tau(\tau) = d(\tau, \tau)$ and $\text{BWD}_\tau(\tau) = d(\tau, \tau)$. The statement follows by a straightforward induction on the levels processed by the algorithm in the top-down pass. Note that the base case on the root follows from the previous item, where the maps $\text{BWD}$ and $\text{FWD}$ store actual distances.

Item 4 For every node $u \in $ and partial node $\tau \in \nabla_\tau(\mathcal{B})$ we have $\text{FWD}^*_\tau(\tau) = d(\tau, \tau)$ and $\text{BWD}^*_\tau(\tau) = d(\tau, \tau)$. The statement follows from Lemma 1 and a straightforward induction on the length of the path from the root of $\mathcal{T}$ to the processed bag $\mathcal{B}$.

Statement 1 of the lemma follows from Item 4. Similarly for statement 2, together with the observation that every strictly partial node $\tau$ appears in the root of $\mathcal{T}$, and thus $\tau \in \nabla_\tau(\mathcal{B})$. Finally, statement 3 follows again from the fact that all strictly partial nodes appear in the root bag of $\mathcal{T}$. The desired result follows.

We now consider the complexity analysis, and we start with a technical lemma on recursion relations.

Lemma 4. Consider the recurrences in Eq. (5) and Eq. (6).

\begin{align*}
T_k(n) &\leq n^{3(k-1)} + 2^\lambda \cdot k \cdot \mathcal{T}_k \left( n \cdot \left( 1 + \frac{\delta}{2} \right)^{\lambda-1} \right) \quad (5) \\
S_k(n) &\leq n^{2(k-1)} + 2^\lambda \cdot k \cdot S_k \left( n \cdot \left( 1 + \frac{\delta}{2} \right)^{\lambda-1} \right) \quad (6)
\end{align*}

Then

1. $T_k(n) = O(n^{3(k-1)})$, and
2. (i) $S_k(n) = O(n^{2(k-1)})$ if $k \geq 3$, and (ii) $S_k(n) = O(n^{2(\lambda^2)})$.

The proof of Lemma 4 is technical, and presented in Appendix D. The following lemma analyzes the complexity of $\text{ConcurPreprocess}$, and makes use of the above recurrences. Recall that $\text{ConcurPreprocess}$ takes as part of its input a desired constant $0 < \epsilon \leq 1$. We choose a $\lambda \in \mathbb{Z}$ and $\delta \in \mathbb{R}$ such that $\lambda \geq 4/\epsilon$ and $\delta \leq \epsilon/18$. Additionally, we set $\alpha = 4 \cdot \lambda / \delta$, $\beta = ((1 + \delta)/2)^{\lambda-1}$ and $\gamma = \lambda$, which are the constants used for constructing an $(\alpha, \beta, \gamma)$ tree-decomposition $T_1 = \text{Tree}(G_1)$ in Item 2a of $\text{ConcurPreprocess}$.

Lemma 5. $\text{ConcurPreprocess}$ requires $O(n^{k-1})$ space and

1. $O(n^{3(k-1)})$ time if $k \geq 3$, and
2. $O(n^{2(\lambda^2)})$ time if $k = 2$.

Proof. We examine each step of the algorithm separately.

1. The time and space required for this step is bounded by the number of nodes introduced in the partial expansion $G$, which is $2 \cdot \sum_{k\in \{1\}} (\lambda^n) = O(n^{k-1})$.
2. By Theorem 2, $\text{ConcurTree}(G)$ is constructed in $O(n^2)$ time and space. In $\mathcal{T}$, the size of each bag $\mathcal{B}$ is increased by constant factor, hence this step requires $O(n^2)$ time and space.
3. In each pass, $\text{ConcurPreprocess}$ spends $|\mathcal{B}|^2$ time to perform an all-pairs algebraic paths computation in each bag $\mathcal{B}$ of $\mathcal{T}$ [36, 50, 58, 74]. The space usage for storing all maps $\text{FWD}_\mathcal{B}$ and $\text{BWD}_\mathcal{B}$ for every node $\tau$ whose root bag is $\mathcal{B}$ is $O(|\mathcal{B}|^2)$, since there are at most $|\mathcal{B}|$ such nodes $\tau$, and each map has size $|\mathcal{B}|$. By the previous item, we have $|\mathcal{B}| = O(|\mathcal{B}|)$, where $\mathcal{B}$ is the corresponding bag of $\mathcal{T}$ before the partial expansion of $G$. By Theorem 2, we have $|\mathcal{B}| = O(n^{\lambda^2})$. Therefore, the above time is $\mathcal{T}(n)$, and $\mathcal{T}(n) = (1 + \delta/2)^{\lambda-1} - 1$. Then, since $\mathcal{T}$ is a full $2^\lambda$-ary tree, the time and space required for preprocessing every $\gamma = \lambda$ levels of $\mathcal{T}$ is given by the following recurrences respectively (ignoring constant factors for simplicity).

\begin{align*}
T_k(n) &\leq n^{k(k-1)} + 2^\lambda \cdot \mathcal{T}_k \left( n \cdot \left( 1 + \frac{\delta}{2} \right)^{\lambda-1} \right) \\
S_k(n) &\leq n^{2(k-1)} + 2^\lambda \cdot \mathcal{S}_k \left( n \cdot \left( 1 + \frac{\delta}{2} \right)^{\lambda-1} \right)
\end{align*}

By the analysis of Eq. (5) and Eq. (6) of Lemma 4, we have that $\mathcal{T}_k(n) = O(n^{3(k-1)})$ and (i) $\mathcal{S}_k(n) = O(n^{2(k-1)})$ if $k \geq 3$, and (ii) $\mathcal{S}_k(n) = O(n^{2(\lambda^2)})$.

4. We first focus on the space usage. Let $\mathcal{B}_u$ denote the ancestor bag of $\mathcal{B}_u$ at level $i$. We have

\begin{align*}
|\mathcal{V}_\mathcal{B}(\mathcal{B}_u)| &\leq c_1 \cdot \sum_i |\mathcal{B}_u| \leq c_2 \cdot \sum_i |\mathcal{B}_u| \leq c_3 \cdot \sum_i \left( n^{k-1} \cdot \beta^i \right) = O(n^{k-1})
\end{align*}

for some constants $c_1, c_2, c_3$. The first inequality comes from expressing the size of all (constantly many) ancestors $\mathcal{B}_u$ with $[1/\gamma] = j$ as a constant factor the size of $\mathcal{B}_u$. The second inequality comes from Item 1 of this lemma, which states that $O(|\mathcal{B}|) = O(|\mathcal{B}|)$ for every bag $\mathcal{B}$. The third inequality comes from Theorem 2. By Item 2, there are $O(n^{\lambda^2})$ such nodes $u$ in $\mathcal{T}$, hence the space required is $O(n^{3(k-1)})$.

We now turn our attention to the time requirement. For every bag $\mathcal{B}$, the algorithm requires $O(|\mathcal{B}|^2)$ time to iterate over all pairs of nodes $u$ and $x$ in Eq. (3) and Eq. (4) to compute the values $\text{FWD}_\mathcal{B}(\tau)$ and $\text{BWD}_\mathcal{B}(\tau)$ for every $\tau \in \nabla_\tau(\mathcal{B})$. Hence the time required for all nodes $u$ and one partial node $\tau \in \nabla_\tau(\mathcal{B})$ to store the maps values $\text{FWD}_\mathcal{B}(\tau)$ and $\text{BWD}_\mathcal{B}(\tau)$ is given by the recurrence

\begin{align*}
T_k(n) &\leq n^{2(k-1)} + 2^\lambda \cdot \mathcal{T}_k \left( n \cdot \left( 1 + \frac{\delta}{2} \right)^{\lambda-1} \right)
\end{align*}

The analysis of Eq. (5) and Eq. (6) of Lemma 4 gives $T_k(n) = O(n^{3(k-1)})$ for $k \geq 3$ and $T_2(n) = O(n^{2(\lambda^2)})$ (i.e., the above time recurrence is analyzed as the recurrence for $\mathcal{S}_k$ of Lemma 4). From the space analysis we have that there exist $O(n^{3(k-1)})$ partial nodes $\tau \in \nabla_\tau(\mathcal{B})$ for every node $u$ whose root bag is $\mathcal{B}$. Hence the total time for this step is $O(n^{3(k-1)})$ for $k \geq 3$, and $O(n^{2(\lambda^2)})$ for $k = 2$.

5. This step requires time linear in the size of $\mathcal{T}$ [41].

The desired result follows.

Algorithm ConcurQuery. In the query phase, ConcurAP answers distance queries using the algorithm ConcurQuery. We distinguish three cases, according to the type of the query.

1. Single-source query. Given a source node $u$, initialize a map data-structure $A : V \rightarrow \Sigma$, and initially set $A(v) = \text{FWD}_u(v)$ for all $v \in \mathcal{B}_u$, and $A(v) = \emptyset$ for all other nodes $v \in V \setminus \mathcal{B}_u$. Perform a BFS on $\mathcal{T}$ starting from $\mathcal{B}_u$, and for every
encyclopedia bag \( B \) and nodes \( x, v \in B \) with \( Lv(v) \leq Lv(x) \), set \( A(v) = A(v) \cup \{ A(x) \cap FwD_x(v) \} \). Return the map \( A \).

2. Pair query. Given two nodes \( u, v \in V \), find the LCA \( B \) of bags \( B_u \) and \( B_v \). Return \( \bigcup_{x \in B_u \cap V} (FwD^+_u(x) \cap FwD^+_v(x)) \).

3. Partial pair query. Given two partial nodes \( \pi, \tau \).
   a) If both \( \pi \) and \( \tau \) are strictly partial, return \( FwD_{\pi \tau} (\pi^2) \), else
   b) If \( \pi \) is strictly partial, return \( BwD^+_\pi (\pi^2) \), else
   c) Return \( FwD^+_\pi (\pi^2) \).

We thus establish the following theorem.

**Theorem 3.** Let \( G = (V, E) \) be a concurrent graph of \( k \) constant-treewidth graphs \((G_i)_{1 \leq i \leq k} \), and \( w_t : E \rightarrow \Sigma \) a weight function of \( G \). For any fixed \( \epsilon > 0 \), the data-structure ConcurAP correctly answers single-source and pair queries and requires:

1. Preprocessing time \( O(n^{3(k-1)}) \) if \( k \geq 3 \), and \( (b) O(n^{3\epsilon}) \) if \( k = 2 \).
2. Preprocessing space \( O(n^{2(k-1)}) \).
3. Single-source query time \( O(n^{2\epsilon}) \) if \( k \geq 3 \), and \( (b) O(n^{2\epsilon+1}) \) if \( k = 2 \).
4. Pair query time \( O(n^{5-\epsilon}) \).
5. Partial pair query time \( O(1) \).

Proof. The correctness of ConcurQuery for handling all queries follows from Lemma 1 and the properties of the preprocessing established in Lemma 3. The preprocessing complexity is stated in Lemma 5. The time complexity for the single-source query comes from the observation that ConcurQuery spends quadratic time in each encountered bag, and the result follows from the recurrence analysis of Eq. (6) in Lemma 4. The time complexity for the pair query follows from the \( O(1) \) time to access the LCA bag \( B \) of \( B_u \) and \( B_v \), and the \( O(|B|) = O(n^{k-1}) \) time required to iterate over all nodes \( x \in B \cap V \). Finally, the time complexity for the partial pair query follows from the \( O(1) \) time lookup in the constructed maps \( FwD, FwD^+ \) and \( BwD^+ \).

Note that a single-source query from a strictly partial node \( \pi \) can be answered in \( O(n^\epsilon) \) time by breaking it down to \( n^\epsilon \) partial pair queries.

The most common case in analysis of concurrent programs is that of two threads, for which we obtain the following corollary.

**Corollary 1.** Let \( G = (V, E) \) be a concurrent graph of two constant-treewidth graphs \( G_1, G_2 \), and \( w_t : E \rightarrow \Sigma \) a weight function of \( G \). For any fixed \( \epsilon > 0 \), the data-structure ConcurAP correctly answers single-source and pair queries and requires:

1. Preprocessing time \( O(n^{3\epsilon}) \).
2. Preprocessing space \( O(n^{3}) \).
3. Single-source query time \( O(n^{2\epsilon}) \).
4. Pair query time \( O(n) \).
5. Partial pair query time \( O(1) \).

Remark 3. In contrast to Corollary 1, the existing methods for handling even one pair query require heuristic time and quartic space [36, 50, 58, 74] by computing the transitive closure. While our improvements are most significant for algebraic path queries, they imply improvements also for special cases like reachability (expressed in Boolean semirings). For reachability, the complete preprocessing requires quartic time, and without preprocessing every query requires quadratic time. In contrast, with almost cubic preprocessing we can answer pair (resp., partial pair) queries in linear (resp. constant) time.

Note that Item 4 of ConcurPreprocess is required for handling pair queries only. By skipping this step, we can handle every (partial) pair query \( \pi, \tau \) similarly to the single source query from \( \pi \), but restricting the BFS to the path \( P : B_{\pi} \longrightarrow B_{\tau} \) and spending \( O(|B|^2) \) time for each bag \( B \) of \( P \). Recall (Theorem 2) that the size of each bag \( B \) in \( T \) (and thus the size of the corresponding bag \( B \) in \( T \)) decreases geometrically every \( \gamma \) levels. Then, the time required for this operation is \( O(|B|^{2\gamma}) = O(n^\gamma) \), where \( |B| \) is the bag of \( P \) with the smallest level. This leads to the following corollary.

**Corollary 2.** Let \( G = (V, E) \) be a concurrent graph of two constant-treewidth graphs \( G_1, G_2 \), and \( w_t : E \rightarrow \Sigma \) a weight function of \( G \). For any fixed \( \epsilon > 0 \), the data-structure ConcurAP (by skipping Item 4 in ConcurPreprocess) correctly answers single-source and pair queries and requires:

1. Preprocessing time \( O(n^3) \).
2. Preprocessing space \( O(n^{2\epsilon}) \).
3. Single-source query time \( O(n^{2\epsilon+1}) \).
4. Pair and partial pair query time \( O(n^\epsilon) \).

Finally, we can use ConcurAP to obtain the transitive closure of \( G \) by performing \( n^\epsilon \) single-source queries. The preprocessing space is \( O(n^{2\epsilon}) \) by Corollary 2, and the space of the output is \( O(n^\epsilon) \), since there are \( n^\epsilon \) pairs for the computed distances. Hence the total space requirement is \( O(n^n) \). The time requirement is \( O(n^{\epsilon+1}) \), since by Corollary 2, every single-source query requires \( O(n^{\epsilon+1}) \) time. We obtain the following corollary.

**Corollary 3.** Let \( G = (V, E) \) be a concurrent graph of two constant-treewidth graphs \( G_1, G_2 \), and \( w_t : E \rightarrow \Sigma \) a weight function of \( G \). For any fixed \( \epsilon > 0 \), the transitive closure of \( G \) wrt \( w_t \) can be computed in \( O(n^{\epsilon+1}) \) time and \( O(n^{3\epsilon}) \) space.

6. **Conditional Optimality for Two Graphs**

In the current section we establish the optimality of Corollary 2 in handling algebraic path queries in a concurrent graph that consists of two constant-treewidth components. The key idea is to show that for any arbitrary graph (i.e., without the constant-treewidth restriction) \( G \) of \( n \) nodes, we can construct a concurrent graph \( G' \) as a 2-self-concurrent asynchronous composition of a constant-treewidth graph \( G'' \) of \( 2 \cdot n \) nodes, such that semiring queries in \( G \) coincide with semiring queries in \( G' \).

**Arbitrary graphs as composition of two constant-treewidth graphs.** We fix an arbitrary graph \( G = (V, E) \) of \( n \) nodes, and a weight function \( w_t : E \rightarrow \Sigma \). Let \( x_i, 1 \leq i \leq n \) range over the nodes of \( V \), and construct a graph \( G'' = (V'', E'') \) such that \( V'' = \{ x_i, y_i : 1 \leq i \leq n \} \) and \( E'' = \{ (x_i, y_i), (y_i, x_i) : 1 \leq i \leq n \} \cup \{(y_{i+1}, y_i) : 1 \leq i < n \} \).

**Claim 2.** The treewidth of \( G'' \) is 1.

Given \( G'' \), we construct a graph \( G' \) as a 2-self-concurrent asynchronous composition of \( G'' \). Informally, a node \( x_i \) of \( G \) corresponds to the node \( \langle x_i, x_i \rangle \) of \( G' \). An edge \( \langle x_j, x_j \rangle \) in \( G \) is simulated by two paths in \( G' \).

1. The first path has the form \( P_1 : \langle x_i, x_i \rangle \longrightarrow \langle x_j, x_j \rangle \), and is used to witness the weight of the edge in \( G \), i.e., \( wt(\langle x_i, x_i \rangle) = \emptyset(P_1) \). It traverses a sequence of nodes, where the first constituent is fixed to \( x_i \), and the second constituent forms the path \( x_i \rightarrow y_i \rightarrow y_i' \rightarrow \cdots \rightarrow y_j \rightarrow x_j \). The last transition will have weight equal to \( wt(x_i, x_j) \), and the other transitions have weight 1.

2. The second path has the form \( P_2 : \langle x_i, x_j \rangle \longrightarrow \langle x_j, x_j \rangle \), it has no weight (i.e., \( \emptyset(P_2) = \emptyset \)), and is used to reach the node \( \langle x_j, x_j \rangle \). It traverses a sequence of nodes, where the second constituent is fixed to \( x_j \), and the first constituent forms the path \( x_i \rightarrow y_i \rightarrow y_i' \rightarrow \cdots \rightarrow y_j \rightarrow x_j \).
This leads to the following corollary.

**Conditional optimality of Corollary 2.** Note that for \( r = O(n) \) pair queries, Corollary 2 yields that the time spent by our data-structure ConcurAP for preprocessing \( G \) and answering \( r \) queries is \( T_{\text{ConcurAP}}(G, r) = O(n^3) \). The long-standing (over five decades) upper bound for answering even one pair query for algebraic path properties in arbitrary graphs of \( n \) nodes is \( O(n^7) \). Theorem 4 implies that any improvement upon our results would yield the same improvement for the long-standing upper bound, which would be a major breakthrough.

**Almost-optimality of Theorem 3 and Corollary 3.** Finally, we highlight some almost-optimality results obtained by variants of ConcurAP for the case of two graphs. By almost-optimality we mean that the obtained bounds are \( O(n^\epsilon) \) factor worse than optimal, for any fixed \( \epsilon > 0 \) arbitrarily close to 0.

1. According to Theorem 3, after \( O(n^{3+\epsilon}) \) preprocessing time, single-source queries are handled in \( O(n^{3+\epsilon}) \) time, and partial pair queries in \( O(1) \) time. The former (resp. later) query time is almost linear (resp. exactly linear) in the size of the output. Hence the former queries are handled almost-optimally, and the latter indeed optimally. Moreover, this is achieved using \( O(n^{3+\epsilon}) \) preprocessing time, which is far less than the \( O(n^5) \) time required for the transitive closure computation (which computes the distance between all \( n^2 \) pairs of nodes).

2. According to Corollary 3, the transitive closure can be computed in \( O(n^{3+\epsilon}) \) time, for any fixed \( \epsilon > 0 \), and \( O(n^3) \) space. Since the size of the output is \( \Theta(n^2) \), the transitive closure is computed in almost-optimal time and optimal space.

7. Experimental Results

In the current section we report on experimental evaluation of our algorithms, in particular of the algorithms of Corollary 3. We test their performance for obtaining the transitive closure on various concurrent graphs. We focus on the transitive closure for a fair comparison with the existing algorithmic methods, which compute the transitive closure even for a single query. Since the contributions of this work are algorithmic improvements for algebraic path properties, we consider the most fundamental representative of this framework, namely, the shortest path problem. Our comparison is done against the standard Bellman-Ford algorithm, which (i) has the best worst-case complexity for the problem, and (ii) allows for practical improvements, such as early termination.

**Basic setup.** We outline the basic setup used in all experiments. We use two different sets of benchmarks, and obtain the controlflow graphs of Java programs using Soot [71], and use LibTW [72] to obtain the tree decompositions of the corresponding graphs. For every obtained graph \( G' \), we construct a concurrent graph \( G' \) as a 2-self asynchronous composition of \( G' \), and then assign random integer weights in the range \([-10^3, 10^3]\), without negative cycles. Although this last restriction does not affect the running time of our algorithms, it allows for early termination of the Bellman-Ford algorithm (and thus only benefits the latter). The 2-self composition is a natural situation arising in practice, e.g. in concurrent datastructures where two threads of the same method access the datastructure. We note that the 2-self composition is no simpler than the composition of any two constant-treewidth graphs, (recall that the lower-bound of Section 6 is established on a 2-self composition).

**DaCapo benchmarks.** In our first setup, we extract controlflow graphs of methods from the DaCapo suit [8]. The average treewidth...
of the input graphs is around 6. This supplies a large pool of 120 concurrent graphs, for which we use Corollary 3 to compute the transitive closure. This allows us to test the scalability of our algorithms, as well as their practical dependence on input parameters. Recall that our transitive closure time complexity is $O(n^{2+\epsilon})$, for any fixed $\epsilon > 0$, which is achieved by choosing a sufficiently large $\lambda \in \mathbb{N}$ and a sufficiently small $\delta \in \mathbb{R}$ when running the algorithm of Theorem 1. We compute the transitive closure for various $\lambda$. In practice, $\delta$ has effects only for very large input graphs. For this, we fix it to a large value ($\delta = 1/3$) which can be proved to have no effect on the obtained running times. Table 2 shows for each value of $\lambda$, the percentage of cases for which that value is at most 5% slower than the smallest time (among all tested $\lambda$) for each examined case. We find that $\lambda = 7$ works best most of the time.

Figure 8 shows the time required to compute the transitive closure on each concurrent graph $G$. Our algorithm is run for $\lambda = 7$. TO denotes that the computation timed out after 30 minutes.

Figure 8 shows the time required to compute the transitive closure on each concurrent graph $G$ by our algorithm (for $\lambda = 7$) and the baseline Bellman-Ford algorithm. We see that our algorithm significantly outperforms the baseline method. Note that our algorithm seems to scale much better than its theoretical worst-case bound of $O(n^{2+\epsilon})$ of Corollary 3.

Concurrency with locks. Our second set of experiments is on methods from containers of the java.util.concurrent library that use locks as their synchronization mechanism. The average treewidth of the input graphs is around 6. In this case, we expand the node set of the concurrent graph $G$ with the lock set $[3]^\ell$, where $\ell$ is the number of locks used by $G$. Intuitively, the $i$-th value of the lock set denotes which of the two components owns the $i$-th lock (the value is 3 if the lock is free). Transitions to nodes that perform lock operations are only allowed wrt the lock semantics. That is, a transition to a node of $G$ where the value of the $i$-th lock is

1. (Lock acquire): $j \in [2]$, is only allowed from nodes where the value of that lock is 3, and the respective graph $G_j$ is performing a lock operation on that edge.
2. (Lock release): 3, is only allowed from nodes where the value of that lock is $j \in [2]$, and the respective graph $G_j$ is performing an unlock operation on that edge.
References


