Algorithms for algebraic path properties in concurrent systems of constant treewidth components

Krishnendu Chatterjee, Rasmus Ibsen-Jensen, Amir Goharshady, Andreas Pavlogiannis


Deposited at: 12 Dec 2018 11:54    ISSN: 2664-1690
Algorithms for Algebraic Path Properties in Concurrent Systems of Constant Treewidth Components

Krishnendu Chatterjee and Rasmus Ibsen-Jensen and Amir Goharshady and Andreas Pavlogiannis

Technical Report No. IST-2015-340-v1+1
Deposited at 15 Jul 2016 09:12
https://repository.ist.ac.at/340/1/main.pdf
Algorithms for Algebraic Path Properties in Concurrent Systems of Constant Treewidth Components

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 properties that arise in program analysis. We consider that each component of the concurrent system is a graph with constant treewidth, and it is known that the controlflow graphs of most programs have constant treewidth. We allow for multiple possible queries, which arise naturally in demand driven dataflow analysis problems (e.g., alias 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 approaches construct the product graph of all components and apply the best-known graph algorithm on the product. In the traditional approach, even the answer to a single query requires the transitive closure computation (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 that show that the worst-case running times of our algorithms cannot be improved without achieving major breakthroughs in graph algorithms (such as improving the worst-case bounds for the shortest path problem in general graphs whose current best-known bound has not been improved in five decades). Finally, we provide a prototype implementation of our algorithms which significantly outperforms the existing algorithmic methods on several benchmarks.

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

General Terms Algorithms, Languages

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 previous known algorithms. We establish conditional optimality results for some of our algorithms in the sense that they cannot be improved without achieving major breakthrough in the algorithmic study of graph problems. We also present experimental results which show that our algorithms perform favorably for several real-world 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 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) \[65\]. The treewidth of a graph is defined based on a tree decomposition of the graph \[62\], 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 \[68\], and it was also shown in \[38\] 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 regular expression of reaching from one node to another. The algebraic path properties subsume the dataflow analysis of the IFDS/IDE frameworks \[64\] \[66\] 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. A well-studied case is that of dataflow analysis of distributive flow functions [23, 26, 32, 63, 66]. Here we elaborate on two other important classes. For a detailed discussion, see Appendix A.

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, 35, 55, 64, 72], phrased on the tropical semiring. For example, the edge weights (positive and negative) can express energy consumption, 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, 24]. 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, 22, 47, 55, 71, 15].

2. **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, 30, 33]. Regular expressions have also been used as algebraic relaxations of interprocedurally valid paths in sequential and concurrent systems [14, 22].

**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 \( v \). Since we consider the concurrent setting, 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 \( p \) 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 components. 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, 23, 29, 33, 58, 62, 63, 66, 75, 76]. Such analysis has several advantages, such as (quoting from [45, 63]) (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, e.g., 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: (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 no preprocessing is done, and 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 programs: typically the controlflow graphs of programs have constant treewidth [17, 30, 33], 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-Kleene [35, 49, 56, 72] style dynamic programming, which requires cubic time. Two well-known special cases of the algebraic path 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 \( O(n \cdot m) \) (where \( n \) is the number of vertices and \( m \) is the number of edges, and since \( m \) can be \( n^2 \), the worst-case bound is \( O(n^3) \)). In the second case, the well-known construction of Kleene’s [49] 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 size \( n \) and there are \( k \) components, then the product graph has size \( O(n^k) \)), and then apply the best-known graph algorithm (thus the overall time complexity is \( O(n^{k+1}) \)). Hence for the important special case of two components we obtain a hexic-time (i.e., \( O(n^6) \)) algorithm. Moreover, for algebraic path properties (even for the special case of shortest path) the current best-known algorithms for one pair query (or one single-source query) computes the entire transitive closure. Hence the existing approach does not allow a tradeoff of preprocessing and query as even for one query the entire transitive closure is computed.

**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.

1. **Improved upper bounds.** We present improved upper bounds both for general \( k \) components, and the important special case of two components.

   - **General case.** We show that for \( k \geq 3 \) components with \( n \) nodes each, after \( O(n^{3\cdot(k-1)}) \) preprocessing time, we can answer (i) single-source queries in \( O(n^{2\cdot(k-1)}) \) 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 \( O(n^{2\cdot(k-1)}) \) space. In contrast,
Table 1: The algorithmic complexity for computing algebraic path queries wrt a closed, complete semiring on a concurrent graph. In contrast, the current best solution requires \( O(n^3) \) time and \( O(n^4) \) space, whereas our results achieve \( \Omega(n^4) \) time and \( \Omega(n^6) \) space. Our results are summarized in Table 1 and the tradeoff is pictorially illustrated in Figure 1.

### 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 that show that our combined preprocessing and query time cannot be improved without achieving a major breakthrough in graph algorithms.

#### Almost optimality
First, note that for 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 \( \Theta(n^2) \). Moreover, the result is achieved with preprocessing time less than \( \Omega(n^4) \), which is a lower bound for computing the transitive closure (which has \( n^4 \) entries). Furthermore, for our third result (obtained from Corollary 3) the \( O(n^4) \) preprocessing time is arbitrarily close to optimal, and the \( O(n^4) \) preprocessing space is indeed optimal, as the transitive closure computes the distance among all \( n^2 \) pairs of nodes (which requires \( O(n^4) \) time and space).

#### Conditional lower bound
In recent years, the conditional lower bound problem has received vast attention in complexity theory, where under the assumption that certain problems (such as matrix multiplication, all-pairs shortest path) cannot be solved faster than the existing upper bounds, lower bounds for other problems (such as dynamic graph algorithms) are obtained [11][12][42]. The current best-known algorithm for algebraic path properties for general (not constant-treewidth) graphs is cubic in the number of nodes. Even for the special case of shortest paths with positive and negative weights, the best-known algorithm (which has not been improved over five decades) is \( O(n^3) \), where \( m \) is the number of edges. Since \( m \) can be \( \Theta(n^2) \), the current best-known worst-case complexity is cubic in the number of nodes. We prove that pair queries require more time in a concurrent graph of two constant-treewidth graphs, with \( n \) nodes each, than in general graphs with \( n \) nodes. This implies that improving the \( O(n^3) \) combined preprocessing and query time over our result (from Corollary 5) for answering \( r \) queries, for \( r \leq O(n) \), would yield the same improvement over the \( O(n^3) \) time for answering \( r \) pair queries in general graphs.

That is, the combination of our preprocessing and query time (from Corollary 5) cannot be improved without equal improvement on the long standing cubic bound for the shortest path and the algebraic path problems in general graphs. Additionally, our result (from Theorem 3) cannot be improved much further even for \( n^2 \) queries, as the combined time for preprocessing and answering \( n^2 \) queries is \( \Omega(n^3) \).
is $O(n^{3.5})$ using Theorem 3, while the existing bound is $O(n^3)$ for general graphs.

3. Experimental results. We provide a prototype implementation of our algorithms which significantly outperforms the baseline methods on several benchmarks.

Technical contributions. The results of this paper rely on several novel technical contributions.


(a) The first key result for our upper bounds 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 $\beta = ((1 + \delta)/2)^{\lambda-1} \gamma = \lambda$.

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 a 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 $G_i$, in Section 5, we present the algorithms for handling algebraic path queries. In particular, we introduce the partial expansion $\tilde{G}$ of $G$ for additionally handling partial queries, and describe the algorithms for preprocessing and querying $\tilde{G}$ 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 B. Several technical proofs are also relegated to the appendix due to lack of space.

1.1 Related Works

Treewidth of graphs. The notion of treewidth for graphs as an elegant mathematical tool to analyze graphs was introduced in [63]. The significance of constant treewidth in graph theory is huge 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 $t$ [9, 11]. Even for problems that can be solved in polynomial time, faster algorithms can be obtained for low treewidth graphs, for example, for the distance problem [25]. 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 [27] showed that for constant-treewidth graphs the decision questions for MSO can be solved in polynomial time; and the result of [31] 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, 59]. The problem of computing a balanced tree decomposition for a constant treewidth graph was considered in [61]. More importantly, in the context of programming languages, it was shown by [68] that the controlflow graphs of goto-free programs from many programming languages have constant treewidth. This theoretical result was subsequently followed up in several practical approaches, and although in the presence of goto’s the treewidth is not guaranteed to be bounded, it has been shown that programs in several programming languages have low treewidth in practice ([17, 35]). The constant-treewidth property of graphs has been used to develop faster algorithms for sequential interprocedural analysis [23], and on the analysis of automata with auxiliary storage (e.g., stacks and queues) [57]. The techniques of [23] do not extend to the concurrent setting because even though the component graphs have constant treewidth, the graph which is the product of the components has treewidth that depends on the number of nodes. Thus none of the previous works exploit the constant-treewidth property for faster algorithms in the concurrent setting.

Concurrent system analysis. The problem of concurrent system analysis has been considered in several works, both for intraprocedural as well context-bound interprocedural analysis [13, 15, 33, 41, 45, 51, 53, 60], and many practical tools have been developed as well [22, 54, 60, 67]. In this work we focus on the intraprocedural analysis with constant-treewidth graphs, and present algorithms with better asymptotic complexity. 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 $X$, 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 associative, commutative, and $0$ is the neutral element,
2. $\otimes$ is associative, and $1$ is the neutral element,
3. $\otimes$ distributes over $\oplus$,
4. $\oplus$ is infinitely associative,
5. $\otimes$ infinitely distributes over $\oplus$,
6. $0$ absorbs in multiplication, i.e., $\forall a \in \Sigma : a \otimes 0 = 0$.

Additionally, we consider that $S$ is equipped with a closure operator $\mathcal{C}$, such that $\forall s \in \Sigma : \mathcal{C}s = \mathcal{C}(s \otimes \mathcal{C}s) = \mathcal{C}(\mathcal{C}s \otimes s)$ (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 $w : E \rightarrow \Sigma$ 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 $P : u \longrightarrow v$ 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 $P$ 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 $\bigotimes(P) = \bigotimes(wt(x_i, x_{i+1}))$, if $|P| \geq 1$ else $\bigotimes(P)$ is $\mathbb{I}$.

Given nodes $u, v \in V$, the semiring distance $d(u, v)$ is defined as $d(u, v) = \bigoplus_{P : u \longrightarrow v} \bigotimes(P)$, and $d(u, v) = \mathbb{U}$ if no such $P$ exists.

Trees. A tree $T = (V, E)$ is an undirected graph with a root node $u_0$, such that between every two nodes there is a unique path. For a node $u$ we denote by $Lv(u)$ the level of $u$ which is defined as the length of the simple path from $u_0$ 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 $u$, any node (including itself) that appears in the path from $u_0$ to $u$ is an ancestor of $u$, and if it is a child 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 subgraph $(X, E')$ of $T$ such that $E' = E \cap (X \times X)$ and for every pair $u, v \in X$, 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 has a child has exactly 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 nodes of $G$ such that the following conditions hold:

1. $\forall T \in V_T, |T| = |B_T|$, the number of bags in $T$. Conventionally, we call $B_0$ the root of $T$, and denote by $Lv(B)$ the level of $B_i$ in $T$. A bag $B$ is a bag at level $i$ if $B \in T$ and denote by $Lv(B)$ the level of $B_i$ in $T$. For a bag $B$ in $T$, we denote by $T[B]$ the subtree of $T$ rooted at $B$. A bag $B$ is called a root bag of a node $u$ if $u \in B$ and every $B'$ that contains $u$ appears in $T[B]$. We often use $B(u)$ to refer to the root bag of $u$, and define $Lv(u) = Lv(B(u))$. Given a bag $B$, we denote by $V_T(B)$ the nodes of $G$ that appear in bags in $T(B)$.

2. $\forall T \in V_T$, the nodes of $G$ that appear in $B$ in $T(B)$.

The width of the tree-decomposition $T$ is the size of the largest bag minus 1. The treewidth $t$ 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 $|T[B]| \leq (t + 1) \cdot |T(B)|$. Given a graph $G$ with treewidth $t$ and a fixed $\alpha \in \mathbb{N}$, a tree-decomposition $T(G)$ is called $\alpha$-approximate if it has width at most $\alpha \cdot (t + 1) - 1$. Figure 2 illustrates the above definitions on a small example.

2.3 Concurrent graphs

Product graphs. A graph $G_{\prod} = (V_{\prod}, E_{\prod})$ is said to be the product graph of $k$ graphs $(G_i = (V_i, E_i))_{1 \leq i \leq k}$ if $V_{\prod} = \bigsqcup V_i$, and $E_{\prod}$ is such that for all $u, v \in V_{\prod}$ with $u = (u_i)_{1 \leq i \leq k}$ and $v = (v_i)_{1 \leq i \leq k}$.

Figure 2: A graph $G$ with treewidth 2 (left) and a corresponding tree-decomposition $T = \text{Tree}(G)$ of 8 bags and width 2 (right).

The distinguished bag $B_0$ is the root bag of node 6. We have $V_T(B_0) = \{6, 7, 9, 4, 5\}$ and $V_T(B_6) = \{6, 7, 9, 8, 10, 2\}$. The subtree $T(B_6)$ is shown in bold.

Concurrent graphs. A graph $G = (V, E)$ is called a concurrent graph of $k$ graphs $(G_i = (V_i, E_i))_{1 \leq i \leq k}$ if $V = \bigsqcup V_i$ and $E_{\prod} = E_{\prod}$, where $G_{\prod} = (V_{\prod}, E_{\prod})$ is the product graph of $(G_i)_{1 \leq i \leq k}$. Given a concurrent graph $G = (V, E)$ and 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 $G'$ if $G_{\prod} = G'$ is the k-self-product of $G'$.

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, various well-known composition notions can be modeled as follows. For any edge $(u, v) \in E$ of the concurrent graph $G$, let $I_{u,v} = \{i \in \{k\} : (u_i, v_i) \in E_i\}$ denote the components that execute a transition in $(u, v)$.

1. In synchronous composition at every step all components make one move each simultaneously. This is captured by $I_{u,v} = \{k\}$.

2. In asynchronous composition at every step only one component makes a move. This is captured by $|I_{u,v}| = 1$ 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 $\prod \{(V_i \cup \{\bot\})\}$, where $\bot \notin \bigcup V_i$. 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 $\overline{\pi} \sqsubseteq \pi$ if for all $1 \leq i \leq k$ either $\pi_i = \bot$ or $\overline{\pi}_i = \pi_i$. We say that the partial node $\pi$ strictly refines $\pi_i$, denoted by $\pi_i \sqsubset \pi$, if $\pi_i \sqsubset \pi$ and $\pi_i \neq \pi$ (i.e., for at least one constituent $i$ we have $\overline{\pi}_i = \bot$ but $\pi_i \neq \bot$). A partial node $\pi$ is said to be strictly partial if it is strictly refined by some node $u \in V$ (i.e., $\pi$ has at least one $\bot$). The notion of semiring distances is extended to partial nodes, and for partial nodes $\pi, \pi'$ of $G$ we define the semiring distance from $\pi$ to $\pi'$ as $d(\pi, \pi') = \bigoplus_{u \in \mathbb{P}, v \in \mathbb{P}} d(u, v)$ 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
Figure 3: A concurrent program (left), its control flow graph (middle), and a tree decomposition of the control flow graph (right).

case as actual nodes, and write $u$ (i.e., without the bar). Distances where one endpoint is a strictly partial node $\pi$ succinctly quantify over all nodes of all the components for which the corresponding constituent of $\pi$ is $\perp$. 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 $\mathcal{wt} : E \to \Sigma$ be a weight function that assigns to every edge of $G$ a weight from a set $\Sigma$ that forms a complete, closed semiring $\Sigma = (\Sigma, \otimes, \oplus, 0, 1)$. The algebraic path problem on $G$ asks the following types of queries:

1. Single-source query. Given a partial node $\pi$ of $G$, return the distance $d(\pi, v)$ to every node $v \in V$. When the partial node $\pi$ is an actual node of $G$, we have a traditional single-source query.
2. Pair query. Given two nodes $u, v \in V$, return the distance $d(u, v)$.
3. Partial pair query. Given two partial nodes $\pi, \tau$ of $G$ where at least one is strictly partial, return the distance $d(\pi, \tau)$.

Figure 3 presents the notions introduced in this section on a toy example on the dining philosophers problem. In Appendix A we discuss the vast modeling capabilities of the algebraic paths framework, and illustrate the importance of pair and partial pair queries in the analysis of the dining philosophers program.

**Input parameters.** For technical convenience, we consider a uniform upper bound $n$ on the number of nodes of each $G_i$ (i.e., $|V_i| \leq n$). Similarly, we let $t$ be an upper bound on the treewidth of each $G_i$. The number $k$ is taken to be fixed and independent of $n$. The input of the problem consists of the graphs $(G_i)_{1 \leq i \leq k}$, together with some representation of the edge relation $E$ of $G$.

**Complexity measures.** The complexity of our algorithms is measured as a function of $n$. In particular, we do not consider the size of the representation of $E$ when considering the size of the input. This has the advantage of obtaining complexity bounds that are independent of the representation of $E$, which be represented implicitly (such as synchronous or asynchronous composition) or explicitly, depending on the modeling of the problem under consideration. The time complexity of our algorithms is measured in number of operations, with each operation being either a basic machine operation, or an application of one of the operations of the semiring.

### 3. Strongly Balanced Tree Decompositions

In this section we introduce the notion of strongly balanced tree decompositions, and present an algorithm for computing them efficiently on constant-treewidth graphs. The following sections make use of this construction.

**Strongly balanced tree decompositions.** Given a tree-decomposition $T$ and constants $0 < \beta < 1, \gamma \in \mathbb{N}^+$, a bag $B$ of $T$ is called $(\beta, \gamma)$-balanced if for every descendant $B_i$ of $B$ with $\text{Lb}(B_i) - \text{Lb}(B) = \gamma$, we have $|T(B_i)| \leq \beta \cdot |T(B)|$, i.e., the number of bags in $T(B_i)$ is at most a $\beta$-fraction of those in $T(B)$. A tree-decomposition $T$ is called a $(\beta, \gamma)$ tree-decomposition if every bag of $T$ is $(\beta, \gamma)$-balanced. A $(\beta, \gamma)$ tree-decomposition that is $\alpha$-approximate is called an $(\alpha, \beta, \gamma)$ tree-decomposition. The following theorem is central to the results obtained in this paper. The proof is technical and relegated to Appendix B and here we provide a sketch of the algorithm for obtaining it.

**Theorem 1.** For every graph $G$ with $n$ nodes and constant treewidth, for any fixed $\delta > 0$ and $\lambda \in \mathbb{N}$ with $\lambda \geq 2$, let $\alpha = 4 \cdot \lambda / \delta$, $\beta = ((1 + \delta) / 2)^{\lambda - 1}$, and $\gamma = \lambda$. A binary $(\alpha, \beta, \gamma)$ tree-decomposition $\text{Tree}(G)$ with $O(n)$ bags can be constructed in $O(n \cdot \log n)$ 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., in $O(n)$ time). Given two parameters $\delta > 0$ and $\lambda \in \mathbb{N}$ with $\lambda \geq 2$. $\text{Tree}(G)$ is turned to an $(\alpha, \beta, \gamma)$ tree-decomposition, for $\alpha = 4 \cdot \lambda / \delta$, $\beta = ((1 + \delta) / 2)^{\lambda - 1}$, and $\gamma = \lambda$, in two conceptual steps.

1. A tree of bags $R_0$ is constructed, which is $(\beta, \gamma)$-balanced.
2. $R_0$ 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
treewidth graphs is linear time \[12\], it is NP-complete for general graphs.

**Remark 1** will be chosen appropriately in later sections.

**Tree** 4. Concurrent Tree Decomposition

those that have only one child, while increasing the size of

\[ \delta \] fraction the size of \[ \delta \]. The union of

\[ \lambda \] is partitioned into two components

\[
\left\{ \begin{array}{l}
\lambda
\end{array} \right. \]

1. If \( \ell > 0 \), then \( C \) is partitioned into components \( \mathcal{Y} = \{ C_1, \ldots, C_r \} \), by removing a list of bags \( \mathcal{X} = \{ B_1, \ldots, B_m \} \), such that \( |C_i| \leq \frac{\ell}{2} \cdot |C| \). The union of \( \mathcal{X} \) yields a new bag \( B \) in \( R_\ell \). Then \( \mathcal{Y} \) is merged into two components \( C_1, C_2 \) with \( |C_i| \leq |C_2| \leq \frac{\ell}{2} \cdot |C| \). Finally, each \( C_\ell \) 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 \( C_1, C_2 \) such that \( |\text{Nh}(C_1) \cap \text{Nh}(C_2)| \leq \frac{|\text{Nh}(C)|}{m} \) by removing a single bag \( B \). This bag becomes a new bag \( B \) in \( R_\ell \), and each \( C_\ell \) is passed on to the next recursive step with \( \ell = (\ell + 1) \mod \lambda \).

**Use of** \((\alpha, \beta, \gamma)\) **tree-decompositions.** For ease of presentation we consider that every \( \text{Tree}(G) \) is a full binary tree. Let \( \alpha = 4 \cdot \lambda \beta \cdot \lambda / \beta + 1 \) and \( \gamma = \lambda \). Since our tree decompositions are \((\beta, \gamma)\)-balanced, we can always attach empty children bags to those that have only one child, while increasing the size of \( \text{Tree}(G) \) by a constant factor only. In the sequel, \( \text{Tree}(G) \) will denote a full binary \((\alpha, \beta, \gamma)\) tree-decomposition of \( G \). The parameters \( \delta \) and \( \lambda \) will be chosen appropriately in later sections.

**Remark 1.** The notion of balanced tree decompositions exists in the literature [13,31], 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. 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 \( O(n^{k-1} \cdot \beta) \); and (ii) for every bag in level at least \( i - \gamma \), the size of the bag is \( O(n^{k-1} \cdot \beta) \).

**Algorithm** \( \text{ConcurTree} \) for concurrent tree decomposition. Let \( G \) be a concurrent graph of \( k \) of graphs \( (G_i)_{1 \leq i \leq k} \). The input consists of a full binary tree-decomposition \( T_\ell \) of \( G \). In the following, \( B_1 \) ranges over bags of \( T_\ell \), and we denote by \( B_{i,r} \), with \( r \in [2^\ell] \), 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 ConcurTree, which operates as follows. On input \( (T_i(B_i))_{1 \leq i \leq k} \), return a tree decomposition where

1. The root bag \( B \) is

\[
B = \bigcup_{1 \leq i \leq k} \left( \left( \prod_{j=1}^{m} V_{T_j}(B_j) \right) \times B_i \times \left( \prod_{j=1}^{m} V_{T_j}(B_j) \right) \right)
\]

\[(1)\]

2. If every \( B_i \) is a non-leaf bag of \( T_i \), for every choice of \( \{ T_{i,1}, \ldots, T_{i,r_i} \} \in [2^\ell]^k \), repeat the procedure for \( (T_i(B_{i,r_i}))_{1 \leq i \leq k} \), and let \( B' \) be the root of the returned tree. Make \( B' \) a child of \( B \).

Let \( B_i \) be the root of the tree-decomposition \( T_i \). We denote by \( \text{ConcurTree}(G) \) the application of the recursive procedure ConcurTree on \( (T_i(B_i))_{1 \leq i \leq k} \). Figure 5 provides an illustration.

**Remark 2.** Recall that for any bag \( B_j \) of a tree-decomposition \( T_j \), we have \( V_{T_j}(B_j) = \bigcup_{j \neq i} B_j^i \), where \( B_j^i \) ranges over bags in \( T_i(B_j) \). Then, for any two bags \( B_i, B_j \), of tree-decompositions \( T_i \) and \( T_j \) respectively, we have

\[
\forall_{B_i^j, B_j^j} \left( B_i^j \times B_j^j \right) = \bigcup_{B_i^j, B_j^j} (B_i^j \times B_j^j)
\]

where \( B_i^j \) and \( B_j^j \) range over bags in \( T_i(B_i) \) and \( T_j(B_j) \) respectively. Since each tree-decomposition \( T_i \) has constant width, it follows that \( |V_{T_i}(B_i) \times V_{T_j}(B_j)| = O(|T_i(B_i)| \cdot |T_j(B_j)|) \). Thus, the size of each bag \( B \) of \( \text{ConcurTree}(G) \) constructed in Eq. (1) on some input \( (T_i(B_i))_{1 \leq i \leq k} \), is \(|B| = O(\sum_{j \neq i} n_j)\), where \( n_i = |T_i(B_i)| \).

In view of Remark 2 the time and space required by ConcurTree to operate on input \( (T_i(B_i))_{1 \leq i \leq k} \) where \( |T_i(B_i)| = n_i \) is given, up to constant factors, by
such that for every $i$ we have that $\sum_{j \in [2]} n_i x_j \leq n_i$. 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 $(\beta, \gamma, \text{balanced})$-balanced, then the size of each bag $B$ of ConcurTree(G) decreases geometrically with its level $\text{Lv}(B)$. See Appendix C for the formal proof.

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

5. Concurrent Algorithmic 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-tree-width graphs $(G_i)_{1 \leq i \leq k}$. To this direction, we develop a data-structure ConcurAP for (concurrent algorithmic paths) which will preprocess $G$ and afterwards support single-source, pair, and partial pair queries in $G$.

**Semiring distances on tree decompositions.** The preprocessing and query of our data-structure exploits a key property of semiring distinct bags $B^d$:

$$T(n_1, \ldots, n_k) \leq \sum_{i \leq k} \sum_{r \in \mathbb{N}} n_i \cdot T(n_i, r_1, \ldots, n_k, r_k)$$

for semiring multiplication) to all new edges in $G$, and two copies for every partial node $\pi$ that is strictly partial.

To this end, we have $|B| = \mathcal{F}(\pi \cdot \pi')$, and a weight function $w: \pi \rightarrow \mathcal{F}(\pi \cdot \pi')$ for all new edges in $G$.

**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 $0 < \epsilon < 1$, choose a $\lambda \in \mathbb{N}$ and $\delta \in \mathbb{R}$ such that $\lambda \geq 2$ and $\delta \leq \frac{2}{|V|}$. Let $\alpha = 4 - \lambda \cdot \beta$, $\beta = ((1 + \delta) / \lambda)^{\lambda - 1}$ and $\gamma = \lambda$. On input $G = (V, E)$, where $G$ is a concurrent graph of $k$ constant-tree-width graphs $(G_i = (V_i, E_i))_{1 \leq i \leq k}$, and a weight function $w: E \rightarrow \mathcal{F}(\pi \cdot \pi')$ ConcurPreprocess operates as follows:

1. Construct the partial expansion $\mathcal{U} = (\overline{V}_t, \overline{E}_t)$ of $G$ as follows.

(a) The node set is $\overline{V} = V \cup \left\{ \overline{v}, \overline{u} \mid \exists u \in V \text{ s.t. } u \in \overline{\pi} \right\}$; i.e., $\overline{V}$ consists of nodes in $V$ and two copies for every partial node $\pi$ that is strictly refined by a node $u$ of $G$.

(b) The edge set is $\overline{E} = E \cup \left\{ (\overline{u}, \overline{v}) \mid \forall u, v \in V \text{ s.t. } u \in \overline{\pi}, v \in \overline{\pi}' \right\}$, 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 $\overline{w}(\overline{v}, \overline{v}) = w(\pi \cdot \pi')$ if $\overline{\pi}, \overline{\pi}' \in \overline{V}$, and $\overline{w}(\overline{v}, \pi) = \mathcal{F}(\pi)$ otherwise. That is, the original weight function is extended with value $\mathcal{F}(\pi)$ (which is neutral for semiring multiplication) to all new edges in $\mathcal{U}$.

2. Construct the tree-decomposition $T = (\overline{V}_t, \overline{E}_t)$ of $\mathcal{U}$ 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 $\mathcal{T}$ be identical to $T_i$ with the following exception: For every bag $B$ of $T$ and $\overline{B}$ the corresponding bag in $\mathcal{T}$, for every node $u \in B$, insert in $\overline{B}$ all strictly partial nodes $\overline{\pi}'$, $\overline{v} \in \overline{V}$ that $u$ refines. Formally, set $\overline{B} = B \cup \left\{ \overline{v} \mid \exists u \in B \text{ s.t. } u \in \overline{\pi} \right\}$. Note that also $\overline{B}$.

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

(a) Traverse $\mathcal{T}$ bottom-up, and for every bag $\overline{B}$, execute an all-pairs algebraic path computation on $\overline{\mathcal{G}(\overline{B})}$ with weight function $w_{\overline{\pi}, \overline{\pi'}}$. This is done using classical algorithms, e.g., [13, 49, 50, 52]. For every pair of partial nodes $\overline{\pi}, \overline{\pi'}$ with $\text{Lv}(\overline{\pi}) \leq \text{Lv}(\overline{\pi'})$, set $\text{BWD}_{\overline{\pi}}(\overline{\pi}) = d(\overline{\pi} \cdot \pi)$ and $\text{FWD}_{\overline{\pi}}(\pi) = \text{BWD}_{\overline{\pi}}(\pi)$. 

Figure 6: Illustration of Lemma II. If $P$ is the unique simple path $B_1 \leftrightarrow B_3$ in Tree(G), then there exist (not necessarily distinct) $x_i \in B_i - 1 \cap B_i$ with $1 < i < 4$ such that $d(u, v) = d(u, x_2) \otimes d(x_2, x_3) \otimes d(x_3, x_4) \otimes d(x_4, v)$.
For the algorithm processes a bag $\mathcal{B}$. All cases depend on Lemma 3.

**Item 5.** For every pair of partial nodes $\pi, \pi' \in \mathcal{B}$ such that $L_C(\pi, \pi') \leq L_C(\pi')$ we have $FWD^u_\pi(\pi') = FWD^u_\pi(\pi') + wt^+(\pi',\pi)$ and $BWD^u_\pi(\pi') = BWD^u_\pi(\pi') + wt^+(\pi',\pi)$. The statement follows by a straightforward induction on the levels processed by the algorithm in the bottom-up pass. Note that if $\pi$ and $\pi'$ are partial nodes in the root of $\mathcal{T}$, the statement yields $FWD^u(\pi') = d(u, v)$ and $BWD^u(\pi') = d(u, v)$.

**Item 7.** The invariant is similar to the previous, except that $P_1$ and $P_2$ range over all $\pi \leftarrow \pi'$ and $\pi' \leftarrow \pi$ paths in $\mathcal{T}$ respectively. Hence now $FWD^u(\pi') = d(u, \pi)$ and $BWD^u(\pi') = d(\pi, \pi')$. 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 $\mathcal{B}$ and $\mathcal{W}$ store actual distances.

**Item 9.** For every node $u \in \mathcal{B}$ and partial node $\pi \in \mathcal{F}(\mathcal{B})$ we have $FWD^u_\mathcal{B}(\pi) = d(u, \pi)$ and $BWD^u_\mathcal{B}(\pi) = d(\pi, \pi')$. 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 7. Similarly for statement 2, together with the observation that every strictly partial node $\pi$ appears in the root of $\mathcal{T}$, and thus $\pi \in \mathcal{F}(\mathcal{B})$. 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 recurrence relations.

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

\[ T_k(n) \leq n^{3(k-1)} + 2^{k} \cdot T_k \left( n \left( \frac{1 + \delta}{2} \right)^{\lambda - 1} \right) \]  

\[ S_k(n) \leq n^{2(k-1)} + 2^{k} \cdot S_k \left( n \left( \frac{1 + \delta}{2} \right)^{\lambda - 1} \right) \]  

Then
1. $T_k(n) = O(n^{3(k-1)})$, and
2. $S_k(n) = O(n^{2(k-1)})$.

The proof of Lemma 4 is technical, and presented in Appendix 2. The following lemma analyzes the complexity of ConcurPreprocess, and makes use of the above recurrences.

**Lemma 5.** ConcurPreprocess requires $O(n^{k-1})$ space and 1. $O(n^{3(k-1)} \cdot \delta)$ time if $k = \delta$, and 2. $O(n^{k+1})$ time if $k \geq 3$.

**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 $\mathcal{T}$, which is $2 \cdot \sum_{i=k}^{\infty} (n^i) = O(n^{k+1})$.
2. By Theorem 2, ConcurTree($\mathcal{G}$) is constructed in $O(n^k)$ time and space. In $\mathcal{B}$, the size of each bag $\mathcal{B}$ is increased by constant factor, hence this step requires $O(n^k)$ time and space.
3. In each pass, ConcurPreprocess spends $|\mathcal{B}|^4$ time to perform an all-pairs algebraic paths computation in each bag $\mathcal{B}$ of $T$. The space usage for storing all maps $\text{FWD}_u$ and $\text{BWD}_u$ for every node $\pi$ whose root bag is $\mathcal{B}$ is $O(|\mathcal{B}|^2)$, since there are at most $|\mathcal{B}|$ such nodes $\pi$, and each map has size $|\mathcal{B}|$. By the previous item, we have $|\mathcal{B}| = O(|B|)$, where $B$ is the corresponding bag of $T$ before the partial expansion of $G$. By Theorem 3.3, we have $|B| = O(n^{k-1} \cdot 2^\delta)$, where $\mathcal{L}(B) = \gamma - (\gamma + 1)$ and $\beta = ((1+\delta)/2)^{\lambda-1}$. Then, since $T$ is a full $2^k$-ary tree, the time and space required for preprocessing every $\gamma$-level of $T$ is given by the following recurrences respectively (ignoring constant factors for simplicity).

$$T_k(n) \leq n^{2(k-1)} + 2^{\lambda \cdot k} \cdot T_k(n \cdot (1 + \frac{\delta}{2})^{\lambda-1})$$

$$S_k(n) \leq n^{2(k-1)} + 2^{\lambda \cdot k} \cdot S_k(n \cdot (1 + \frac{\delta}{2})^{\lambda-1})$$

By the analysis of Eq. 3 and Eq. 4 of Lemma 4.1, we have that $T_k(n) = O(n^{2(k-1)})$ and $S_k(n) = O(n^{2(k-1)})$ if $k \geq 3$, and (ii) $S_k(n) = O(n^{2(k-1)})$.

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

$$|\mathcal{P}_u(\mathcal{B}_u)| = \sum_{i} |\mathcal{B}_u(i)| \leq c_1 \cdot \sum_{i} |\mathcal{B}_u(i)\gamma| \leq c_2 \cdot \sum_{i} |\mathcal{B}_u(i)|\gamma| \leq c_3 \cdot \sum_{i} n^{|\gamma| \cdot 2^\delta} = O(n^{k-1})$$

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 $|\mathcal{P}_u(\mathcal{B}_u)| = j$ as a constant factor the size of $\mathcal{B}_u$.

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 $v$ in $\mathcal{B}$, and to compute the values $\text{FWD}_u(\mathcal{B})$ and $\text{BWD}_u(\mathcal{B})$ for every $\pi \in \mathcal{P}_u(\mathcal{B})$. Hence the time required for all nodes $u$ and one partial node $\pi \in \mathcal{P}_u(\mathcal{B})$ to store the maps values $\text{FWD}_u(\pi)$ and $\text{BWD}_u(\pi)$ is $O(n^{k-1})$.

$$T_k(n) \leq n^{2(k-1)} + 2^{\lambda \cdot k} \cdot T_k(n \cdot (1 + \frac{\delta}{2})^{\lambda-1})$$

The analysis of Eq. 3 and Eq. 4 of Lemma 4.1 gives $T_k(n) = O(n^{2(k-1)})$ if $k \geq 3$ and $T_k(n) = O(n^{k-1})$ (i.e., the above time recurrence is analyzed as the recurrence for $S_k$ of Lemma 4.1). From the space analysis we have that there exist $O(n^{k-1})$ partial nodes $\pi \in \mathcal{P}_u(\mathcal{B})$ for every node $u$ whose root bag is $\mathcal{B}$. Hence the total time for this step is $O(n^{k-1})$ for $k \geq 3$, and $O(n^{k-1})$ for $k = 2$.

5. This step requires time linear in the size of $T$.

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 $A(v) = \emptyset$ for all nodes $v \in V$. Perform a BFS on $T$ starting from $\mathcal{B}_u$, and for every encountered bag $\mathcal{B}$ and nodes $u, v \in B$ with $L(v) \leq L(u)$, set $A(v) = A(A(x) \otimes \text{FWD}_v(x))$. Return the map $A$.

2. **Pair query.** Given two nodes $u, v \in V$, find the LCA $\mathcal{B}$ of bags $\mathcal{B}_u$ and $\mathcal{B}_v$. Return $\bigoplus_{\pi \in \mathcal{P}_u(\mathcal{B})} (\text{FWD}_u(x) \otimes \text{BWD}_u(x))$.

3. **Partial pair query.** Given two partial nodes $\pi, \pi'$. If both $\pi$ and $\pi'$ are strictly partial, return $\text{FWD}_u(\pi')$, else if $\pi$ is strictly partial, return $\text{BWD}_u(\pi')$, else return $\text{FWD}_u(\pi')$.

We thus establish the following theorem.

**Theorem 3.** Let $G = (V, E)$ be a concurrent graph of $k$ constant-trieewidth graphs $(G_i)_{1 \leq i \leq k}$, and $\omega : E \rightarrow \Sigma$ 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**
   - (a) $O(n^{3(k-1)})$ if $k \geq 3$, and (b) $O(n^{k-1})$ if $k = 2$.

2. **Preprocessing space** $O(n^{2(k-1)})$.

3. **Single-source query time**
   - (a) $O(n^{3(k-1)})$ if $k \geq 3$, and (b) $O(n^{k-1})$ if $k = 2$.

4. **Pair query time** $O(n^{k-1})$.

5. **Partial pair query time** $O(1)$.

**Proof.** The correctness of ConcurQuery for handling all queries follows from Lemma 4 and the properties of the preprocessing established in Lemma 5. The preprocessing complexity is stated in Lemma 5. The time complexity for the single-source query follows by the observation that ConcurQuery spends quadratic time in each encountered bag, and the result follows from the recurrence analysis of Eq. (9) in Lemma 4. The time complexity for the pair query follows from the O(1) time to access the LCA bag $\mathcal{B}$ of $\mathcal{B}_u$ and $\mathcal{B}_v$, and the O(1) time required to iterate over all nodes $x \in \mathcal{B} \cap V$. Finally, the time complexity for the partial pair query follows from the O(1) time lookup in the constructed maps $\text{FWD}, \text{BWD}_u$, and $\text{BWD}_v$.

Note that a single-source query from a strictly partial node $\pi$ can be answered in $O(n^k)$ time by breaking it down to $n^k$ 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-trieewidth graphs $G_1, G_2$, and $\omega : E \rightarrow \Sigma$ 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** $O(n^{2(k-1)})$.

2. **Preprocessing space** $O(n^{k-1})$.

3. **Single-source query time** $O(n^{k-1})$.

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 $\text{(5)} \text{(12)}$ 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 3 of ConcurPreprocess is required for handling pair queries only. By skipping this step, we can handle every (partial) pair query $\pi, \pi'$ similarly to the single source query from $\pi$, but restricting the BFS to the path $P : \mathcal{B}_\pi \rightarrow \mathcal{B}_\pi$, and spending
Corollary 2. Let $G = (V, E)$ be a concurrent graph of two constant-treewidth graphs $G_1, G_2$, and $\omega : E \rightarrow \Sigma$ a weight function of $G$. For any fixed $\epsilon > 0$, the data-structure $\text{ConcurAP}$ (by skipping Item 4) correctly answers single-source and pair queries and requires:

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

Finally, we can use ConcurAP to obtain the transitive closure of $G$ by performing $n^2$ single-source queries. The preprocessing space is $O(n^{4+\epsilon})$ by Corollary 3 and the space of the output is $O(n^5)$. Since there are $n^2$ pairs for the computed distances. Hence the total space requirement is $O(n^5)$. The time requirement is $O(n^{4+\epsilon})$, since by Corollary 3 every single-source query requires $O(n^{2+\epsilon})$ 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 $\omega : E \rightarrow \Sigma$ a weight function of $G$. For any fixed $\epsilon > 0$, the transitive closure of $G$ wrt ConcurAP can be computed in $O(n^{4+\epsilon})$ time and $O(n^5)$ 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 $\omega : E \rightarrow \Sigma$. Let $x_i, 1 \leq i \leq n$ range over the nodes $V$ of $G$, and construct a graph $G'' = (V'', E'')$ such that $V'' = \{x_i, y_i : 1 \leq i \leq n\}$ and $E'' = \{(x_i, y_j), (y_i, x_j) : 1 \leq i \leq n\}$.

Claim 2. The treewidth of $G''$ is $2$.

We construct a graph $G' = (V', E')$ as a 2-self-concurrent asynchronous composition of $G''$, by including the following edges.

1. Black edges. For all $1 \leq i < n$ and $1 \leq j < n$ we have $(x_i, y_{i+1}), (x_{i+1}, y_i), (x_i, y_j), (x_j, y_i) \in E'$, and for all $1 \leq i < n$ and $1 \leq j \leq n$ we have $(y_i, x_j), (y_{i+1}, x_j), (y_j, x_i) \in E'$.
2. Blue edges. For all $1 \leq i < n$ we have $(x_i, x_j), (x_i, y_j), (y_i, x_j), (x_j, y_i) \in E'$.
3. Red edges. For all $(x_i, x_j) \in E$ we have $(x_i, y_j), (x_j, y_i) \in E'$.
4. Green edges. For all $1 \leq i, j \leq n$ with $i \neq j$ we have $(x_i, x_j), (x_i, y_j), (x_j, y_i) \in E'$.

Additionally, we construct a weight function such that $\omega'(x_i, y_j, x_j, x_i) = \omega(x_i, x_j)$ for every red edge $(x_i, y_j, x_j, x_i)$, and $\omega'(u, v) = 1$ for every other edge $(u, v)$.

Figure 7: A graph $G$ (left), and $G'$ that is a 2-self-product of a graph $G''$ of treewidth 2 (right). The weighted edges of $G$ correspond to weighted red edges on $G'$. The distance $d(x_i, x_j)$ in $G$ equals the distance $d((x_i, x_j), (x_j, x_i)) = d((x_i, x_i), (x_j, x_j))$ in $G'$.

Lemma 6. For every $x_i, x_j \in V$, there exists a path $P : x_i \rightarrow x_j$ with $\omega(P) = z$ in $G$ if and only if there exists a path $P' : (x_i, x_j) \rightarrow (x_j, x_j)$ with $\omega(P') = z$ in $G'$.
Table 2: Percentage of cases for which the transitive closure of the graph \( G \) for the given value of \( \lambda \) is at most 5\% slower than the time required to obtain the transitive closure of \( G \) for the best \( \lambda \).

2. According to Corollary [3], the transitive closure can be computed in \( O(n^{\lambda+1}) \) time, for any fixed \( \epsilon > 0 \), and \( O(n^3) \) space. Since the size of the output is \( \Theta(n^3) \), 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 use two different sets of benchmarks, and obtain the controlflow graphs of Java programs using Soot [69], and use LibTW [70] 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).

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 \( O(n^{\lambda+1}) \), 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 3 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 \( G \) 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 performs lock operations are only allowed wrt the lock semantics. Similarly as before, we compare our transitive closure time with the standard Bellman-Ford algorithm. Table 3 shows a time comparison between our algorithms and the baseline method. We observe that our transitive closure algorithm is significantly faster, and also scales better.

<table>
<thead>
<tr>
<th>Java method</th>
<th>( n )</th>
<th>( T_o(s) )</th>
<th>( T_b(s) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>ArrayBlockingQueue: poll</td>
<td>19</td>
<td>19</td>
<td>60</td>
</tr>
<tr>
<td>ArrayBlockingQueue: peek</td>
<td>29</td>
<td>29</td>
<td>81</td>
</tr>
<tr>
<td>LinkedBlockingDeque: advance</td>
<td>25</td>
<td>25</td>
<td>195</td>
</tr>
<tr>
<td>PriorityBlockingQueue: removeEQ</td>
<td>25</td>
<td>32</td>
<td>76</td>
</tr>
<tr>
<td>ArrayBlockingQueue: init</td>
<td>26</td>
<td>47</td>
<td>249</td>
</tr>
<tr>
<td>LinkedBlockingDeque: remove</td>
<td>26</td>
<td>49</td>
<td>290</td>
</tr>
<tr>
<td>ArrayBlockingQueue: offer</td>
<td>26</td>
<td>56</td>
<td>304</td>
</tr>
<tr>
<td>ArrayBlockingQueue: clear</td>
<td>28</td>
<td>33</td>
<td>389</td>
</tr>
<tr>
<td>ArrayBlockingQueue: contains</td>
<td>32</td>
<td>205</td>
<td>881</td>
</tr>
<tr>
<td>DelayQueue: remove</td>
<td>42</td>
<td>267</td>
<td>3792</td>
</tr>
<tr>
<td>ConcurrentHashMap: scanAndLockForPut</td>
<td>46</td>
<td>375</td>
<td>2176</td>
</tr>
<tr>
<td>ArrayBlockingQueue: next</td>
<td>46</td>
<td>407</td>
<td>3915</td>
</tr>
<tr>
<td>ConcurrentHashMap: put</td>
<td>72</td>
<td>1895</td>
<td>TO</td>
</tr>
</tbody>
</table>

Table 3: Time required for the transitive closure on concurrent graphs extracted from methods of the java.util.concurrent library. \( T_o(s) \) and \( T_b(s) \) correspond to our method and the baseline method respectively. TO denotes timeout after 8 hours.

8. Conclusions

We have considered the fundamental algorithmic problem of computing algebraic path properties in a concurrent interprocedural setting, where component graphs have constant treewidth. We have presented algorithms that significantly improve the existing theoretical complexity of the problem, and provide a variety of tradeoffs between preprocessing and query times for on-demand analyses. Moreover, we have proved that further theoretical improvements over our algorithms must achieve major breakthroughs. Given that our theoretical results are almost tight in the interprocedural concurrent setting, interesting directions of future work is to extend to the interprocedural setting. However, in that case even the basic problem of reachability is undecidable, and other techniques and formulations are required to make the analysis tractable, such as context-bounded formulations and regular approximations of interprocedural paths [15, 52, 60]. The effect of constant-treewidth components in such formulations is an interesting theoretical direction to pursue, with potential for practical use.


M. Thorup. All Structured Programs Have Small Tree Width and Good Register Allocation. *Information and Computation*, 1998.


APPENDIX

A. Modeling power

The algebraic paths framework considered in this work has a rich expressive power, as it can model a wide range of path problems arising in the static analysis of programs.

Reachability. The simplest path problem asks whether there exists a path between two locations of a concurrent system. The problem can be formulated on the boolean semiring \((\{\text{True}, \text{False}\}, \lor, \land, \text{False}, \text{True})\).

Dataflow problems. A wide range of dataflow problems has an algebraic paths formulation, expressed as a “meet-over-all-paths” analysis [48]. Perhaps the most well-known case is that of distributive flow functions considered in the IFDS framework [64, 65].

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 \((\lor)\) of the weights of all \(u \rightarrow v\) paths. The problem can be formulated on the meet-composition semiring \((F, \lor, \land, 2, I)\), 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 intra-procedural analysis of concurrent programs. The dataflow analysis of concurrent programs has been a problem of intensive study (e.g. [26, 28, 32, 37, 45, 50]), where (part of) the underlying analysis is based on an algebraic, “meet-over-all-paths” approach.

Weighted problems. The algebraic paths framework subsumes several problems on weighted graphs. The most well-known such problem is the shortest path problem [63, 33, 34, 42]. Phrased on the tropical semiring \((\mathbb{R} \cup \{-\infty, \infty\}, \inf, +, \infty, 0)\). A number of other fundamental problems on weighted graphs can be reduced to various instances of the shortest-path problem, e.g. the most probable path, the mean-payoff and the minimum initial credit problem [16, 23, 43, 55, 71]. Lately, path problems in weighted systems are becoming increasingly important in quantitative verification and quantitative program analysis [16, 18, 21, 24, 73].

Kleene algebras. Finally, a well-known family of closed semirings used in program analysis is that of Kleene algebras [33]. A common instance is when edges of the controlflow graph are annotated with observations or actions. In such case the set of observations/actions instance is when edges of the controlflow graph are annotated with use in program analysis is that of Kleene algebras [34]. A common

The concurrent graph \(G\) is associated with a weight function \(w\) of dataflow functions \(f : 2^D \rightarrow 2^D\). The dataflow function \(w(e)\) along an edge \(e\) behaves as follows on input datafact \(F\) (we only describe the case where \(F = \text{fork}\), as the other case is symmetric).

1. If \(e\) transitions to a node in which the second thread acquires \(\text{fork}\) or the first thread releases \(\text{fork}\), then \(w(e)(\text{fork}) \rightarrow \text{null}\) (i.e., \(\text{fork}\) is removed from the datafacts).

2. Else, if \(e\) transitions to a node in which the first thread acquires \(\text{fork}\), then \(w(e)(\text{null}) \rightarrow \text{fork}\) (i.e., \(\text{fork}\) is inserted to the datafacts).

Similarily for the \(F = \text{knife}\) datafact. The “meet-over-all-paths” operation is set intersection. Then the question is answered by testing whether \(d(\langle 1, 1, 3 \rangle, \langle 1, 1, 3 \rangle) = \{\text{fork}, \text{knife}\}\), i.e., by performing a partial pair query, in which the node of the second thread is unspecified.

Global property as a pair query. Assume that we are interested in determining whether the two threads can cause a deadlock. Because of symmetry, we look for a deadlock in which the first thread may hold the \(\text{fork}\), and the second thread may hold the \(\text{knife}\). A typical datafact set is \(D = \{\text{fork, knife}\}\). For a datafact \(F \in \mathcal{D}\) we have

1. \(\text{fork} \in F\) if \(\text{fork}\) may be acquired by the first thread.
2. \(\text{knife} \in F\) if \(\text{knife}\) may be acquired by the second thread.

The concurrent graph \(G\) is associated with a weight function \(w\) of dataflow functions \(f : 2^D \rightarrow 2^D\). The dataflow function \(w(e)\) along an edge \(e\) behaves as follows on input datafact \(F\).

1. If \(e\) transitions to a node in which the second thread acquires \(\text{fork}\) or the first thread releases \(\text{fork}\), then \(w(e)(\text{fork}) \rightarrow F\\{\text{fork}\}\) (i.e., the first thread no longer owns \(\text{fork}\)).
2. If \(e\) transitions to a node in which the first thread acquires \(\text{fork}\), then \(w(e)(\text{null}) \rightarrow F\\{\text{fork}\}\) (i.e., the first thread now owns \(\text{fork}\)).
3. If \(e\) transitions to a node in which the first thread acquires \(\text{knife}\) or the second thread releases \(\text{knife}\), then \(w(e)(\text{knife}) \rightarrow F\\{\text{knife}\}\) (i.e., the second thread no longer owns \(\text{knife}\)).
4. If \(e\) transitions to a node in which the second thread acquires \(\text{knife}\), then \(w(e)(\text{knife}) \rightarrow F\\{\text{knife}\}\) (i.e., the second thread now owns \(\text{knife}\)).

The “meet-over-all-paths” operation is set union. Then the question is answered by testing whether \(\{\text{fork, knife}\} \in d(\langle 1, 1, 3 \rangle, \langle 2, 2, 3 \rangle)\), i.e., by performing a pair query, and finding out whether the two threads can start the \(\text{while}\) loop with each one holding one resource. Alternatively, we can answer the question by performing a single-source query from \(\langle 1, 1, 3 \rangle\) and finding
out whether there exists any node in the concurrent graph $G$ in which every thread owns one resource (i.e., its distance contains (fork, knife)).

B. Details of Section 3

Throughout this section we fix

$$\alpha = 4 \cdot \lambda \cdot \delta; \quad \beta = (1 + \delta)/2 \cdot \lambda^{-1}; \quad \gamma = \lambda$$

We show how given a graph $G$ of treewidth $t$ and a tree-decomposition $\text{Tree}'(G)$ of $b$ bags and width $t$, we can construct in $O(b \cdot \log b)$ time and $O(b)$ space a $(\alpha, \beta, \gamma)$ tree-decomposition with $b$ bags. That is, the resulting tree-decomposition has width at most $\alpha \cdot (t + 1)$, and for every bag $B$ and descendant $B'$ of $B$ that appears $\gamma$ levels below, we have that $|T(B')| \leq \beta \cdot |T(B)|$ (i.e., the number of bags in $T(B')$ is at most $\beta$ times as large as that in $T(B)$). The result is established in two steps.

**Tree components and operations** Split and Merge. Given a tree-decomposition $T = (V_T, E_T)$, a component $C$ of $T$ is a subset of bags of $T$. The neighborhood $\text{Nh}(C)$ of $C$ is the set of bags in $V_T \setminus C$ that have a neighbor in $C$, i.e.

$$\text{Nh}(C) = \{ B \in V_T \setminus C : \{(B) \times C \} \cap E_T \neq \emptyset \}$$

Given a component $C$, we define the operation Split as $\text{Split}(C) = (X, \mathcal{Y})$, where $X \subseteq C$ is a list of bags $(B_1, \ldots, B_i)$ and $\mathcal{Y}$ is a list of sub-components $(C_1, \ldots, C_{r-1})$ such that removing each bag $B_i$ from $C$ splits $C$ into the subcomponents $\mathcal{Y}$, and for every $i$ we have $|C_i| \leq \frac{1}{2} \cdot |C|$. Note that since $C$ is a component of a tree, we can find a separator bag that splits $C$ into subcomponents of size at most $\frac{1}{2} |C|$. Applying this step recursively for $\log(2/\delta)$ levels yields the desired separator set $X$. For technical convenience, if this process yields less than $2/\delta$ bags, we repeat some of these bags until we have $2/\delta$ many.

Consider a list of components $\mathcal{Y} = (C_1, \ldots, C_i)$, and let $z = \sum_i |C_i|$. Let $j$ be the largest integer such that $\sum_{i=j}^i |C_i| \leq \frac{z}{2}$. We define the operation $\text{Merge}(\mathcal{Y}) = (\mathcal{C}_1, \mathcal{C}_2)$, where $\mathcal{C}_1 = \bigcup_{i=1}^{j-1} C_i$, and $\mathcal{C}_2 = \bigcup_{i=j+1}^i C_i$. The following claim is trivially obtained.

**Claim 3.** If $|C_i| < \frac{\delta}{2} \cdot z$ for all $i$, then $|\mathcal{C}_1| \leq \frac{1 + \delta}{2} \cdot z$.

**Proof.** By construction, $\frac{1 - \delta}{2} \cdot z < |\mathcal{C}_1| \leq \frac{1}{2} \cdot z$, and since $\mathcal{C}_1$ and $\mathcal{C}_2$ partition $\mathcal{Y}$, we have $|\mathcal{C}_1| + |\mathcal{C}_2| = 1$. The result follows. □

**Construction of a $(\beta, \gamma)$-balanced rank tree.** In the following, we consider that $T_G = \text{Tree}'(G) = (V_T, E_T)$ is a tree-decomposition of $G$ and has $|V_T| = b$ bags. Given the parameters $\lambda \in \mathbb{N}$ with $\lambda \geq 2$ and $\delta > 0$, we use the following algorithm Rank to construct a tree of bags $R_G$. Rank operates recursively on inputs $(C, \ell)$ where $C$ is a component of $T_G$ and $\ell \in \{0\} \cup [\lambda - 1]$, as follows.

1. If $|C| \cdot \frac{1}{2} \leq 1$, construct a bag $B = \bigcup_{B \in C} B$, and return $B$.
2. Else, if $\ell > 0$, let $(X, \mathcal{Y}) = \text{Split}(C)$. Construct a bag $B = \bigcup_{B \in X} B$, and let $(\mathcal{C}_1, \mathcal{C}_2) = \text{Merge}(\mathcal{Y})$. Call Rank recursively on input $(\mathcal{C}_1, \ell + 1)$ mod $\lambda$ and $(\mathcal{C}_2, (\ell + 1)$ mod $\lambda$, and let $B_1, B_2$ be the returned bags. Make $B_1$ and $B_2$ the left and right child of $B$.
3. Else, if $\ell = 0$, if $|\text{Nh}(C)| > 1$, find a bag $B$ whose removal splits $C$ into connected components $\mathcal{C}_1, \mathcal{C}_2$ with $|\text{Nh}(C)\cap C_1| \leq \frac{1}{2} |\text{Nh}(C)|$. Call Rank recursively on input $(\mathcal{C}_1, \ell + 1)$ mod $\lambda$ and $(\mathcal{C}_2, (\ell + 1)$ mod $\lambda$, and let $B_1, B_2$ be the returned bags. Make $B_1$ and $B_2$ the left and right child of $B$. Finally, if $|\text{Nh}(C)| \leq 1$, call Rank recursively on input $(C, (\ell - 1)$ mod $\lambda)$.

In the following we use the symbols $B$ and $\mathcal{B}$ to refer to bags of $T_G$ and $R_G$ respectively. Given a bag $B$, we denote by $C(B)$ the input component of Rank when $B$ was constructed, and define the neighborhood of $B$ as $\text{Nh}(B) = \text{Nh}(C(B))$. Additionally, we denote by $\text{Bh}(B)$ the set of separator bags $B_1, \ldots, B_r$ of $C$ that were used to construct $B$. It is straightforward that $\text{Bh}(B) \cap \text{Bh}(B') = \emptyset$ for every distinct $B_1$ and $B_2$.

**Claim 4.** Let $B$ and $B'$ be respectively a bag and its parent in $R_G$. Then $\text{Nh}(B) \subseteq \text{Nh}(B') \cup \text{Bh}(B')$, and thus $|\text{Nh}(B)| \leq |\text{Nh}(B')| + 2/\delta$.

**Proof.** Every bag in $\text{Nh}(C(B'))$ is either a bag in $\text{Nh}(C(B'))$, or a separator bag of $C(B')$, and thus a bag of $\text{Bh}(B')$. □
Note that every bag $B$ of $T_G$ belongs in $Bh(B)$ of some bag $B$ of $R_G$, and thus the bags of $R_G$ already cover all nodes and edges of $G$ (i.e., properties C1 and C2 of a tree decomposition). In the following we show how $R_G$ can be modified to also satisfy condition C3, i.e., that every node $u$ appears in a contiguous subtree of $R_G$. Given a bag $B$, we denote by $Nh(V(B)) = B \cup \bigcup_{i \in Nh(B)} B_i$, i.e., $Nh(V(B))$ is the set of nodes of $G$ that appear in $B$ and its neighborhood. In the sequel, to distinguish between paths in different trees, given a tree of bags $T$ (e.g., $T_G$ or $R_G$) and bags $B_1$, $B_2$ of $T$, we write $B_1 \leadsto_T B_2$ to denote the unique simple path from $B_1$ to $B_2$ in $T$.

We say that a pair of bags $(B_1, B_2)$ form a gap of some node $u$ in a tree of bags $T$ (e.g., $R_G$) if $u \in B_1 \cap B_2$ and for the unique simple path $P' : B_1 \leadsto_{T'} B_2$ (e.g., there is at least one intermediate bag in $P'$) and for all intermediate bags $P \in P'$ we have that $u \notin P$. The following crucial lemma shows that if $B_1$ and $B_2$ form a gap of a bag in $R_G$, then for every intermediate bag $P$ in the path $P : B_1 \leadsto_{R_G} B_2$, $u$ must appear in some bag of $Nh(V(B))$.

**Lemma 7.** For every node $u$, and pair of bags $(B_1, B_2)$ that form a gap of $u$ in $R_G$, such that $B_1$ is an ancestor of $B_2$, for every intermediate bag $B$ in $P : B_1 \leadsto_{R_G} B_2$ in $R_G$, we have that $u \in Nh(V(B))$.

**Proof.** Fix any such a bag $B$, and since $B_1$ and $B_2$ form a gap of $u$, there exist bags $B_1 \in Bh(B_1)$ and $B_2 \in Bh(B_2)$ with $u \in B_1 \cap B_2$. Let $B'$ be the rightmost bag of the path $P : B_1 \leadsto_{R_G} B_2$ that has been chosen as a separator when $B$ was constructed. Note that $B_1$ has been chosen as such a separator, therefore $B'$ is well defined. We argue that $B' \in Nh(V(B))$, which implies that $u \in Nh(V(B'))$. This is done in two steps.

1. Since $B_2$ is a descendant of $B$, we have that $B_2 \subseteq C(B)$, i.e., $B_2$ is a bag of the component when $B$ was constructed.
2. By the choice of $B'$, for every intermediate bag $B'$ in the path $B' \leadsto_{R_G} B_2$ we have $B' \subseteq C(B)$. Hence $B'$ is incident to the component where $B_2$ belonged at the time $B$ was constructed.

These two points imply that $B' \in Nh(V(B))$, as desired. Figure 11 provides an illustration of the argument.

**Turning the rank tree to a tree decomposition.** Lemma 7 suggests a way to turn the rank tree $R_G$ to a tree-decomposition. Let $\bar{R}_G = \text{Replace}(R_G)$ be the tree obtained by replacing each bag $B$ of $R_G$ with $Nh(V(B))$. For a bag $B$ in $R_G$ let $\bar{B}$ be the corresponding bag in $\bar{R}_G$ and vice versa.

Figure 10: Illustration of one recursive step of Rank on a component $C$ (gray). $C$ is split into two sub-components $\bar{C}_1$ and $\bar{C}_2$ by removing a list of bags $X' = (B_i)$. Once every $\lambda$ recursive calls, $X'$ contains one bag, such that the neighborhood $Nh(C_i)$ of each $C_i$ is at most half the size of $Nh(C)$ (i.e., the red area is split in half).

In the remaining $\lambda - 1$ recursive calls, $X'$ contains 2/3 bags, such that the size of each $\bar{C}_i$, is at most $\frac{3}{4\lambda}$ fraction of the size of $C$. (i.e., the gray area is split in almost half).

Figure 11: Illustration of Lemma 7. Since $B_2$ belongs to $C(B)$ and the blue sub-component has not been split yet, the bag $B'$ is in the neighborhood of the blue sub-component, and thus in the neighborhood of $C(B)$.

**Claim 5.** If there is a pair of bags $\bar{B}_1, \bar{B}_2$ that form a gap of some node $u$ in $R_G$, then there is a pair of bags $\bar{B}_1', \bar{B}_2'$ that also form a gap of $u$, and $\bar{B}_1'$ is ancestor of $\bar{B}_2'$.

**Proof.** Assume that neither of $\bar{B}_1, \bar{B}_2$ is ancestor of the other.

1. If for some $i \in \{1, 2\}$ there is no bag $B_i \in Bh(B_i)$ such that $u \in B_i$, then $u \in Nh(V(B))$ and hence there is an ancestor $B_i'$ of $B_i$ such that $u \in Nh(V(B'))$. Thus $\bar{B}_1'$ and $\bar{B}_2$ form a gap of $u$ in $R_G$.
2. Else, there exists a $B_1 \in Bh(B_1)$ and $B_2 \in Bh(B_2)$ such that $u \in B_1 \cap B_2$. Let $B$ be first bag in the path $B_1 \leadsto_{R_G} B_2$ that was chosen as a separator we have $B \in Bh(B)$ for some ancestor $B$ of $B_1$ and $B_2$, therefore $u \in Nh(V(B))$.

It follows that there exists an ancestor $B_i'$ of some $\bar{B}_i$ so that the two form a gap of $u$ in $R_G$.

The following lemma states that $\bar{R}_G$ is a tree decomposition of $G$.

**Lemma 8.** $\bar{R}_G = \text{Replace}(R_G)$ is a tree-decomposition of $G$.

**Proof.** It is straightforward to see that the bags of $\bar{R}_G$ cover all nodes and edges of $G$ (properties C1 and C2 of the definition of tree-decomposition), because for each bag $B$, we have that $B \subseteq B$. It remains to show that every node $u$ appears in a contiguous subtree of $R_G$ (i.e., that property C3 is satisfied).

Assume towards contradiction otherwise, and by Claim 5 it follows that there exist bags $B_1$ and $B_2$ in $\bar{R}_G$ that form a gap of some node $u$ such that $\bar{B}_1$ is an ancestor of $\bar{B}_2$. Let $P : B_1 \leadsto_{\bar{R}_G} B_2$ be the path between them, and $P : B_1 \leadsto_{R_G} B_2$ the corresponding path in $R_G$. By Lemma 7 we have that $u \notin B_1 \cap B_2$, otherwise for every intermediate bag $B \in P$ we would have $u \in Nh(V(B))$ and thus $u \in B$. Additionally, we have $u \in B_2$, otherwise by Claim 4 we would have $u \in Nh(V(B'_2))$, where $B'_2$ is the parent of $B_2$, and thus $u \in B'_2$, contradicting the assumption that $\bar{B}_1$ and $\bar{B}_2$ form a gap of $u$. Hence $u \notin B_1$. A similar argument as that of Claim 5 shows that there exists an ancestor $B'_1$ of $B_1$ such that $u \in B'_1$, and WLOG, take $B'_1$ to be the lowest ancestor of $B_1$ with this property. Then $B'_1$ is also an ancestor of $B_2$, and $B'_1$ and $B_2$ form a gap of $u$ in $R_G$. Since $B_1$ is an intermediate bag in $B'_1 \leadsto_{\bar{R}_G} B_2$, by Lemma 7 we have that $u \in Nh(V(B))$, thus $u \in B$. We have thus arrived at a contradiction, and the desired result follows.

**Properties of the tree-decomposition $\bar{R}_G$.** Lemma 8 states that $\bar{R}_G$ obtained by replacing each bag of $R_G$ with $Nh(V(B))$ is a tree-decomposition of $G$. The remaining of the section focuses on...
We show that ConcurTree constructed out of a tree-decomposition.

Proof. We prove each item separately.

1. For every bag $B$ of $\mathcal{R}_G$ is $(\beta, \gamma)$-balanced.

2. For every bag $\mathcal{B}$ of $\mathcal{R}_G$, we have $|\mathcal{B}| \leq \alpha \cdot (t + 1)$.

Lemma 9. The following assertions hold:

1. Every bag $B$ of $\mathcal{R}_G$ is $(\beta, \gamma)$-balanced.
2. For every bag $\mathcal{B}$ of $\mathcal{R}_G$, we have $|\mathcal{B}| \leq \alpha \cdot (t + 1)$.

Proof. We prove each item separately.

1. For every bag $B$ of $\mathcal{R}_G$, $\mathcal{B}$ is $(\beta, \gamma)$-balanced in $\mathcal{R}_G$, and hence $\mathcal{B}$ is $(\beta, \gamma)$-balanced in $\mathcal{R}_G$.

2. It suffices to show that for every bag $B$, we have $|\text{Nh}(\mathcal{B})| \leq \alpha - 1 = 2 \cdot (2/\delta) \cdot \lambda - 1$. Assume towards contradiction otherwise. Let $B$ be the first bag that $\text{Rank}$ assigned a rank such that $|\text{Nh}(\mathcal{B})| \geq 2 \cdot (2/\delta) \cdot \lambda$. Let $\mathcal{B}'$ be the lowest ancestor of $\mathcal{B}$ in $\mathcal{R}_G$ that was constructed by $\text{Rank}$ on some input $(C, \ell)$ with $\ell = 1$, and let $\mathcal{B''}$ be the parent of $\mathcal{B}'$ in $\mathcal{R}_G$ (note that $\mathcal{B''}$ can be $\mathcal{B}$ itself). By Item 3 of $\text{Rank}$, it follows that $|\text{Nh}(\mathcal{B}'')| \leq \frac{|\text{Nh}(\mathcal{B}'')|}{2} + 1$. Note that $\mathcal{B''}$ is at most $\lambda - 1$ levels above $\mathcal{B}$ (as we allow $\mathcal{B''}$ to be $\mathcal{B}$). By Claim 4, the neighborhood of a bag can increase by at most $\frac{1}{\lambda}$ from the neighborhood of its parents, hence $|\text{Nh}(\mathcal{B}'')| \geq (2/\delta) \cdot (\lambda + 1)$. The last two inequalities lead to $|\text{Nh}(\mathcal{B}'')| \geq 2 \cdot (2/\delta) \cdot \lambda$, which contradicts our choice of $\mathcal{B}$.

The desired result follows.

A minimal example. Figure 12 illustrates an example of $\mathcal{R}_G$ constructed out of a tree-decomposition $\text{Tree}'(G)$. First, $\text{Tree}'(G)$ is turned into a binary and balanced tree $\mathcal{R}_G$ and then into a binary and balanced tree $\mathcal{R}_G$. If the numbers are pointers to bags, such that $\text{Tree}'(G)$ is a tree-decomposition for $\mathcal{G}$, then $\mathcal{R}_G$ is a binary and balanced tree-decomposition of $\mathcal{G}$. The values of $\lambda$ and $\delta$ are immaterial for this example, as $\mathcal{R}_G$ becomes perfectly balanced (i.e., $(1/2, 1)$-balanced).

Theorem 1. For every graph $G$ with $n$ nodes and constant tree-width, for any fixed $\delta > 0$ and $\lambda \in \mathbb{N}$ with $\lambda \geq 2$, let $\alpha = 4 \cdot \lambda / \delta$, $\beta = \frac{(1 + \delta)}{2} \lambda - 1$, and $\gamma = \lambda$. A binary $(\alpha, \beta, \gamma)$ tree-decomposition $\text{Tree}'(G)$ with $O(n)$ bags can be constructed in $O(n \cdot \log n)$ time and $O(n)$ space.

Proof. By 12 an initial tree-decomposition $\text{Tree}'(G)$ of $G$ with width $t$ and $b = O(n)$ bags can be constructed in $O(n)$ time. Lemma 8 and Lemma 9 prove that the constructed $\mathcal{R}_G$ is a $(\alpha, \beta, \gamma)$ tree-decomposition of $G$. The time and space complexity come from the construction of $\mathcal{R}_G$ by the recursion of $\text{Rank}$. It can be easily seen that every level of the recursion processes disjoint components $C_i$ of $\text{Tree}'(G)$ in $O(|C_i|)$ time, thus the time complexity of the recursion requires $O(b)$ time in total. There are $O(\log b)$ such levels, since every $\lambda$ levels, the size of each component has been reduced to at most a factor $(1 + \delta) \lambda - 1$. Hence the time complexity is $O(\cdot \log b) = (n \cdot \log n)$. The space complexity is that of processing a single level of the recursion, hence $O(b) = O(n)$.

C. Details of Section 4

Lemma 10. ConcurTree$(G)$ is a tree-decomposition of $G$.

Proof. We show that $T$ satisfies the three conditions of a tree decomposition.
by factoring out every $n_{i,r_i}$ of the sum. Similarly,
\[
Y = \sum_{(r_i,j) \in [2]^k} \left( \sum_{1 \leq i \leq k, j \neq i} n_{j,r_j} \right)
\]
\[
= \sum_{1 \leq i \leq k} \left( \sum_{r_i,j \in [2]} n_{j,r_j} \right)
\]
\[
= 2 \cdot \sum_{1 \leq i \leq k} \left( \frac{n_{1,r_1}}{r_1} \prod_{r_i,j \in [2]} n_{j,r_j} \right)
\]
\[
\geq 2 \cdot \frac{\prod_{1 \leq j \leq k} n_{j}}{1}(11)
\]
as the inner sum in the second line is independent of $i$, and $r_i \in [2]$.
\[
T(n_1, \ldots, n_k) \leq \sum_{1 \leq i \leq k} \prod_{j \neq i} n_j + X - Y
\]
\[
\leq c_1 \cdot \prod_{1 \leq i \leq k} n_i - c_2 \cdot \sum_{1 \leq i \leq k} \prod_{j \neq i} n_j
\]
for appropriate choices of the constants $c_1$, $c_2$, and thus
\[
T(n_1, \ldots, n_k) = O(n_1 \cdot \ldots \cdot n_k) = O(n^k),
\]
as desired.

**Lemma 12.** ConcurTree requires $O(n^k)$ time and space.

Proof. It is easy to verify that ConcurTree$(G)$ performs a constant number of operations per node per bag in the returned tree decomposition. Hence we will bound the time taken by bounding the size of ConcurTree$(G)$. Consider a recursion step of ConcurTree on input $(T_i(B_i))_{1 \leq i \leq k}$. Let $n_i = |T_i(B_i)|$ for all $1 \leq i \leq k$, and $n_{i,r_i} = |T_i(B_{i,r_i})|$, $r_i \in [2]$, where $B_{i,r_i}$ is the $r_i$-th child of $B_i$. In view of Remark 2, the time required by this input is given by the recurrence in Eq. [7], up to a constant factor. The desired result follows from Lemma [11].

**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 $2^k$-ary tree-decomposition ConcurTree$(G)$ of $G$ in $O(n^k)$ time and space, with the following property. For every $i \in [n]$ and bag $B$ at level $Lv(B) \geq i \cdot \gamma$, we have $|B| = O(n^{k-1} \cdot \beta^i)$.

Proof. Lemma [10] proves the correctness and Lemma [12] the complexity. Here we focus on bounding the size of a bag $B$ with $Lv(B) \leq i \cdot \gamma$. Let $(T_i(B_i))_{1 \leq i \leq k}$ be the input on ConcurTree when it constructed $B$ using Eq. [1] and $n_i = |T_i(B_i)|$. Observe that $Lv(B) = Lv(B_i)$ for all $i$, and since each $T_i$ is $(\beta, \gamma)$-balanced, we have that $n_i \leq O(n \cdot \beta^i)$. Since each $T_i$ is $\alpha$-approximate, $|B_i| = O(1)$ for all $i$. It follows from Eq. [1] and Remark 2 that $|B| = O(n^{k-1} \cdot \beta^i)$. \hfill \qedsymbol

**D. Details of Section 5**

**Lemma 1.** Consider a graph $G = (V, E)$ with a weight function $wt : E \rightarrow \Sigma$, and a tree-decomposition Tree$(G)$. Let $u, v \in V$, and $P : B_1, B_2, \ldots, B_j$ be a simple path in $T$ such that $u \in B_1$ and $v \in B_j$. Let $A = \{u\} \times (\prod_{i < j}(B_{i-1} \cap B_i)) \times \{v\}$. Then
\[
d(u, v) = \bigoplus_{(x_1, \ldots, x_{j+1}) \in A} \bigotimes_{i=1}^{j} d(x_i, x_{i+1}).
\]

Proof. By Lemma 1, for every bag $B_i$ with $i > 1$ and path $P : u \rightsquigarrow v$, there exists a node $x_i \in B_i \cap B_{i-1}$ Denote by $P_{x,y}$ a path $x \rightsquigarrow y$ in $G$. Then
\[
d(u, v) = \bigoplus_{P_{u,v}} \bigotimes_{P_{u,v}}
\]
and the proof follows an easy induction on $i$. \hfill \qedsymbol

**Claim 1.** For every partial node $\overline{u}$ and strictly partial node $\overline{v}$ we have $d(\overline{u}, \overline{v}) = d(\overline{u}, \overline{v})$ and $d(\overline{u}, \overline{v}) = d(\overline{u}, \overline{v})$.

Proof. By construction, for every node $v \in V$ that strictly refines $\overline{u}$ (i.e., $v \subset \overline{u}$), we have $\overline{wt}(\overline{v}) = d(\overline{v}, \overline{v}) = \overline{d}(\overline{v}, \overline{v}) = d(v, v)$. Hence, every such $v$ can reach (resp. be reached from) $\overline{v}$ (resp. $\overline{v}$) without changing the distance from $\overline{u}$. The claim follows easily. \hfill \qedsymbol

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

Proof. By Theorem 2 ConcurTree$(G)$ is a tree decomposition of $G$. To show that $T$ is a tree decomposition of the partial expansion $\overline{G}$, it suffices to show that the conditions C1-C3 are met for every pair of nodes $\overline{u}$, $\overline{v}$ that correspond to a strict partial node $\overline{u}$ of $G$. We only focus on $\overline{u}$, as the other case is similar.

C1 This condition is met, as $\overline{u}$ appears in every bag of $T$ that contains a node $u$ that refines $\overline{u}$.

C2 Since every node $\overline{u}$ is connected only to nodes $u$ of $G$ that refine $\overline{u}$, this condition is also met.
C3 If $\Pi' = \langle \emptyset, \ldots, \emptyset \rangle$, then $\Pi'$ appears in all bags of $T$, hence the condition is met. Otherwise, let $1 \leq i \leq k$ be any index such that for the $i$-th constituent $u_i$ of $\Pi'$, we have $u_i \notin \emptyset$ (i.e., $u_i \in V_j$), and note that every node $u \in V$ that refines $\Pi$ has $u_i$ as its $i$-th constituent. Let $B_i$ be the root bag of Tree($G_i$), and we have that $u \in V_{T_i}(B_i)$. Additionally, if for some bag $B_i'$ of Tree($G_i$) we have $u \notin V_{T_i}(B_i')$, then $u \notin V_{T_i}(B_i')$ for any descendant $B_i''$ of $B_i'$. Hence, by the construction of ConcurTree($G$), the node $u_i$ appears as a constituent in a contiguous subtree of ConcurTree($G$), thus the set of nodes $u$ that refine $\Pi'$ appears in a contiguous subtree of ConcurTree($G$). It follows that $\Pi'$ appears in a contiguous subtree of $T$, as desired.

The desired result follows.

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

$$T_k(n) \leq n^{3(k-1)} + 2^{\lambda k} \cdot T_k \left( n \cdot \left( \frac{1 + \delta}{2} \right)^{\lambda-1} \right)$$

(5)

$$S_k(n) \leq n^{2(k-1)} + 2^{\lambda k} \cdot S_k \left( n \cdot \left( \frac{1 + \delta}{2} \right)^{\lambda-1} \right)$$

(6)

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) $S_k(n) = O(n^{2+k})$.

**Proof.** We analyze each recurrence separately. First we consider Eq. (5). Note that

$$\left( n \cdot \left( \frac{1 + \delta}{2} \right)^{\lambda-1} \right)^{3(k-1)} = \left( \frac{1 + \delta}{2} \right)^{3(\lambda-1)(k-1)} \cdot n^{3(k-1)}$$

(12)

and

$$2^{\lambda k} \cdot \left( \frac{1 + \delta}{2} \right)^{3(\lambda-1)(k-1)} = \frac{2^{2+k}}{2^{2+k} \cdot \lambda + 3} \cdot \left( \frac{1 + \delta}{2} \right)^{3(\lambda-1)(k-1)}$$

(13)

and since

$$\log(1 + \delta) = \frac{\ln(1 + \delta)}{\ln 2} < \frac{\delta}{\ln 2} < 2 \cdot \delta,$$

we have

$$(1 + \delta)^{3(\lambda-1)(k-1)} = 2^{3\log(1 + \delta)} \cdot 3(\lambda-1)(k-1) < 2^{2+k} \cdot (\lambda-1)(k-1).$$

Hence the expression in Eq. (13) is bounded by $2^k$ with

$$x \leq 2 \cdot \delta \cdot (\lambda - 1) \cdot (k - 1) - 2 \cdot k \cdot \lambda + 3 \cdot (\lambda + k - 1)$$

and note that since $\lambda \geq \frac{2}{\epsilon} > 4$ and $\delta \leq \frac{\delta}{\ln 2} < \frac{1}{n^{2/\epsilon}}$, $f(k)$ is decreasing, and thus maximized for $k = 2$, for which we obtain

$$f(2) = \frac{-2 \cdot \lambda \cdot k \cdot (1 - 3 \cdot \delta) + 3 \cdot (\lambda + k - 1) \cdot (1 - 2 \cdot \delta) < 0}{\delta \leq \frac{1}{n^{2/\epsilon}}},$$

which yields that there exists a constant $c \leq 1$ for which

$$2^{\lambda k} \cdot T_k \left( n \cdot \left( \frac{1 + \delta}{2} \right)^{\lambda-1} \right) \leq c \cdot n^{3(k-1)}$$

which yields that Eq. (5) follows a geometric series, and thus $T_k(n) = O(n^{3(k-1)})$.

We now turn our attention to Eq. (6). When $k \geq 3$, an analysis similar to Eq. (5) yields the bound $O(n^{2+k})$. When $k = 2$, since $\epsilon > 0$, we write Eq. (6) as

$$S_2(n) \leq n^{2+k} + 2^{2\lambda} \cdot S_2 \left( n \cdot \left( \frac{1 + \delta}{2} \right)^{\lambda-1} \right)$$

(14)

Similarly as above, we have

$$\left( n \cdot \left( \frac{1 + \delta}{2} \right)^{\lambda-1} \right)^{2+k} = \left( \frac{1 + \delta}{2} \right)^{2+k} \cdot n^{2+k}$$

(15)

and

$$2^{2\lambda} \cdot \left( \frac{1 + \delta}{2} \right)^{2+k} \cdot \left( \frac{1 + \delta}{2} \right)^{\lambda-1} = \left( \frac{1 + \delta}{2} \right)^{2+k} \cdot \left( \frac{1 + \delta}{2} \right)^{\lambda-1}$$

(16)

and since

$$\log(1 + \delta) = \frac{\ln(1 + \delta)}{\ln 2} < \frac{\delta}{\ln 2} < 2 \cdot \delta,$$

we have

$$\left( \frac{1 + \delta}{2} \right)^{\lambda-1} < 2^{2\lambda} \cdot (\lambda-1)$$

Hence the expression in Eq. (16) is bounded by $2^k$ with

$$x \leq \frac{2 \cdot \delta \cdot (\lambda - 1)}{2} + 2 \cdot \epsilon \cdot (\lambda - 1)$$

$$= \frac{(\lambda - 1)(2 \cdot \delta \cdot (2 + \epsilon) - \epsilon)}{2} + 2$$

$$\leq \frac{(\lambda - 1)(2 \cdot \epsilon - 2 \cdot 2) + 2}{18}$$

$$\leq \frac{2}{3} + 2 \leq 0$$

since $\delta \leq \frac{1}{n^{2/\epsilon}}$ and $\lambda \geq \frac{2}{\epsilon}$ and $\epsilon \leq 1$. It follows that there exists a constant $c \leq 1$ for which

$$2^{2\lambda} \cdot S_2(n) \leq c \cdot n^{2+k}$$

which yields that Eq. (14) follows a geometric series, and thus $S_2(n) = O(n^{2+k})$.

**E. Details of Section 6**

**Claim 2.** The treewidth of $G''$ is 2.

**Proof.** Construct a bag $B_i$ and let $x_i, y_i \in B_i$ for all $1 \leq i \leq n$, and $y_{i+1} \in B_i$ for all $1 \leq i < n$. Make a tree $T$ where $B_i$ is an ancestor of $B_i+1$. Then $T$ is a tree decomposition of $G''$ and has width 2.

**Lemma 6.** For every $x_i, x_j \in V$, there exists a path $P : x_i \rightsquigarrow x_j$ with $\text{wt}(P) = z$ in $G$ iff there exists a path $P' : \langle x_i, x_j \rangle \rightsquigarrow \langle x_j, x_j \rangle$ with $\text{wt}(P') = z$ in $G'$.

**Proof.** Recall that only red edges contribute to the weights of paths in $G'$. We argue that there is a path $P : \langle x_i, x_j \rangle \rightsquigarrow \langle x_j, x_j \rangle$ in $G'$ that traverses a single red edge iff there is an edge $(x_i, x_j)$ in $G$ with $\text{wt}(P) = \text{wt}(x_i, x_j)$.

1. Given the edge $(x_i, x_j)$, the path $P$ is formed by traversing the red edge $(\langle x_i, y_i \rangle, \langle x_j, y_j \rangle)$ as

$$\langle x_i, y_i \rangle \rightsquigarrow \langle x_i, y_i \rangle \rightsquigarrow \langle x_j, y_j \rangle \rightsquigarrow \langle x_j, y_j \rangle \rightsquigarrow \langle x_j, x_j \rangle$$

Since $\text{wt}(\langle x_i, y_i \rangle, \langle x_j, y_j \rangle) = \text{wt}(x_i, x_j)$ and all other edges of $P$ have weight $T$, we have that $\text{wt}(P) = \text{wt}(x_i, x_j)$.

2. Every path $P$ that traverses a red edge $(\langle x_i, y_i \rangle, \langle x_j, y_j \rangle)$ has to traverse a blue edge to $(x_j, x_j)$. Then $x_j$ must be $x_i$, otherwise $P$ will traverse a second red edge before reaching $(x_j, x_j)$.

The result follows easily from the above.
F. Formal Pseudocode of Our Algorithms

The current section presents formally (in pseudocode) the algorithms that appear in the main part of the paper.

Algorithm 1: Rank

Input: A component \( C \) of \( T \), a natural number \( \ell \in [\lambda] \)
Output: A rank tree \( R_C \)

1. Assign \( T \leftarrow \) an empty tree
2. if \( |C| \leq 2 \)
3. Assign \( B \leftarrow \bigcup_{B \in C} B \) and make \( B \) the root of \( T \)
4. else if \( \ell > 0 \)
5. Assign \( (X, Y) \leftarrow \text{Split}(C) \)
6. Assign \( B \leftarrow \bigcup_{B \in X} B_i \)
7. Assign \( (T_1, T_2) \leftarrow \text{Merge}(Y) \)
8. Assign \( T_1 \leftarrow \text{Rank}(T_1, (\ell + 1) \mod \lambda) \)
9. Assign \( T_2 \leftarrow \text{Rank}(T_2, (\ell + 1) \mod \lambda) \)
10. Make \( B \) the root of \( T \) and \( T_1 \) and \( T_2 \) its left and right subtree
11. else
12. if \( |\text{Nh}(C)| > 1 \)
13. Let \( B \leftarrow \) a bag of \( C \) whose removal splits \( C \) into \( T_1, T_2 \) with \( |\text{Nh}(T_1) \cap \text{Nh}(T_2)| \leq \frac{|\text{Nh}(C)|}{2} \)
14. Assign \( B \leftarrow B \)
15. Assign \( T_1 \leftarrow \text{Rank}(T_1, (\ell + 1) \mod \lambda) \)
16. Assign \( T_2 \leftarrow \text{Rank}(T_2, (\ell + 1) \mod \lambda) \)
17. Make \( B \) the root of \( T \) and \( T_1 \) and \( T_2 \) its left and right subtree
18. else
19. Assign \( T \leftarrow \text{Rank}(C, (\ell - 1) \mod \lambda) \)
20. end
21. end
22. return \( T \)

Algorithm 2: Concur Tree

Input: Tree-decompositions \( T_i = (V_{T_i}, E_{T_i})_{1 \leq i \leq k} \) with root bags \((B_1)_{1 \leq i \leq k}\)
Output: A tree decomposition \( T \) of the concurrent graph

1. Assign \( B \leftarrow \emptyset \)
2. Assign \( T \leftarrow \) a tree with the single bag \( B \) as its root
3. for \( i \in [k] \) do
4. Assign \( B \leftarrow B \cup \left( \prod_{1 \leq j < i} V_{T_j}(B_j) \times B_i \times \prod_{i < j \leq k} V_{T_j}(B_j) \right) \)
5. end
6. if none of the \( B_i \)'s is a leaf in its respective \( T_i \), then
7. for every sequence of bags \( B_1', \ldots, B_k' \) such that each \( B_i' \) is a child of \( B_i \) in \( T_i \), do
8. Assign \( T'_i \leftarrow \text{ConcurTree}(T_1(B_1'), \ldots, T_k(B_k')) \)
9. Add \( T'_i \) to \( T_i \), setting the root of \( T'_i \) as a new child of \( B \)
10. end
11. end
12. return \( T \)

Algorithm 3: Concur Preprocess Item 1

Input: Graphs \( (G_i = V_i, E_i)_{1 \leq i \leq k} \), a concurrent graph \( G(V, E) \) of \( G_i \)'s and a weight function \( \text{wt} : E \rightarrow \Sigma \)

/* Construct the partial expansion \( \overline{G} \) of \( G \) */
1. Assign \( V \leftarrow V \)
2. Assign \( E \leftarrow E \)
3. Create a map \( \text{wt} : E \rightarrow \Sigma \)
4. Assign \( \overline{G} \leftarrow \text{wt} \)
5. foreach \( u' \in \bigcap_i (V_i \cup \{\perp\}) \) do
6. Let \( u \in V \) such that \( u \sqsubseteq u' \)
7. Assign \( \overline{V} \leftarrow \overline{V} \cup \{(\pi^1, \pi^2)\} \)
8. Assign \( \overline{E} \leftarrow \overline{E} \cup \{(\pi^1, u), (u, \pi^2)\} \)
9. Set \( \overline{G}(\pi^1, u) \leftarrow (\overline{V}, \overline{E}) \)
10. Set \( \overline{G}(u, \pi^2) \leftarrow (\overline{V}, \overline{E}) \)
11. end
12. return \( \overline{G} = (\overline{V}, \overline{E}) \) and \( \overline{G} \)

Algorithm 4: Concur Preprocess Item 2

Input: A tree-decomposition \( T = \text{Tree}(G) = (V_T, E_T) \) and the partial expansion \( \overline{G} = (\overline{V}, \overline{E}) \)

/* Construct the tree-decomposition \( T \) of \( \overline{G} \) */
1. Assign \( \overline{V} \leftarrow \emptyset \)
2. foreach \( B \in V_T \) do
3. Assign \( \overline{B} \leftarrow B \) foreach \( u \in B \) do
4. Assign \( \overline{u} \leftarrow \overline{u} \) such that \( u \sqsubseteq \overline{u} \)
5. Assign \( \overline{B} \leftarrow \overline{B} \cup \{(\pi^1, \pi^2)\} \)
6. end
7. end
8. Assign \( \overline{V} \leftarrow \overline{V} \cup \{\overline{B}\} \)
9. end
10. return \( T = (\overline{V}, \overline{E}) \)
Algorithm 5: ConcurPreprocess Item

Input: The partial expansion tree-decomposition $T = (V_T, E_T)$, and weight function $\omega_T$

/* Local distance computation */

1. foreach partial node $\pi$ do
   2. Create two maps $\text{FWD}_T, \text{BWD}_T : B_T \rightarrow \Sigma$
   3. for $v \in B_T$ do
      4. Assign $\text{FWD}_T(v) \leftarrow \omega_T(\pi, v)$
      5. Assign $\text{BWD}_T(v) \leftarrow \omega_T(\sigma, \pi)$
   6. end
   7. end

8. foreach bag $B$ of $T$ in bottom-up order do
   9. Assign $d' \leftarrow$ the transitive closure of $\text{FWD}_T[B]$ with wrt $\omega_T$
   10. foreach $\pi, \sigma \in B$ do
        if $\text{Lvl}(\pi) \leq \text{Lvl}(\sigma)$ then
            11. Assign $\text{BWD}_T(\pi) \leftarrow d'(\sigma, \pi)$
            12. Assign $\text{FWD}_T(\sigma) \leftarrow d'(\pi, \sigma)$
        end
   13. end
   14. end

15. foreach bag $B$ of $T$ in top-down order do
   16. Assign $d' \leftarrow$ the transitive closure of $\text{FWD}_T[B]$ with wrt $\omega_T$
   17. foreach $\pi, \sigma \in B$ do
        if $\text{Lvl}(\sigma) \leq \text{Lvl}(\pi)$ then
            18. Assign $\text{BWD}_T(\pi) \leftarrow d'(\pi, \sigma)$
            19. Assign $\text{FWD}_T(\sigma) \leftarrow d'(\sigma, \pi)$
        end
   20. end
   21. end

Algorithm 6: ConcurPreprocess Item

Input: The partial expansion tree-decomposition $T = (V_T, E_T)$ and maps $\text{FWD}_T, \text{BWD}_T : B_u \rightarrow \Sigma$ for every partial node $\pi$

/* Ancestor distance computation */

1. foreach node $u \in V$ do
   2. Create two maps $\text{FWD}_u, \text{BWD}_u : \text{P}_u \rightarrow \Sigma$
   3. end

4. foreach bag $B$ of $T$ in DFS order starting from the root do
   5. Let $B'$ be the parent of $B$
   6. foreach node $u \in B' \cap V$ such that $B'$ is the root of $u$ do
      7. Assign $\text{FWD}_u(\pi) \leftarrow \bigoplus_{x \in \text{P}_u \cap B'} \text{FWD}_u(\pi) \otimes \omega_u^+(x, \pi)$
      8. Assign $\text{BWD}_u(\pi) \leftarrow \bigoplus_{x \in \text{P}_u \cap B'} \text{BWD}_u(\pi) \otimes \omega_u^+(T, x)$
   9. end
 10. end
 11. end
 12. end

Algorithm 7: ConcurQuery Single-source query

Input: A source node $u \in V$
Output: A map $A : V \rightarrow \Sigma$ that contains distances of vertices from $u$

1. Create a map $A : V \rightarrow \Sigma$
2. for $v \in V$ do
   3. Assign $A(v) \leftarrow \emptyset$
   4. end
5. for every bag $B$ of $T$ in bottom-up order do
   6. for $x, v \in B \cap V$ do
      7. if $\text{Lvl}(v) \leq \text{Lvl}(x)$ then
          8. Assign $A(v) \leftarrow A(v) \oplus A(x) \otimes \text{FWD}_u(v)$
       end
   9. end
10. end
11. return $A$

Algorithm 8: ConcurQuery Pair query

Input: Two nodes $u, v \in V$
Output: The distance $d(u, v)$

1. Let $B'$ be the LCA of $B_u$ and $B_v$ in $T$
2. Assign $d \leftarrow \emptyset$
3. for $x \in B' \cap V$ do
   4. Assign $d \leftarrow d \oplus \text{FWD}_u(x) \otimes \text{BWD}_v(x)$
5. end
6. return $d$

Algorithm 9: ConcurQuery Partial pair query

Input: Two partial nodes $\pi, \tau \in V$, at least one of which is strictly partial
Output: The distance $d(\pi, \tau)$

1. if both $\pi$ and $\tau$ are strictly partial then
   2. return $\text{FWD}_u(\pi)$
3. else if $\pi$ is strictly partial then
   4. return $\text{BWD}_u(\tau)$
5. else
   6. return $\text{FWD}_u(\tau)$
7. end