Algorithms for Algebraic Path Properties in Concurrent Systems 13; 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

IST Austria (Institute of Science and Technology Austria)
Am Campus 1
A-3400 Klosterneuburg, Austria
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 hecix 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 [69], 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 IDFS/IDE frameworks [38, 55] in the intraprocedural setting, which consider compositions of distributive dataflow functions, and meet-over-all-paths as the semiring plus operator. Since IDFS/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 intra-procedural 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 \[23, 26, 32, 64, 66\]. 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 \[6, 55, 54, 72\], phrased on the tropical semiring. For example, the edge weights (positive and negative) can express energy consumptions, and the shortest path problem asks for the least energy consuming path. Another important quantitative property is the mean-payoff property, where each edge weight represents a reward or cost, and the problem asks for a path that minimizes the average of the weights along a path. Many quantitative properties of relevance for program analysis (e.g., to express performance or resource consumption) can be modeled as mean-payoff properties \[15, 24\]. 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 \[16, 18, 21, 22, 47, 55, 71, 73\].

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 \(n\) in the product graph only specifies nodes of a nonempty strict subset of all the components. Our work also considers partial pair and partial single-source queries, where the input nodes are partial nodes. Queries on partial nodes are very natural, as they capture properties between local locations in a component, that are shaped by global paths in the whole concurrent system. For example, constant propagation and dead code elimination are local properties in a program, but their analysis requires analyzing the concurrent system as a whole.

**Preprocess vs query.** A topic of widespread interest in the programming languages community is that of on-demand analysis \[5, 23, 29, 33, 58, 62, 63, 66, 75, 76\]. Such analysis has several advantages, such as (quoting from \[43, 63\]): (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 pre-

processing 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 problems: typically the controlflow graphs of programs have constant treewidth \[17, 30, 64\], 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 \(O(n^k)\)). Hence for the important special case of two components we obtain a heur-time (i.e., \(O(n^k)\)) 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 \(O(n^3(k^{-1}))\) preprocessing time, we can answer (i) single-source queries in \(O(n^2(k^{-1}))\) time, (ii) pair queries in \(O(n(k^{-1}))\) time, (iii) partial single-source queries in \(O(n^k)\) time, and (iv) partial pair queries in \(O(1)\) time; while using at all times \(O(n^2k^{-1})\) space. In contrast,
the existing methods \cite{35,49,56,72} compute the transitive closure even for a single query, and thus require $O(n^{2+\epsilon})$ time and $O(n^{2+\epsilon})$ space.

**Two components.** For the important case of two threads, the existing methods require $O(n^6)$ time and $O(n^{2+\epsilon})$ space even for one query. In contrast, we establish a variety of tradeoffs between preprocessing and query times, and the best choice depends on the number of expected queries. In particular, for any fixed $\epsilon > 0$, we establish the following three results.

**Three results.** First, we show (Corollary 2) that with $O(n^3)$ preprocessing time and using $O(n^{2+\epsilon})$ space, we can answer single-source queries in $O(n^{2+\epsilon})$ time, and pair and partial pair queries require $O(n^2)$ time. Second, we show (Theorem 3) that with $O(n^{3+\epsilon})$ preprocessing time and using $O(n^{2+\epsilon})$ space, we can answer the pair and the partial pair queries in time $O(n)$ and $O(1)$, respectively. Third, we show (Corollary 3) that the transitive closure can be computed using $O(n^{2+\epsilon})$ preprocessing time and $O(n^{3+\epsilon})$ space, after which single-source queries require $O(n^2)$ time, and pair and partial pair queries require $O(1)$ time (i.e., all queries require linear time in the size of the output).

**Tradeoffs.** Our results provide various tradeoffs: The first result is best for answering $O(n^{1+\epsilon})$ pair and partial pair queries; the second result is best for answering between $\Omega(n^{1+\epsilon})$ and $O(n^{3+\epsilon})$ pair and partial pair queries; and the third result is best when answering $\Omega(n^{3+\epsilon})$ pair queries. Observe that the transitive closure computation is preferred when the number of queries is large, in sharp contrast to the existing methods that compute the transitive closure even for a single query. Our results are summarized in Table 1 and the tradeoff is pictorially illustrated in Figure 1.

<table>
<thead>
<tr>
<th>Previous results</th>
<th>Our result Corollary 2</th>
<th>Our result Theorem 3</th>
<th>Our result Corollary 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$O(n^6)$</td>
<td>$O(n^{2+\epsilon})$</td>
<td>$O(n^{2+\epsilon})$</td>
<td>$O(n^{2+\epsilon})$</td>
</tr>
<tr>
<td>$O(n^{2+\epsilon})$</td>
<td>$O(n^2)$</td>
<td>$O(n^{3+\epsilon})$</td>
<td>$O(n^2)$</td>
</tr>
<tr>
<td>$O(1)$</td>
<td>$O(1)$</td>
<td>$O(n)$</td>
<td>$O(n)$</td>
</tr>
</tbody>
</table>

Table 1: The algorithmic complexity for computing algebraic path queries wrt a closed, complete semiring on a concurrent graph $G$ which is the product of two constant-treewidth graphs $G_1, G_2$, with $n$ nodes each.

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

- **Almost optimality.** First, note that 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 $\Omega(n^3)$, which is a lower bound for computing the transitive closure (which has $n^3$ entries). Furthermore, for our third result (obtained from Corollary 3) the $O(n^{3+\epsilon})$ preprocessing time is arbitrarily close to optimal, and the $O(n^{3+\epsilon})$ 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 \cite{1,2,42}. 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 $O(n\cdot m)$, where $m$ is the number of edges. Since $m$ can be $\Omega(n^2)$, 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 3) for answering $r$ queries, for $r \leq O(n)$, would yield the same improvement over the $O(n^{3+\epsilon})$ time for answering $r$ pair queries in general graphs. That is, the combination of our preprocessing and query time (from Corollary 3) 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 4) cannot be improved much further even for $n^2$ queries, as the combined time for preprocessing and answering $n^2$ queries.
is $O(n^{3+b})$ using Theorem \ref{thm:main} while the existing bound is $O(n^b)$ 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 $\beta > 0$ and $\gamma \in \mathbb{N}$ with $\lambda \geq 2$, let $\beta = ((1 + \beta)/2)^{-1}$ and $\gamma = \lambda$. We show that a $(\beta, \gamma)$-balanced tree decomposition of a constant-treewidth graph with $n$ nodes can be constructed in $O(n \cdot \log n)$ time and $O(n)$ space. To our knowledge, this is the first algorithm that constructs a tree decomposition with such a strong notion of balance. This property is crucial for achieving the resource bounds of our algorithms for algebraic paths. The construction is presented in Section \ref{sec:upperbounds}.

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

(c) Given the above tree-decomposition algorithm for concurrent graphs $G$, in Section \ref{sec:upperbounds} 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 \ref{app:code}. 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 \cite{Ohtsuki91}. 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 $t$ \cite{Ohtsuki91}. 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 \cite{Fomin}. 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 \cite{Courcelle90} showed that for constant-treewidth graphs the decision questions for MSO can be solved in polynomial time; and the result of \cite{Courcelle90} 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 \cite{Karpinski2002}. The problem of computing a balanced tree decomposition for a constant treewidth graph was considered in \cite{Bodlaender92}. More importantly, in the context of programming languages, it was shown by \cite{Krivine96} that the control-flow 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 gotos the treewidth is not guaranteed to be bounded, it has been shown that programs in several programming languages have low treewidth in practice (\cite{Godefroid96}). The constant-treewidth property of graphs has been used to develop faster algorithms for sequential interprocedural analysis \cite{Brendel96}, and on the analysis of automata with auxiliary storage (e.g., stacks and queues) \cite{Datta08}. The techniques of \cite{Brendel96} 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 \cite{Brendel96} \cite{Krivine96} \cite{Brandt97} \cite{Desel98} \cite{Godefroid96} \cite{Krivine96} \cite{Krivine96}, and many practical tools have been developed as well \cite{Krivine96} \cite{Krivine96} \cite{Krivine96} \cite{Krivine96}. 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 $x_1, \ldots, x_n$ is denoted for short by $(x_i)_{1 \leq i \leq k}$, or $(x_i)$, when $k$ is implied from the context. Given a sequence $Y$, we denote by $y \in Y$ the fact that $y$ appears in $Y$.

2.1 Complete, closed semirings

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

1. $\otimes$ is associative, commutative, and $\mathbf{0}$ is the neutral element,
2. $\uplus$ is associative, and $\mathbf{1}$ is the neutral element,
3. $\otimes$ distributes over $\uplus$,
4. $\uplus$ is infinitely associative,
5. $\otimes$ infinitely distributes over $\uplus$,
6. $\mathbf{0}$ absorbs in multiplication, i.e., $\forall a \in \Sigma : a \otimes \mathbf{0} = \mathbf{0}$.

Additionally, we consider that $S$ is equipped with a closure operator $s^*$, such that $\forall s \in \Sigma : s^* = \mathbf{1} \uplus (s \otimes s^*) = \mathbf{1} \uplus (s^* \otimes s)$ (i.e., the semiring is closed).
2.2 Graphs and tree decompositions

Graphs and weighted paths. Let $G = (V, E)$ be a weighted finite directed graph (henceforth called simply a graph) where $V$ is a set of $n$ nodes and $E \subseteq V \times V$ is an edge relation, along with a weight function $w : E \to \Sigma$ 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 $P : u \leadsto v$ 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 $\otimes(P) = \otimes(\{w(x_i, x_{i+1})\})$ if $|P| \geq 1$ else $\otimes(P) = 1$. Given nodes $u, v \in V$, the semiring distance $d(u, v)$ is defined as $d(u, v) = \sum_{u \leadsto v} \otimes(P)$, and $d(u) = \emptyset$ 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 $L(u)$ 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 $L(v) = L(u) + 1$ 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 $v$ is an ancestor of $u$, then $u$ is a descendant of $v$. Given two nodes $u, v$, the lowest common ancestor (LCA) is the common ancestor of $u$ and $v$ with the highest level. Given a tree $T$, a contiguous subtree is a 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 $T$ is k-acyclic 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-acyclic tree, every node has a child exactly k children.

Tree decompositions. A tree-decomposition $T = (\mathcal{T}, \mathcal{B})$ 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. $\mathcal{V}_T = \{B_0, \ldots, B_k\}$ with $B_0 \subseteq V$, and $\bigcup_{B_i \in \mathcal{T}} B_i = V$ (every node is covered).
2. For all $(u, v) \in E$ there exists $B_i \in \mathcal{T}$ such that $u, v \in B_i$ (every edge is covered).
3. For all $i, j, k$ such that there is a bag $B_k$ that appears in the simple path $B_i \leadsto B_j$ in $T$ (Tree(G)), we have $B_i \cap B_j \subseteq B_k$ (every node appears in a contiguous subtree of $T$).

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

1. $V_T(B)$ the nodes of $G$ that appear in bags in $T[B]$,
2. $\mathcal{V}_T(B)$ the nodes of $G$ that appear in $B$ and its ancestors in $T$.

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

2.3 Concurrent graphs

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

![Figure 2: A graph G with treewidth 2 (left) and a corresponding tree-decomposition T = Tree(G) of 8 bags and width 2 (right).](image-url)

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

we have $(u, v) \in E_{\cong}$ iff there exists a set $I \subseteq [k]$ such that (i) $(u_i, v_i) \in E_i$ for all $i \in I$, and (ii) $u_i = v_i$ for all $i \notin I$. In words, an edge $(u, v) \in E_{\cong}$ is formed in the product graph by traversing a set of edges $\{(u_i, v_i) \in E_i\}_{i \in I}$ in some component graphs $\{G_i\}_{i \in I}$, and traversing no edges in the remaining $\{G_j\}_{j \notin I}$.

Concurrent graphs. A graph $G = (V, E)$ is called a concurrent graph of $k$ graphs $G_1 = (V_1, E_1), \ldots, G_k = (V_k, E_k)$, where $G_\cong = (V_{\cong}, E_{\cong})$ is the product graph of $\{G_i\}_{i \in I}$. 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 $G_\cong$ is a k-self-product of a graph $G_i' \cong G_i$ if $G_{\cong}$ is the k-self-product of $G_i$.

Various notions of composition. The framework we consider is quite general as it captures various different notions of concurrent composition. Indeed, the edge set of the concurrent graph is any possible subset of the edge set of the corresponding product graph. Then, various well-known composition notions can be modeled as follows. For any edge $(u, v) \in E$ of the concurrent graph $G$, let $T_{\cong, uv} = \{i \in [k] : (u_i, v_i) \in E_i\}$ 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 $T_{\cong, uv} = |k|$.
2. In asynchronous composition at every step only one component makes a move. This is captured by $|T_{\cong, uv}| = 1$ 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 $\overline{v}$ of a concurrent graph $G$ is an element of $\bigcup_{i \in I} (V_i \cup \{\perp\})$, where $\perp \notin \bigcup_{i \in I} V_i$. Intuitively, $\perp$ is a fresh symbol to denote that a component is unspecified. A partial node $\overline{v}$ is said to refine a partial node $\overline{v}$, denoted by $\overline{v} \sqsubseteq \overline{v}$ if for all $1 \leq i \leq k$ either $\overline{v}_i = \perp$ or $\overline{v}_i = \overline{v}_i$. We say that the partial node $\overline{v}$ strictly refines $\overline{v}$, denoted by $\overline{v} \sqsubseteq ^* \overline{v}$, if $\overline{v} \sqsubseteq \overline{v}$ and $\overline{v} \sqsubseteq \overline{v}$ (i.e., for at least one constituent $i$ we have $\overline{v}_i = \perp$ but $\overline{v}_i \neq \perp$). A partial node $\overline{v}$ is called strictly partial if it is strictly refined by some node $u \in V$ (i.e., $\overline{v}$ has at least one $\perp$). The notion of semiring distances is extended to partial nodes, and for partial nodes $\overline{u}, \overline{v}$ of $G$ we define the semiring distance from $\overline{u}$ to $\overline{v}$ as $d(\overline{u}, \overline{v}) = \sum_{u \sqsubseteq \overline{v}, v \sqsubseteq \overline{v}} d(u, v)$ where $u, v \in V$. In the sequel, a partial node $\overline{v}$ will be either (i) a node of $V$, or (ii) a strictly partial node. We refer to nodes of the first
Method: DiningPhilosophers
1 while True do
2 while fork not mine or knife not mine do
3 if fork is free then
4 lock(fork)
5 acquire(fork)
6 unlock(fork)
7 end
8 if knife is free then
9 lock(knife)
10 acquire(knife)
11 unlock(knife)
12 end
13 end
14 dine(fork, knife) // for some time
15 lock(fork)
16 release(fork)
17 release(knife)
18 unlock(fork)
19 discuss() // for some time
20 end

Figure 3: A concurrent program (left), its 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 $\omega : E \to \Sigma$ be a weight function that assigns to every edge of $G$ a weight from a set $\Sigma$ that forms a complete, closed semiring $\Sigma = (\Sigma, \circ, \odot, 1, 0, i, T)$. The algebraic path problem on $G$ asks the following types of queries:

1. Single-source query. Given a partial node $\pi \subseteq G$, 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 $\pi, \sigma \subseteq G$ where at least one is strictly partial, return the distance $d(\pi, \sigma)$.

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 are 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 $|V(B_i)| - |V(B)| = \gamma$, 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 $Tree(G)$ 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 $Tree'(G)$ of width $t$ and $O(n)$ bags is given (which can be obtained using e.g. Algorithm 12 in $O(n)$ time). Given two parameters $\delta > 0$ and $\lambda \in \mathbb{N}$ with $\lambda \geq 2$, $Tree'(G)$ 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 $R_\alpha$ is constructed, which is $(\beta, \gamma)$-balanced.
2. $R_\alpha$ is turned to an $\alpha$-approximate tree decomposition of $G$.
Figure 4: Illustration of one recursive step of Rank on a component \( C \) (gray). \( C \) is split into two sub-components \( \overline{C}_1 \) and \( \overline{C}_2 \) by removing a list of bags \( X = (B_i)_i \). Once every \( \lambda \) recursive calls, \( X \) contains one bag, such that the neighborhood \( N_h(C)_i \) of each \( C_i \) is at most half the size of \( N_h(C) \) (i.e., the red area is split in half). In the remaining \( \lambda - 1 \) recursive calls, \( X \) contains \( m \) bags, such that the size of each \( \overline{C}_i \), at most \( \frac{2}{\lambda} \times \frac{1}{2} \) fraction the size of \( C \), (i.e., the gray area is split in almost half).

Tree'(G), and \( \ell \in [\lambda] \) specifies the type of operation the algorithm performs on \( G \). Given such a component \( C \), we denote by \( N_h(C) \) the neighborhood of \( C \), defined as the set of bags of Tree'(G) that are incident to \( C \). Informally, on input \((C', \ell')\), the algorithm partitions \( C \) into two sub-components \( \overline{C}_1 \) and \( \overline{C}_2 \) such that either (i) the size of each \( \overline{C}_i \) is approximately half the size of \( C \), or (ii) the size of the neighborhood of each \( \overline{C}_i \) is approximately half the size of the neighborhood of \( C \). More specifically,

1. If \( \ell > 0 \), then \( C \) is partitioned into components \( \mathcal{V} = (C_1, \ldots, C_r) \), by removing a list of bags \( X = (B_1, \ldots, B_m) \), such that \( |C_i| \leq \frac{\ell}{\lambda} |C| \). The union of \( X \) yields a new bag \( B \) in \( R_C \). Then \( \mathcal{V} \) is merged into two components \( \overline{C}_1, \overline{C}_2 \) with \( |\overline{C}_1| \leq |\overline{C}_2| \leq \frac{\ell+1}{\lambda} |C| \). Finally, each \( \overline{C}_i \) is passed on to the next recursive step with \( \ell = (\ell + 1) \mod \lambda \).

2. If \( \ell = 0 \), then \( C \) is partitioned into two components \( \overline{C}_1, \overline{C}_2 \) such that \(|N_h(\overline{C}_1) \cap N_h(\overline{C}_2)| \leq \frac{|N_h(C)|}{\lambda} \) by removing a single bag \( B \). This bag becomes a new bag \( B \) in \( R_C \), and each \( \overline{C}_i \) is passed on to the next recursive step with \( \ell = (\ell + 1) \mod \lambda \).

Figure 4 provides an illustration. The second construction is obtained simply by inserting in each bag \( B \) in \( R_C \) the nodes contained in the neighborhood \( N_h(\overline{C}) \) of the component \( C \) upon which \( B \) was constructed. In Appendix B, we prove that these two steps yield an \((\alpha, \beta, \gamma)\) tree-decomposition of \( C \).

Use of \((\alpha, \beta, \gamma)\) tree-decompositions. For ease of presentation we consider that every Tree'(G) is a full binary tree. Let \( \alpha = 4 \cdot \lambda/\beta, \beta = (1 + \delta)/2 \beta - 1\) 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 Tree'(G) by a constant factor only. In the sequel, 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 of the 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 ConcurTree which constructs a tree-decomposition ConcurTree(G) of \( G \), given a \((\alpha, \beta, \gamma)\) tree-decomposition of the components, in \( O(n^k) \) time and space (i.e., linear in the size of \( G \)), such that the following properties hold: (i) the width is \( O(n^{k-1}) \); and (ii) for every bag in level \( \ell \), the size of the bag is \( O(n^{k-1} \cdot \beta^\ell) \).

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

1. The root bag \( B \) is

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

(1)

2. If every \( B \) is a non-leaf bag of \( T \), for every choice of \((r_1, \ldots, r_k) \in [2]^k\), repeat the procedure for \((T_i(B_{i,r_i}))_{i \in [k]}\), 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 ConcurTree(G) the application of the recursive procedure ConcurTree on \((T_i(B_i))_{i \in [k]}\). Figure 5 provides an illustration.

Remark 2. Recall that for any bag \( B \) of a tree-decomposition \( T \), we have \( V_{T_j}(B_j) = \bigcup_{j \neq i} B_j \) where \( B_j \) ranges over bags in \( T_j(B_i) \). Then, for any two bags \( B_{i1}, B_{i2} \), of tree-decompositions \( T_{i1} \) and \( T_{i2} \) respectively, we have

\[
V_{T_{i1}}(B_{i1}) \times V_{T_{i2}}(B_{i2}) = \bigcup_{B'_{i1}, B'_{i2}} (B'_{i1} \times B'_{i2})
\]

where \( B'_{i1} \) and \( B'_{i2} \) range over bags in \( T_{i1}(B_{i1}) \) and \( T_{i2}(B_{i2}) \) respectively. Since each tree-decomposition \( T_i \) has constant width, it follows that \( |V_{T_{i1}}(B_{i1}) \times V_{T_{i2}}(B_{i2})| = O(|T_{i1}(B_{i1})| \cdot |T_{i2}(B_{i2})|) \). Thus, the size of each bag \( B \) of ConcurTree(G) constructed in Eq. (1) on some input \((T_i(B_i))_i\), is \( |B| = O(\sum_{i \neq j} n_j) \), where \( n_i = |T_i(B_i)| \).

In view of Remark 2 the time and space required by ConcurTree to operate on input \((T_i(B_i))_{i \in [k]}\) where \( |T_i(B_i)| = n_i \) is given, up to constant factors, by
5. Concurrent Algebraic Paths

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

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

\[\text{Lemma 1.} \] Consider a graph \( G = \langle V, E \rangle \) with a weight function \( wt : E \rightarrow \Sigma \) and a tree-decomposition \( \text{Tree}(G) \). 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 \( A = \{u\} \times \left( \bigcup_{1 \leq i < j} (B_{i+1} \setminus B_i) \right) \times \{v\} \). Then \( d(u, v) = \bigoplus_{(x_1, \ldots, x_{j+1}) \in A} \bigotimes_{i=1}^{j} d(x_i, x_{i+1}) \).

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

1. First, the partial expansion \( \Gamma \) of \( G \) is constructed by introducing a pair of strictly partial nodes \( \pi, \pi' \) for every strictly partial node \( \pi \) of \( G \), and edges between strictly partial nodes and the corresponding nodes of \( G \) that refine them.

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

3. Third, a standard, two-way pass of \( T \) is performed to compute local distances. In this step, for every bag \( B \) in \( T \) and all partial nodes \( \pi, \pi' \in B \), the distance \( d(\pi, \pi') \) is computed (i.e., all-pair distances in \( B \)). Since we compute distances between nodes that are local in a bag, this step is called local distance computation. This information is used to handle (i) single-source queries and (ii) partial pair queries in which both nodes are strictly partial.

4. Finally, a top-down pass of \( T \) is performed in which for every node \( u \) and partial node \( \pi \in T \), the distances \( d(\pi, u) \) and \( d(\pi, u) \) are computed. This information is used to handle pair queries in which at least one node is a node of \( G \) (i.e., not strictly partial).

Algorithm ConcurPreprocess. We now formally describe algorithm ConcurPreprocess for preprocessing the concurrent graph \( G = \langle V, E \rangle \) for the purpose of answering algebraic path queries. For any desired \( \epsilon < \gamma \), choose a \( \lambda \in \mathbb{N} \) and \( \delta \in \mathbb{R} \) such that \( \lambda \geq 2 \delta \) and \( \delta \leq \epsilon \). Let \( \alpha = 4 \cdot \lambda \cdot \delta \cdot \beta = ((1 + \delta) / \lambda)^{\gamma - 1} \) and \( \gamma = \lambda \). On input \( G = \langle V, E \rangle \), where \( G \) is a concurrent graph of \( k \) constant-treewidth graphs \( (G_i = \langle V_i, E_i \rangle)_{1 \leq i \leq k} \) and a weight function \( wt : E \rightarrow \Sigma \), ConcurPreprocess operates as follows:

1. Construct the partial expansion \( \Gamma = \langle \overline{V}, \overline{E} \rangle \) of \( G \) together with an extended weight function \( \overline{wt} : E \rightarrow \Sigma \) as follows.

2. Construct the tree-decomposition \( T = \langle \overline{V}, \overline{E} \rangle \) of \( \Gamma \) as follows.

3. Perform the local distance computation on \( T \) as follows. For every partial node \( \pi \), maintain two data-structures \( \text{FWD}^\pi, \text{BWD}^\pi \) for all partial nodes \( \pi \) in \( \overline{E} \). For every pair of partial nodes \( \pi, \pi' \) with \( \text{Lvl}(\pi) \leq \text{Lvl}(\pi') \) we have that \( \overline{wt}(\pi, \pi') \) is computed (i.e., all-pair distances in \( B \)). Since we compute distances between nodes that are local in a bag, this step is called local distance computation. This information is used to handle (i) single-source queries and (ii) partial pair queries in which both nodes are strictly partial.

4. Finally, a top-down pass of \( T \) is performed in which for every node \( u \) and partial node \( \pi \in T \), the distances \( d(\pi, u) \) and \( d(\pi, u) \) are computed. This information is used to handle pair queries in which at least one node is a node of \( G \) (i.e., not strictly partial).
have

Lemma 2. \( \mathcal{T} \) is a tree decomposition of the partial expansion \( \mathcal{G} \).

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

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

1. For all nodes \( u, v \in V \) such that \( \mathcal{B}_u \) is in \( \mathcal{T}(\mathcal{B}_v) \), we have \( \text{FWD}_d^u(v) = d(u, v) \) and \( \text{BWD}_d^u(v) = d(v, u) \).
2. For all strictly partial nodes \( \tau \in \mathcal{V} \) and nodes \( u \in V \) we have \( \text{FWD}_d^d(\tau^2) = d(\pi, u) \) and \( \text{BWD}_d^d(\tau^2) = d(u, \pi) \).
3. For all strictly partial nodes \( \tau, \pi \in \mathcal{V} \) we have \( \text{FWD}_d^d(\tau^2) = d(\pi, \tau) \) and \( \text{BWD}_d^d(\tau^2) = d(\tau, \pi) \).

Proof. We describe the key invariants that hold during the traversals of \( \mathcal{T} \) by ConcurPreprocess in Item [3a] Item [3b] and Item [4]. After the algorithm processes a bag \( \mathcal{B} \), all cases depend on Lemma [4].

\( \mathbf{Item \ [3a]} \) For every pair of partial nodes \( \tau, \pi \in \mathcal{B} \) such that \( \text{Ltv}(\tau) \leq \text{Ltv}(\pi) \) we have \( \text{FWD}_d(\tau) = \bigoplus \text{FWD}_d(\pi) \) and \( \text{BWD}_d(\tau) = \bigoplus \text{BWD}_d(\pi) \), where \( \bigoplus \) represents the union of the maps. The statement follows by a straightforward induction on the levels processed by the algorithm in the bottom-up pass. Note that if \( \pi \) and \( \tau \) are partial nodes in the root of \( \mathcal{T} \), the statement yields \( \text{FWD}_d(\tau) = d(u, v) \) and \( \text{BWD}_d(\tau) = d(v, u) \).

\( \mathbf{Item \ [3b]} \) For every node \( u \in \mathcal{B} \) and partial node \( \pi \in \mathcal{V} \), we have \( \text{FWD}_d(\tau) = d(u, \pi) \) and \( \text{BWD}_d(\tau) = d(\pi, u) \). 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 [3a]. Similarly for statement 2, together with the observation that every strictly partial node \( \tau \) appears in the root of \( \mathcal{T} \), and thus \( \tau \in \mathcal{V}(\mathcal{B}_u) \). 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.

\( \mathbf{Lemma \ 4. \ Consider \ the \ recurrences \ in \ Eq. \ (5) \ and \ Eq. \ (6).} \)

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

Then

1. \( \mathcal{T}_k(n) = O(n^{3(k-1)}) \), and
2. (i) \( \mathcal{S}_k(n) = O(n^{2(k-1)}) \) if \( k \geq 3 \), and (ii) \( \mathcal{S}_2(n) = O(n^{2.5}) \).

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

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

Proof. We examine each step of the algorithm separately.

1. The time and space required for this step is bounded by the number of nodes introduced in the partial expansion \( \mathcal{G} \), which is \( 2 \cdot \sum_{i<k} i \cdot n^i = O(n^{k-1}) \).
2. By Theorem [2], ConcurTree(\( \mathcal{G} \)) is constructed in \( O(n^k) \) time and space. In \( \mathcal{T} \), 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 $|\mathcal{B}|^3$ time to perform an all-pairs algebraic paths computation in each bag $B$ of $T$ [33 [49] [50] [22]. The space usage for storing all maps $FWD_B$ and $BWD_B$ for every node $\pi$ whose root bag is $B$ is $O(|\mathcal{B}|^2)$, since there are at most $|\mathcal{B}|$ such bags $\pi$, and each map has size $|\mathcal{B}|$. By the previous item, we have $|\{B\}| = O(|\mathcal{B}|)$, where $B$ is the corresponding bag of $T$ before the partial expansion of $G$. By Theorem 2, we have $|B| = O(n^{\beta-1})$, where $l_T(v) \leq O(x)\leq O(n^{\beta-1})$, 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 $\gamma = \lambda$ levels of $T$ is given by the following recurrences respectively (ignoring constant factors for simplicity).

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

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

By the analysis of Eq. (3) and Eq. (4) of Lemma 3, 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 $B_u$ denote the ancestor bag of $B_u$ at level $i$. We have

$$|\mathcal{B}(B_u)| = \sum_i |B_u| \leq c_1 \cdot \sum_i |B_u^{i(\gamma)}| \leq c_2 \cdot \sum_i |B_u^{i(\lambda)}|$$

for some constants $c_1, c_2, c_3$. The first inequality comes from expressing the size of all (constantly many) ancestors $B_u$ with $[i/\gamma] = j$ as a constant factor the size of $B_u^{i(\gamma)}$. The second inequality comes from the analysis of Eq. (3) and Eq. (4) of Lemma 3, which states that $O(|\mathcal{B}|) = O(|\mathcal{B}|)$ for every bag $B$. The third inequality comes from Theorem 2. By Item 2, there are $O(n^k)$ such nodes $u$ in $T$, hence the space required is $O(n^{2^{(k-1)}})$.

We now turn our attention to the time requirement. For every bag $B_u$, the algorithm requires $O(|\mathcal{B}|^2)$ time to iterate over all pairs of nodes $u$ and $v$ in $B_u$ and $B_v$ in $T$, and for every $\pi \in T_v(B_u)$, the data-structure $FWD_{\pi}(\pi)$ and $BWD_{\pi}(\pi)$ are stored in $T_v(B_u)$. Hence the time required for all nodes $u$ and every node partial $\pi \in T_v(B_u)$ to store the maps $FWD_{\pi}(\pi)$ and $BWD_{\pi}(\pi)$ is given by the recurrence

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

The analysis of Eq. (3) and Eq. (4) of Lemma 3 gives $T_k(n) = O(n^{2^{(k-1)}})$ and $T_3(n) = O(n^{2^{(k-1)}})$ (i.e., the above time recurrence is analyzed as the recurrence for $S_k$ of Lemma 3). From the space analysis we have that there exist $O(n^{2^{(k-1)}})$ partial nodes $\pi \in T_v(B_u)$ for every node $u$ whose root bag is $B_u$. Hence the total time for this step is $O(n^{2^{(k-1)}})$ for $k \geq 3$, and $O(n^{2^{(k-1)}})$ for $k = 2$.

5. This step requires time linear in the size of $T$ [40].

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 $B_u$, and for every encountered bag $B$ and nodes $x$, $v \in B$ with $l_T(x) \leq l_T(v)$, set $A(v) = A(v) \cup (A(x) \cap FWD_B(x))$. Return the map $A$. 

2. Pair query. Given two nodes $u, v, x \in B$, find the LCA $B_u$ of bags $B_u$ and $B_v$. Return $\bigoplus_{\pi \in \mathcal{V}} (FWD_{\pi}(x) \cap BWD_{\pi}(x))$.

3. Partial pair query. Given two partial nodes $\pi, \eta$. (a) If both $\pi$ and $\eta$ are strictly partial, return $FWD_{\eta}(\pi)^2$, else (b) if $\pi$ is strictly partial, return $BWD_{\eta}(\pi)^2$, else (c) Return $FWD_{\eta}(\eta)$. 

We thus establish the following theorem.

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

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

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

4. Pair query time $O(n^{2^{(k-1)}})$.

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

Proof. The correctness of ConcurQuery for handling all queries follows from Lemma 3, and the properties of the preprocessing established in Lemma 3. The preprocessing complexity is stated in Lemma 3. 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. (4) in Lemma 3. The time complexity for the pair query follows from the O(1) time to access the LCA bag $B_u$ of $B_u$ and $B_v$, and the $O(|\mathcal{B}|) = O(n^{2^{(k-1)}})$ time required to iterate over all nodes $x \in B \cap V$. Finally, the time complexity for the partial pair query follows from the O(1) time lookup in the constructed maps $FWD, BWD^+$.

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-treewidth graphs $G_1, G_2$, and $\psi : E \rightarrow \Sigma$ 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 $O(n^k)$.

3. Single-source query time $O(n^{2^{(k-1)}})$.

4. Pair query time $O(n^{2^{(k-1)}})$.

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 [31 [49] [50] [22] 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 2 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 $\pi \rightarrow \pi$, and spending
O(|B|^2) time for each bag B of P. Recall (Theorem 1) that the size of each bag B in T (and thus the size of the corresponding bag B in T) decreases geometrically every γ times. Then, the time required for this operation is O(|B|^2) = O(n^α), where B is the bag of P with the smallest level. This leads to the following corollary.

**Corollary 2.** Let G = (V, E) be a concurrent graph of two constant-treewidth graphs G_1, G_2, and ωt : E → Σ a weight function of G. For any fixed ε, the data-structure 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).
3. Single-source query time O(n^2).
4. Pair and partial pair query time O(n^2).

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^2) by Corollary 2 and the space of the output is O(n^4), since there are n^4 pairs for the computed distances. Hence the total space requirement is O(n^4). The time requirement is O(n^2), since by Corollary 2 every single-source query requires O(n^2) 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 ωt : E → Σ a weight function of G. For any fixed ε > 0, the transitive closure of G with respect to a semiring can be computed in O(n^4) time and O(n^4) 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·n nodes, such that querying semiring queries in G coincide with querying 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 wt : E → Σ. Let x_i, 1 ≤ i ≤ n run over the nodes V of G, and construct a graph G'' = (V'', E'') such that V'' = \{x_i, y_i : 1 ≤ i ≤ n\} and E'' = \{(x_i, y_{i+1}), (y_{i+1}, x_{i+1}) : 1 ≤ i < n\} ∪ \{(y_i, x_{i+1}) : 1 ≤ i ≤ n\}. Let G'' be the corresponding graph.

**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 ≤ i ≤ n and 1 ≤ j ≤ n we have \((x_i, y_j), (x_j, y_{j+1}), (x_j, y_i), (x_i, y_{i+1})\) ∈ E', and for all 1 ≤ i < n and 1 ≤ j ≤ n we have \((x_i, x_{i+1}), (y_i, x_{i+1}), (y_{i+1}, x_{i+1})\) ∈ E'.

2. **Blue edges.** For all 1 ≤ i ≤ n we have \((x_i, x_i), (x_i, y_i), (y_i, x_i), (x_i, x_{i+1})\) ∈ E'.

3. **Red edges.** For all \((x_i, x_j) ∈ E\) we have \((x_i, y_j), (x_j, y_i)\) ∈ E'.

4. **Green edges.** For all 1 ≤ i, j ≤ n with i ≠ j we have \((x_i, x_j), (x_j, x_i)\) ∈ E'.

Additionally, we construct a weight function such that wt'(\(x_i, y_j), (x_j, x_i)\) = wt(x_i, x_j) for every red edge \((x_i, x_j), (x_j, x_i)\), and wt'(u, v) = \(\overline{1}\) 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) in G''.

**Lemma 6.** For every x_i, x_j ∈ V, there exists a path P : x_i ↦ x_j with wt(P) = z in G if and only if there exists a path P' : x_i ↦ x_j with wt(P') = z in G''.

This leads to the following corollary.

**Corollary 4.** Let T_{\mathcal{S}}(n) = \(Ω(n^2)\) be a lower bound on the time required to answer a single algebraic paths query wrt to a semiring S on arbitrary graphs of n nodes. Consider any concurrent graph G which is an asynchronous self-composition of two constant-treewidth graphs of n nodes each. For any data-structure DS, let T_{\mathcal{DS}}(G, r) be the time required by DS to preprocess G and answer r parallel queries. We have T_{\mathcal{DS}}(G, 1) = \(Ω(T_{\mathcal{S}}(n))\).

**Conditional optimality of Corollary 2.** Note that for r = O(n) parallel queries, Corollary 2 yields that the time spent by our data-structure ConcurAP for preprocessing G and answering r queries on a 2-self-concurrent asynchronous composition of a constant-treewidth graph together with a weight function wt : E → Σ is such that for all u, v ∈ V, and (u, w), (v, w) ∈ V' we have d(u, v) = d(u, w) + d(w, v), and hence parallel queries in G also coincide with partial parallel queries in G''.

**Theorem 4.** For every graph G = (V, E) and weight function wt : E → Σ there exists a graph G' = (V × E, E') such that is a 2-self-concurrent asynchronous composition of a constant-treewidth graph, together with a weight function wt' : E' → Σ, such that for all u, v ∈ V, and (u, w), (v, w) ∈ V' we have d(u, v) = d(u, w) + d(w, v), and moreover, the graph G' can be constructed in quadratic time in the size of G.

This leads to the following corollary.

**Corollary 4.** Let T_{\mathcal{S}}(n) = \(Ω(n^2)\) be a lower bound on the time required to answer a single algebraic paths query wrt to a semiring S on arbitrary graphs of n nodes. Consider any concurrent graph G which is an asynchronous self-composition of two constant-treewidth graphs of n nodes each. For any data-structure DS, let T_{\mathcal{DS}}(G, r) be the time required by DS to preprocess G and answer r parallel queries. We have T_{\mathcal{DS}}(G, 1) = \(Ω(T_{\mathcal{S}}(n))\).

Almost-optimality of Theorem 3 and Corollary 3. Finally, we highlight some almost-optimality results obtained by variants of ConcurAP for the case of two graphs. By almost-optimality we mean that the obtained bounds are O(n^4) factor worse that optimal, for any fixed ε > 0 arbitrarily close to 0.

1. According to Theorem 3, after O(n^{3+ε}) preprocessing time, single-source queries are handled in O(n^{4+ε}) time, and partial parallel queries in O(1) time. The former (resp. later) query time is almost linear (resp. exactly linear) in the size of the output. Hence the former queries are handled almost-optimally, and the latter indeed optimally. Moreover, this is achieved using O(n^{3+ε}) preprocessing time, which is far less than the \(Ω(n^4)\) time required for the transitive closure computation (which computes the distance between all n^4 pairs of nodes).
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 $\lambda$.

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>%</td>
<td>6</td>
<td>7</td>
<td>16</td>
<td>22</td>
<td>25</td>
<td>67</td>
<td>17</td>
</tr>
</tbody>
</table>

2. According to Corollary 3, the transitive closure can be computed in $O(n^{3+\epsilon})$ time, for any fixed $\epsilon > 0$, and $O(n^4)$ space. Since the size of the output is $O(n^3)$, 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 $O(n^{3+\epsilon})$, for any fixed $\epsilon > 0$, which is achieved by choosing a sufficiently large $\lambda \in \mathbb{N}$ and a sufficiently small $\delta \in \mathbb{R}$ when running the algorithm of Theorem 1. We compute the transitive closure for various $\lambda$. In practice, $\delta$ has effects only for very large input graphs. For this, we fix it to a large value ($\delta = 1/3$) which can be proved to have no effect on the obtained running times. Table 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 seems to scale much better than its theoretical worst-case bound of $O(n^{3+\epsilon})$ of Corollary 3.

Concurrence with locks. Our second set of experiments is on methods from containers of the java.util.concurrent library that use locks as their synchronization mechanism. The average treewidth of the input graphs is around 8. In this case, we expand the node set of the concurrent graph $G$ with the lock set $[3]^\ell$, where $\ell$ is the number of locks used by $G'$. Intuitively, the $i$-th value of the lock set denotes which one of the two components owns the $i$-th lock (the value is 3 if the lock is free). Transitions to nodes that perform lock operations are only allowed wrt the lock semantics. 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.

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 intraprocedural 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.
References


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 \{True, False\}. We note, however, that case (1) still requires an analysis on the concurrent graph \(G\).

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 \[64, 66\]. Given a finite domain \(D\) and a universe \(F\) of distributive dataflow functions \(f : 2^D \rightarrow 2^D\), a weight function \(w_t : E \rightarrow F\) 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 \((\cap)\) (union or intersection) of the weights of all \(u \xrightarrow{e} v\) paths. The problem can be formulated on the meet-composition semiring \((F, \cap, \lor, \downarrow, I)\), where \(\downarrow\) is the identity function. We note, however, that the IFDS/IDE framework considers interprocedural paths in sequential programs. In contrast, the current work focuses on intraprocedural analysis of concurrent programs. The dataflow analysis of concurrent programs has been a problem of intensive study (e.g. \[26, 28, 32, 45, 50\]), 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 \[6, 15, 35, 44, 12\], 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 \[16, 23, 47, 55, 71\]. 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 the system makes in a path between two nodes of the controlflow graph forms a regular language \[19, 30, 33\]. Kleene algebras have also been used as algebraic relaxations of interprocedurally valid paths in sequential and concurrent systems \[13, 74\].

Modeling example

Figure 2 illustrates the introduced notions in a small example of the well-known \(k\) dining philosophers problem. For the purpose of the example, lock is considered a blocking operation. Consider the case of \(K = 2\) threads being executed in parallel. The graphs \(G_1\) and \(G_2\) that correspond to the two threads have nodes of the form \((i, \ell)\), where \(i \in \{0\}\) is a node of the controlflow graph, and \(\ell \in \{0\}\) denotes the thread that controls the lock \(\ell = 3\) denotes that \(\ell\) is free, whereas \(\ell = 1 \in \{2\}\) denotes that it is acquired by thread \(i\). The concurrent graph \(G\) is taken to be the asynchronous composition of \(G_1\) and \(G_2\), and consists of nodes \((x, y)\), where \(x\) and \(y\) is a node of \(G_1\) and \(G_2\) respectively, such that \(x\) and \(y\) agree on the value of \(\ell\) (all other nodes can be discarded). For brevity, we represent nodes of \(G\) as triplets \((x, y, \ell)\) where now \(x\) and \(y\) are nodes in the controlflow graphs \(G_1\) and \(G_2\) (i.e., without carrying the value of the lock), and \(\ell\) is the value of the lock. A transition to a node \((x, y, \ell)\) in which one component \(G_i\) performs a lock is allowed only from a node where \(\ell = 3\), and sets \(\ell = i\) in the target node (i.e., \((x, y, i)\)). Similarly, a transition to a node \((x, y, \ell)\) in which one component \(G_i\) performs an unlock is allowed from a node where \(\ell = i\), and sets \(\ell = 3\) in the target node (i.e., \((x, y, 3)\)).

Suppose that we are interested in determining (1) whether the first thread can execute dine(fork, knife) without owning fork or knife, and (2) whether a deadlock can be reached in which each thread owns one resource. These questions naturally correspond to partial pair and pair queries respectively, as in case (1) we are interested in a local property of \(G_i\), whereas in case (2) we are interested in a global property of \(G\). We note, however, that case (1) still requires an analysis on the concurrent graph \(G\).

Local property as a partial pair query. Assume that we are interested in determining whether the first thread can execute dine(fork, knife) without owning fork or knife. A typical datafact set is \(D = \{\text{fork, knife, null}\}\), where each datafact denotes that the corresponding resource must be owned by the first thread. The concurrent graph \(G\) is associated with a weight function \(w_t\) of dataflow functions \(f : 2^D \rightarrow 2^D\). The dataflow function \(w_t(e)\) 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 fork or the first thread releases fork, then \(w_t(e)(\text{fork}) \rightarrow \text{null}\) (i.e., fork is removed from the datafacts).
2. Else, if \(e\) transitions to a node in which the first thread acquires fork, then \(w_t(e)(\text{null}) \rightarrow \text{fork}\) (i.e., fork is inserted to the datafacts).

Similarly for the \(F = \text{knife}\) datafact. The “meet-over-all-paths” operation is set intersection. Then the question is answered by testing whether \(d(\langle 1, 1, 3 \rangle, \langle 1, 1, 3 \rangle) = \{\text{fork, knife}\}\), 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 fork, and the second thread may hold the knife. A typical datafact set is \(D = 2^\{(\text{fork, knife})\}\). For a datafact \(F \in D\) we have

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

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

1. If \(e\) transitions to a node in which the second thread acquires fork or the first thread releases fork, then \(w_t(e)(\text{fork}) \rightarrow F \cup \{\text{fork}\}\) (i.e., the first thread now owns fork).
2. If \(e\) transitions to a node in which the first thread acquires fork, then \(w_t(e)(\text{null}) \rightarrow F \cup \{\text{fork}\}\) (i.e., the first thread now owns fork).
3. If \(e\) transitions to a node in which the first thread acquires knife or the second thread releases knife, then \(w_t(e)(\text{knife}) \rightarrow F \cup \{\text{knife}\}\) (i.e., the second thread now owns knife).
4. If \(e\) transitions to a node in which the second thread acquires knife, then \(w_t(e)(\text{null}) \rightarrow F \cup \{\text{knife}\}\) (i.e., the second thread now owns 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 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

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

We show how to construct a graph $G$ of treewidth $t$ and a tree-decomposition $T$ of $G$ in time $O(b \cdot \log b)$ and space $O(b)$ for 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, 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 of $T$ is a subset of bags of $T$. The neighborhood $Nh(C)\subseteq C$ of the set of bags in $V_T \setminus C$ that have a neighbor in $C$, i.e.

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

Given a component $C$, we define the operation Split as $Split(C) = (X, Y)$, where $X \subseteq C$ is a list of bags $(B_1, \ldots, B_{\ell-1})$ and $Y$ is a list of sub-components $(C_1, \ldots, C_{\ell})$ such that removing each bag $B_i$ from $C$ splits $C$ into subcomponents $Y_i$ and, for every $i$ we have $|C_i| \leq \frac{\ell}{4 \cdot |C|}$. 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 $\frac{\ell}{2}$. 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 $Y = (C_1, \ldots, C_{\ell})$, and let $z = \sum |C_i|$. Let $j$ be the largest integer such that $\sum_{i=1}^{j} |C_i| \leq \frac{\ell}{4}$. We define the operation $Merge(Y) = (\overline{C}_1, \overline{C}_2)$, where $\overline{C}_1 = \bigcup_{i=1}^{j} C_i$, and $\overline{C}_2 = \bigcup_{i=j+1}^{\ell} C_i$. The following claim is trivially obtained.

**Claim 3.** If $|C_i| < \frac{\delta}{2} \cdot z$ for all $i$, then $\overline{C}_1 \leq \overline{C}_2 \leq \frac{1+\delta}{2} \cdot z$.

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

**Construction of a $(\beta, \gamma)$-balanced rank tree.** In the following, we consider that $T_G = Tree(G) = (V_T, E_T)$ is a tree-decomposition of $G$ and has $|V_T| = b$ bags. Given the parameters $\lambda \in N$ with $\lambda \geq 2$ and $\delta > 0$, we use the following algorithm Rank to construct a tree of bags $R_C$. 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 $|C| \cdot \frac{\delta}{2} \leq 1$, construct a bag $B = \bigcup_{B \in C} B$, and return $B$.
2. Else, if $\ell > 0$ let $(X, Y) = Split(C)$. Construct a bag $B = \bigcup_{B \in x} B_1$, and let $(\overline{C}_1, \overline{C}_2) = Merge(Y)$.
   Call Rank recursively on input $(\overline{C}_1, (\ell+1) \mod \lambda)$ and $(\overline{C}_2, (\ell+1) \mod \lambda)$, 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 $|Nh(C)| > 1$, find a bag $B$ whose removal splits $C$ into connected components $\overline{C}_1, \overline{C}_2$ with $Nh(C) \subseteq Nh(\overline{C}_1) \cup Nh(\overline{C}_2) \leq \frac{|Nh(C)|}{2}$.
   Call Rank recursively on input $(\overline{C}_1, (\ell+1) \mod \lambda)$ and $(\overline{C}_2, (\ell+1) \mod \lambda)$, 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 $|Nh(C)| = 1$, call Rank recursively on input $(\overline{C}, (\ell-1) \mod \lambda)$.

In the following we use the symbols $B$ and $C$ to refer to bags of $T_G$ and $R_C$ 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 $Nh(B) = Nh(C(B))$. Additionally, we denote by $Bh(B)$ the set of separator bags $B_1, \ldots, B_2$ of $C$ that were used to construct $B$. It is straightforward that $Bh(B) \setminus Bh(B_2) = \emptyset$ for every distinct $B_1$ and $B_2$.

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

**Proof.** Every bag in $Nh(C(B'))$ is either a bag in $Nh(C(B'))$, or a separator bag of $C(B')$, and thus a bag of $Bh(B')$. \qed
Note that every bag \( B \) of \( \mathcal{T}_G \) belongs in \( \text{Bh}(B) \) of some bag \( B \) of \( \mathcal{R}_G \) and thus the bags of \( \mathcal{R}_G \) already cover all nodes and edges of \( G \) (i.e., properties C1 and C2 of a tree decomposition). In the following we show how \( \mathcal{R}_G \) can be modified to also satisfy condition C3, i.e., that every node \( u \) appears in a contiguous subtree of \( \mathcal{R}_G \). Given a bag \( B \), we denote by \( \text{NhV}(B) = B \cup \bigcup_{u \in \text{Nh}(B)} P \) (i.e., \( \text{NhV}(B) \) 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 \( T (\text{e.g.} \mathcal{T}_G \text{ or } \mathcal{R}_G) \) and bags \( B_1, B_2 \) of \( T \), we write \( B_1 \sim_{\mathcal{T}_G} B_2 \) 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., \( \mathcal{R}_G \)) if \( u \in B_1 \cap B_2 \) and for the unique simple path \( P^\prime : B_1 \sim_{\mathcal{T}_G} B_2 \) we have \(|P^\prime| \geq 2\) (i.e., there is at least one intermediate bag in \( P^\prime \)) and for all intermediate bags \( B \) in \( P^\prime \) we have \( u \not\in B \). The following crucial lemma shows that if \( B_1 \) and \( B_2 \) form a gap of \( u \) in \( \mathcal{R}_G \), then for every intermediate bag \( B \) in the path \( P : B_1 \sim_{\mathcal{T}_G} B_2 \), \( u \) must appear in some bag of \( \text{NhV}(B) \).

**Lemma 7.** For every node \( u \), and pair of bags \((B_1, B_2)\) that form a gap of \( u \) in \( \mathcal{R}_G \), such that \( B_1 \) is an ancestor of \( B_2 \), for every intermediate bag \( B \) in \( P : B_1 \sim_{\mathcal{T}_G} B_2 \) in \( \mathcal{R}_G \), we have that \( u \in \text{NhV}(B) \).

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

1. Since \( B_2 \) is a descendant of \( B_1 \), we have that \( B_2 \in \mathcal{C}(B) \), i.e., \( B_2 \) is a bag of the component when \( B \) was constructed.
2. By the choice of \( B^\prime \), for every intermediate bag \( B^\prime \) in the path \( B^\prime \sim_{\mathcal{T}_G} B_2 \) we have \( B^\prime \in \mathcal{C}(B) \). Hence \( B^\prime \) is incident to the component where \( B_2 \) belonged at the time \( B \) was constructed.

These two points imply that \( B^\prime \in \text{Nh}(B) \), as desired. Figure 11 suggests a way to turn the rank tree \( \mathcal{R}_G \) to a tree-decomposition. Let \( \mathcal{R}_G = \text{Replace}(\mathcal{R}_G) \) be the tree obtained by replacing each bag \( B \) of \( \mathcal{R}_G \) with \( \text{NhV}(B) \). For a bag \( B \) in \( \mathcal{R}_G \) let \( \mathcal{B} \) be the corresponding bag in \( \mathcal{R}_G \) and vice versa.

**Properties of the tree-decomposition \( \mathcal{R}_G \).** Lemma 8 states that \( \mathcal{R}_G \) obtained by replacing each bag of \( \mathcal{R}_G \) with \( \text{NhV}(B) \) is a tree-decomposition of \( G \). The remaining of the section focuses on
showing that \( \mathcal{R}_G \) is a \((\alpha, \beta, \gamma)\) tree-decomposition of \( G \), and that it can be constructed in \( O(n \cdot \log b) \) time and \( O(b) \) space.

**Lemma 9.** The following assertions hold:

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

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

**Proof.** We prove each item separately.

1. For every bag \( B \) constructed by \text{Rank} in at least \( \gamma - 1 \) out of every \( \gamma \) levels, Item 2 of the algorithm applies, and by Claim 2 the recursion proceeds on components \( \mathcal{C}_1 \) and \( \mathcal{C}_2 \) that are at most \( \frac{1 + \delta}{\beta} \) factor as large as the input component \( C \) in that recursion step. Thus \( B \) is \((\beta, \gamma)\)-balanced in \( \mathcal{R}_G \), and hence \( B \) is \((\beta, \gamma)\)-balanced in \( \mathcal{R}_G \).

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

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 \( G \), then \( \mathcal{R}_G \) is a binary and balanced tree-decomposition of \( G \). 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 treewidth, for any fixed \( \delta > 0 \) and \( 0 < \lambda < 2 \), 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.

**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 one level 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 \(((1 + \delta) / 2) \lambda - 1\). Hence the time complexity is \( O(b \cdot \log b) = (n \cdot \log n) \). 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.** \( \text{ConcurTree}(G) \) is a tree decomposition of \( G \).

**Proof.** We show that \( T \) satisfies the three conditions of a tree decomposition.
The desired result follows from Lemma 11. In view of Remark 2, the time required by Bplex is \( O(n \cdot \beta^k) \). Here we focus on bounding the size of a bag \( |B| \) of treewidth graphs. Let \( \sum_{i=1}^{k} n_i = |B_i| \). Observe that \( Lv(B) = Lv(B_i) \) for all \( i \), and since each \( T_i \) is \((\alpha, \beta, \gamma)\)-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 \( |B| = O(n^{k-1} \cdot \beta^k) \).

D. Details of Section 5

Lemma 1. Consider a graph \( G = (V, E) \) with a weight function \( wt : E \to \Sigma \) and a tree-decomposition \( ConcurTree(G) \). 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 \( A = \{u\} \times (\prod_{1 \leq i,j \leq j} (B_{i-1} \cap B_i)) \times \{v\} \). Then \( d(u, v) = \bigoplus_{(x_1, \ldots,x_j)} d(x_i, x_{i+1}) \).

Proof. By Lemma 1, for every bag \( B_i \) with \( i \geq 1 \) and path \( P : u \sim v \), there exists a node \( x_i \in B_i \cap B_{i-1} \). Denote by \( P_{x,y} \) a path \( x \sim y \) in \( G \). Then \( d(u, v) = \bigoplus_{P_{u,v}} (P_{u,x}) \bigoplus_{P_{x,v}} (P_{x,v}) \).

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

Proof. By construction, for every node \( v \in V \) that strictly refines \( \overline{\pi} \) (i.e., \( v \subset \overline{\pi} \)), we have \( \overline{\pi}(v, v') = d(\overline{\pi}, v) = d(\overline{\pi}, v') \).

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 \( \overline{\pi}', \overline{\pi}'' \) that correspond to a strict partial node \( \pi \) of \( G \). We only focus on \( \overline{\pi}' \), as the other case is similar.

C1 This condition is met, as \( \overline{\pi}' \) appears in every bag of \( T \) that contains a node \( u \) that refines \( \overline{\pi}' \).

C2 Since every node \( \overline{\pi}' \) is connected only to nodes \( u \) of \( G \) that refine \( \overline{\pi}' \), this condition is also met.

Figure 12: Given the tree-decomposition \( \text{Tree}(G) \) on the left, the graph in the middle is the corresponding \( R_G \) and the one on the right is the corresponding tree-decomposition \( \hat{R}_G \) = Replace\( (R_G) \) after replacing each bag \( B \) with \( \text{NhV}(B) \).
If \( \pi^1 = \langle \bot, \ldots, \bot \rangle \), then \( \pi^1 \) 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 \( \pi^1 \), we have \( u_i \neq \bot \) (i.e., \( u_i \in V_{\pi} \)), 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 \( u \in V_{\pi}(B_i) \). Additionally, if for some bag \( B'_i \) of Tree(\( G_i \)) we have \( u \notin V_{\pi}(B'_i) \), then \( u \notin V_{\pi}(B'_i') \) 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 \( \pi^1 \) appears in a contiguous subtree of ConcurTree(\( G \)). It follows that \( \pi^1 \) appears in a contiguous subtree of \( T \), as desired.

The desired result follows.

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

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

(5)

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

(6)

Then

1. \( T_{\pi}(n) = O(n^{3(k-1)}) \), and
2. (i) \( S_{\pi}(n) = O(n^{2(k-1)}) \) if \( k \geq 3 \), and (ii) \( S_{\pi}(n) = O(n^{2+\epsilon}) \).

**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

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

(13)

and since \( \log(1 + \delta) = \frac{\log(1 + \epsilon)}{\ln 2} \), we have

\[ (1 + \delta)^{3(\lambda-1)(k-1)} \leq 2^{\ln(1 + \epsilon)(3(\lambda-1)(k-1))} \leq 2^{\delta^3(\lambda-1)(k-1)} \]

(16)

Hence the expression in Eq. (13) is bounded by \( 2^{\delta^3} \) with

\[ x \leq 2 \cdot \delta \cdot (2 + \epsilon) \cdot (\lambda - 1) + 2 - \epsilon \cdot (\lambda - 1) = (\lambda - 1) \cdot (2 \cdot \delta \cdot (2 + \epsilon) - \epsilon) + 2 \]

\[ \leq (\lambda - 1) \cdot \frac{4 + 2 + 2 \cdot 2^3 - 18 \cdot \epsilon}{18} + 2 \]

\[ \leq (1 - \delta) \cdot \frac{2}{3} + 2 \]

\[ \leq -4 \cdot \frac{2}{3} + 2 \leq 0 \]

which yields that there exists a constant \( c < 1 \) for which

\[ 2^{\delta^3} \cdot S_{\pi}(n) \leq c \cdot n^{2+\epsilon} \]

Similarly as above, we have

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

(15)

and

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

(16)

Hence the expression in Eq. (16) is bounded by \( 2^{\delta^3} \) with

\[ x \leq 2 \cdot \delta \cdot (2 + \epsilon) \cdot (\lambda - 1) + 2 - \epsilon \cdot (\lambda - 1) = (\lambda - 1) \cdot (2 \cdot \delta \cdot (2 + \epsilon) - \epsilon) + 2 \]

\[ \leq (\lambda - 1) \cdot \frac{4 + 2 + 2 \cdot 2^3 - 18 \cdot \epsilon}{18} + 2 \]

\[ \leq (1 - \delta) \cdot \frac{2}{3} + 2 \]

\[ \leq -4 \cdot \frac{2}{3} + 2 \leq 0 \]

which yields that there exists a constant \( c < 1 \) for which

\[ 2^{\delta^3} \cdot S_{\pi}(n) \leq c \cdot n^{2+\epsilon} \]

which yields that Eq. (14) follows a geometric series, and thus \( S_{\pi}(n) = O(n^{2+\epsilon}) \).

**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 \( P : x_i \rightarrow x_j \) with \( wt(P) = z \) in \( G \) iff there exists a path \( P' : \langle x_i, x_j \rangle \rightarrow \langle x_i, x_j \rangle \) with \( wt(P') = z \) in \( G' \).

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

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

\[ \langle x_i, x_j \rangle \rightarrow \langle x_i, y_i \rangle \rightarrow \langle y_i, x_j \rangle \rightarrow \langle x_i, y_j \rangle \rightarrow \langle y_j, x_j \rangle \rightarrow \langle x_j, x_j \rangle \]

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

2. Every path \( P \) that traverses a red edge \( \langle x_i, y_j \rangle \) has to traverse a blue edge to \( \langle x_j, x_j \rangle \). Then \( x_j \) must be \( x_j \).

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 \( |C| \cdot \frac{\lambda}{2} \leq 1 \) then
3. Assign \( B \leftarrow \bigcup_{B \in C} B \) and make \( B \) the root of \( T \)
4. else if \( \ell > 0 \) then
5. Assign \((X, Y) \leftarrow \) Split\((C)\)
6. Assign \( B \leftarrow \bigcup_{B \in Y} B_i \)
7. Assign \((\overline{C}_1, \overline{C}_2) \leftarrow \) Merge\((Y)\)
8. Assign \( T_1 \leftarrow \text{Rank}\((\overline{C}_1, (\ell + 1) \mod \lambda)\)\)
9. Assign \( T_2 \leftarrow \text{Rank}\((\overline{C}_2, (\ell + 1) \mod \lambda)\)\)
10. Make \( B \) the root of \( T \) and \( T_1 \) and \( T_2 \) its left and right subtree
11. else if \( |\text{Nh}(C)| > 1 \) then
12. Let \( B \leftarrow \) a bag of \( C \) whose removal splits \( C \) to \( \overline{C}_1, \overline{C}_2 \) with \(|\text{Nh}(\overline{C}_1) \cap \text{Nh}(C)| \leq \frac{|\text{Nh}(C)|}{2}\)
13. Assign \( B \leftarrow B \)
14. Assign \( T_1 \leftarrow \text{Rank}\((\overline{C}_1, (\ell + 1) \mod \lambda)\)\)
15. Assign \( T_2 \leftarrow \text{Rank}\((\overline{C}_2, (\ell + 1) \mod \lambda)\)\)
16. Make \( B \) the root of \( T \) and \( T_1 \) and \( T_2 \) its left and right subtree
17. else
18. Assign \( T \leftarrow \text{Rank}\((C, (\ell - 1) \mod \lambda)\)\)
19. return \( T \)

Algorithm 2: ConcurTree

**Input:** Tree-decompositions \( T_i = (V_{T_i}, E_{T_i}) \) of \( G \) with root bags \((B_j)_{1 \leq j \leq k}\)

**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 \( B \leftarrow B \cup \left(\bigcap_{1 \leq j \leq i} V_{T_j}(B_j) \times B_i \times \bigcap_{1 < j \leq k} V_{T_j}(B_j)\right)\)
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 \( T'_i \leftarrow \text{ConcurTree}(T_i(B'_1), \ldots, T_k(B'_k))\)
9. Add \( T'_i \) to \( T \), setting the root of \( T'_i \) as a new child of \( B \)
10. end
11. end
12. return \( T \)

Algorithm 3: ConcurPreprocess 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 \( \overline{V} \leftarrow V \)
2. Assign \( \overline{E} \leftarrow E \)
3. Create a map \( \overline{\text{wt}} : \overline{E} \rightarrow \Sigma \)
4. Assign \( \overline{\text{wt}} \leftarrow \text{wt} \)
5. foreach \( u' \in \bigcup_i \{V_i \cup \{\bot\}\} \) do
6. Let \( u \in V \) such that \( u \supset u' \)
7. Assign \( \overline{V} \leftarrow \overline{V} \cup \{\overline{\pi}, \overline{\pi'}\} \)
8. Assign \( \overline{E} \leftarrow \overline{E} \cup \{(\overline{\pi}, u), (u, \pi')\} \)
9. Set \( \overline{\text{wt}}(\overline{\pi}, u) \leftarrow \overline{\pi} \)
10. Set \( \overline{\text{wt}}(u, \pi') \leftarrow \overline{\pi} \)
11. end
12. return \( \overline{G} = (\overline{V}, \overline{E}) \) and \( \overline{\text{wt}} \)

Algorithm 4: ConcurPreprocess 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 \( \overline{T} \) of \( \overline{G} \) */
1. Assign \( \overline{\pi_T} \leftarrow \emptyset \)
2. foreach \( B \in V_T \) do
3. Assign \( \overline{B} \leftarrow B \)
4. foreach \( u \in B \) do
5. Assign \( \overline{B} \leftarrow \overline{B} \cup \{\overline{B}', \overline{B}''\} \)
6. end
7. end
8. Assign \( \overline{\pi_T} \leftarrow \overline{\pi_T} \cup \{\overline{B}\} \)
9. end
10. return \( \overline{T} = (\overline{\pi_T}, E_T) \)
### Algorithm 5: ConcurPreprocess Item 1

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

/* Local distance computation */

1. foreach partial node $\pi$ do
2.   Create two maps $\text{FWD}_T$, $\text{BWD}_T : \mathcal{B}_T \rightarrow P$ for $\pi \in \mathcal{B}_T$
3.   Assign $\text{FWD}_T(\pi) \leftarrow \text{wt}(\pi)$
4.   Assign $\text{BWD}_T(\pi) \leftarrow \text{wt}(\pi)$
5. end
6. end

### Algorithm 6: ConcurPreprocess Item 2

**Input:** The partial expansion tree-decomposition $T = (V_T, E_T)$ and maps $\text{FWD}_u$, $\text{BWD}_u : \mathcal{B}_u \rightarrow P$ for every partial node $\pi$

/* Ancestor distance computation */

1. foreach node $u \in V$ do
2.   Create two maps $\text{FWD}_u$, $\text{BWD}_u : \mathcal{B}_u \rightarrow P$
3. end
4. foreach bag $B$ of $T$ in DFS order starting from the root do
5.   Assign $d' \leftarrow$ the transitive closure of $\mathcal{C}(B)$ with wrt $wt_T$
6.   foreach node $\pi \in B$ do
7.     if Lv(\pi) \leq Lv(\pi) then
8.       Assign $\text{BWD}_T(\pi) \leftarrow d'(\pi, \pi)$
9.       Assign $\text{FWD}_T(\pi) \leftarrow d'(\pi, \pi)$
10. end
11. end
12. end

### Algorithm 7: ConcurQuery Single-source query

**Input:** A source node $u \in V$

**Output:** A map $A : V \rightarrow P$ that contains distances of vertices from $u$

1. Create a map $A : V \rightarrow P$
2. for $v \in V$ do
3.   Assign $A(v) \leftarrow 0$
4. end
5. for every bag $B$ of $T$ in BFS order starting from $B_u$ do
6.   for $x, v \in B \cap V$ do
7.     if Lv(v) \leq Lv(x) then
8.       Assign $A(x) \leftarrow A(v) \oplus A(x) \oplus \text{FWD}_x(v)$
9.     end
10. end
11. end
12. 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 $d' \leftarrow 0$
3. for $x \in B \cap V$ do
4.   Assign $d' \leftarrow d' \oplus \text{FWD}_x^+(x) \oplus \text{BWD}_x^+(x)$
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 $\text{FWD}_x^+(\pi''')$
3. else if $\pi$ is strictly partial then
4.   return $\text{BWD}_x^+(\tau'')$
5. else
6.   return $\text{FWD}_x^+(\pi'')$
7. end