Abstract

A linear graph is a graph whose vertices are linearly ordered. This linear ordering allows pairs of disjoint edges to be either preceding (<), nesting (□) or crossing (≀). Given a family of linear graphs, and a non-empty subset $R \subseteq \{<, \sqsubseteq, \bowtie\}$, we are interested in the Maximum Common Structured Pattern (MCSP) problem: find a maximum size edge-disjoint graph, with edge-pairs all comparable by one of the relations in $R$, that occurs as a subgraph in each of the linear graphs of the family. The MCSP problem generalizes many structure-comparison and structure-prediction problems that arise in computational molecular biology.

We give tight hardness results for the MCSP problem for $\{<, \bowtie\}$-structured patterns and $\{\sqsubseteq, \bowtie\}$-structured patterns. Furthermore, we prove that the problem is approximable within ratios: (i) $2\mathcal{H}(k)$ for $\{<, \bowtie\}$-structured patterns, (ii) $k^{1/2}$ for $\{\sqsubseteq, \bowtie\}$-structured patterns, and (iii) $O(\sqrt{k\log k})$ for $\{<, \sqsubseteq, \bowtie\}$-structured patterns, where $k$ is the size of the optimal solution and $\mathcal{H}(k) = \sum_{i=1}^{k} 1/i$ is the $k$-th harmonic number. Also, we provide combinatorial results concerning the different types of structured patterns that are of independent interest in their own right.
1 Introduction

Many biological molecules such as RNA and proteins exhibit a three-dimensional structure that determines most of their functionality. This three dimensional structure can be modeled in two dimensions by an edge-disjoint linear graph, \textit{i.e.}, a graph with linearly ordered vertices that are incident to exactly one edge. The corresponding structure-similarity or structure-prediction problems that arise in such contexts usually translate to finding common edge-disjoint subgraphs, or common \textit{structured patterns}, that occur in a family of general linear graphs. Examples of such problems are the \textsc{Longest Common Subsequence} \cite{21,22} problem, the \textsc{Maximum Common Ordered Tree Inclusion} \cite{2,8,23} problem, the \textsc{Arc-Preserving Subsequence} \cite{4,16,19} problem, and the \textsc{Maximum Contact Map Overlap} \cite{17} problem. In this paper, we study a general framework for such problems which we call the \textsc{Maximum Common Structured Pattern} (MCSP) problem.

The MCSP problem was originally introduced (under a different name) by Davydov and Batzoglou \cite{10} in the context of (non-coding) RNA secondary structure prediction via multiple structural alignment. There, an RNA sequence of \( n \) nucleotides is represented by a linear graph with \( n \) vertices, and an edge connects two vertices if and only if their corresponding nucleotides are complementary. A family of linear graphs is then used to represent a family of functionally-related RNAs, and a common structured pattern in such a family is considered to be a putative common secondary structure element of the family.

The ordering amongst the vertices of a linear graph allows a pair of disjoint edges in the graph to be either preceding (\(<\)), nesting (\(\sqsubseteq\)), or crossing (\(\bowtie\)). Since most RNA secondary structures translate to linear graphs with non-crossing edges, Davydov and Batzoglou \cite{10} focused on the variant of the MCSP problem where the common structured pattern is required to be non-crossing. In other words, they focus on finding maximum common \(\{<,\sqsubseteq\}\)-structured patterns. However, there are known RNAs which have secondary structures that translate to linear graphs with a few edge-crossings (pseudo-
knotted RNA secondary structures). Also, when predicting proteins rather than RNA structures, the non-crossing restriction becomes an even bigger limitation since the folding structures of proteins are often more complex than those of RNAs. In [18], it is argued that many important protein secondary structure elements like alpha helices and anti-parallel beta sheets exhibit \(\{<, \dot{\square}\}\)-structured patterns, i.e., patterns which are non-nesting rather than non-crossing.

In the following, we present a framework which extends the work of Davydov and Batzoglou [10] by considering different types of common structured patterns. Following [33], we consider structured patterns that are allowed to have crossing edges, and which might also be restricted to be non-nesting or non-preceding. More specifically, the MCSP problem receives as input a family of linear graphs and a non-empty subset \(\mathcal{R} \subseteq \{<, \sqsubset, \dot{\square}\}\), and the goal is to find a maximum common \(\mathcal{R}\)-structured pattern (see Figure 1). We study the combinatorics behind the structures of these different types of patterns, with a focus on approximation algorithms for the MCSP problem.

![Figure 1](image)

Figure 1. Four linear graphs \(G_1, G_2, G_3\) and \(G_4\) and a \(\{<, \sqsubset, \dot{\square}\}\)-comparable common structured pattern (depicted also as \(G_{sol}\) in the bottom part). The occurrence of the structured pattern \(G_{sol}\) in each graph is emphasized in bold. Edges \(e_2, e_3, e_4\) and \(e_5\) are nested in edge \(e_1\); edges \(e_2\) and \(e_3\) precede edge \(e_5\); edge \(e_2\) precedes edge \(e_4\) and crosses edge \(e_3\), while edge \(e_3\) crosses both edges \(e_2\) and \(e_4\).

The paper is organized as follows. In the remaining part of this section we briefly review related work and notations that will be used throughout the paper. In Section 2, we discuss simple structured patterns, i.e., \(R\)-structured patterns, where \(R \in \{<, \sqsubset, \dot{\square}\}\), and \(\{<, \sqsubset\}\)-structured patterns. Following this, we discuss the more complex \(\{<, \dot{\square}\}\)-structured patterns and \(\{\sqsubset, \dot{\square}\}\)-structured patterns in Section 3 and Section 4, respectively. In Section 5, we deal with general structured patterns, i.e., \(\{<, \sqsubset, \dot{\square}\}\)-structured patterns. An overview of the paper, along with some open problems, is given in Section 6.
1.1 Related work

There are many structural comparison problems that are closely related to the MCSP problem. First, as mentioned previously, the MCSP problem for \(<, \sqsubseteq\)-structured patterns has been studied by Davydov and Batzoglou in [10] (coined as the Maximum Common Nested Subgraph problem). Recently, new results concerning this problem appeared in [27]. We discuss the results of both these works in Section 2. Below we list other related problems.

Closely related to the MCSP problem are the Arc-Preserving Subsequence [4,16,19], and Maximum Contact Map Overlap [17] problems. Both are concerned with finding maximum common subgraphs in a pair of linear graphs, except that in the Arc-Preserving Subsequence problem the vertices of the linear graphs are assigned letters from some given alphabet, and an occurrence of a common subgraph in each of the linear graphs is required to preserve the letters, as well as their arc structure. Another closely related problem is the Pattern Matching over 2-Interval Set problem [33], where one asks whether a given structured pattern occurs in a given 2-interval set, which is a generalization of a linear graph. The 2-Interval Pattern problem [5,9,33] asks to find the maximum \(R\)-structured pattern, for some given \(R \subseteq \{<, \sqsubseteq, \bowtie\}\), in a single family of 2-interval sets.

There is a well-known bijective correspondence between \(<, \sqsubseteq\)-structured patterns and ordered forests – the nesting relation corresponds to the ancestor/predecessor relationship between the nodes, and the precedence relation corresponds to their order. Hence, the MCSP problem for \(<, \sqsubseteq\)-structured patterns can be viewed as the problem of finding a tree which is included in all trees of a given tree family, the Maximum Common Ordered Tree Inclusion problem. Determining whether a tree is included in another is studied in [2,8,23]. Finding the maximum common tree included in a pair of trees can be done using the algorithms given in [11,24,31]. The MCSP problem for \(<, \sqsubseteq\)-structured patterns has been studied in [10,27]. We discuss the results there in Section 2.

Like \(<, \sqsubseteq\)-structured patterns, \(\sqsubseteq, \bowtie\)-structured patterns also correspond to natural combinatorial objects, namely permutations (see Section 4). In [7], the authors studied the problem of determining whether a permutation occurs in another permutation as a pattern, the so called Pattern Matching for Permutations problem. This problem corresponds to determining whether a \(\sqsubseteq, \bowtie\)-structured pattern is a subpattern of another \(\sqsubseteq, \bowtie\)-structured pattern. Bose, Buss, and Lubiw proved that Pattern Matching for Permutations is \(\text{NP}\)-complete [7].

Determining whether a given \(<, \bowtie\)-structured pattern occurs in a general
linear graph has been studied in [18,28]. Gramm [18] gave a polynomial-time algorithm for this problem. Recently, Li and Li [28] proved that this algorithm was incorrect and showed the problem was in fact \textbf{NP}-complete. Prior to this, Blin \textit{et al.} [5] proved that a generalization of this problem, where the linear graph is replaced by a 2-interval set, is \textbf{NP}-complete.

Finally, probably the oldest and most famous problem related to the \textit{MCSP} problem is the \textbf{Longest Common Subsequence (LCS)} problem [21,22], where one wishes to find the longest common subsequence in two or more sequences. Important developments of the initial algorithms of [21,22] can be found in [3,14,30]. Maier [29] proved that the \textit{LCS} problem for multiple sequences is \textbf{NP}-hard.

\subsection{Terminology and basic definitions}

For a graph $G$, we denote $V(G)$ as the set of vertices and $E(G)$ as the set of edges. The \textit{order} and the \textit{size} of $G$ stand for $|V(G)|$ and $|E(G)|$, respectively. A \textit{linear graph} of order $n$ is a vertex-labeled graph where each vertex is labeled by a distinct label from $\{1, 2, \ldots, n\}$. Thus, it can be viewed as a graph with vertices embedded on the integral line, yielding a total order amongst them. In case of linear graphs, we write an edge between vertices $i$ and $j$, $i < j$, as the pair $(i,j)$. Two edges of a linear graph are \textit{disjoint} if they do not share a common vertex. A linear graph $G$ is said to be \textit{edge-disjoint} if it is composed of disjoint edges, \textit{i.e.}, $G$ is a matching. Of particular interest are the relations between pairs of disjoint edges [33]. Let $e = (i,j)$ and $e' = (i',j')$ be two disjoint edges in a linear graph $G$; we write:

- $e < e'$ (\textit{e precedes} $e'$) if $i < j < i' < j'$,
- $e \sqsubset e'$ (\textit{e is nested in} $e'$) if $i' < i < j < j'$, and
- $e \not\sqsubseteq e'$ (\textit{e and} $e'$ \textit{cross}) if $i < i' < j < j'$.

Two edges $e$ and $e'$ are $R$-\textit{comparable}, for some $R \subseteq \{<, \sqsubset, \not\sqsubseteq\}$, if $e R e'$ or $e' R e$. For a subset $\mathcal{R} \subseteq \{<, \sqsubset, \not\sqsubseteq\}$, $\mathcal{R} \neq \emptyset$, $e$ and $e'$ are said to be $\mathcal{R}$-\textit{comparable} if $e$ and $e'$ are $R$-comparable for some $R \in \mathcal{R}$. A set of edges $E$ (or a linear graph $G$ with $E(G) = E$) is $\mathcal{R}$-\textit{comparable} if any pair of distinct edges $e, e' \in E$ are $\mathcal{R}$-comparable. A \textit{subgraph} of a linear graph $G$ is a linear graph $H$ which can be obtained from $G$ by a series of vertex and edge deletions, where a deletion of vertex $i$ results in removing vertex $i$ and all edges incident to it from the graph, and then relabeling all vertices $j$ with $j > i$ to $j - 1$. An edge-disjoint subgraph of a linear graph is called a \textit{structured pattern}. For a vertex subset $V \subseteq V(G)$ (resp. edge subset $E \subseteq E(G)$), we let $G[V]$ (resp. $G[E]$) denote the subgraph \textit{induced} by $V$ (resp. $E$), \textit{i.e.} the subgraph obtained by deleting all vertices $V(G) \setminus V$ (resp. all edges $E(G) \setminus E$). By convention, we let $G[i \ldots j]$,
\[1 \leq i \leq j \leq |V(G)|,\] denote the subgraph induced by all vertices labeled \(k\) with \(i \leq k \leq j\). For a family of linear graphs \(G = G_1, \ldots, G_n\), a common structured pattern of \(G\) is an edge-disjoint linear graph \(H\) that is a subgraph of \(G_i\), for all \(1 \leq i \leq n\). Following the above notation, \(H\) is called an \(R\)-structured pattern, for some non-empty \(R \subseteq \{<, \sqsubseteq, \sqsupseteq\}\), if \(E(H)\) is \(R\)-comparable. We are now in position to formally define the MCSP problem.

**Definition 1** Given a family of linear graphs \(G = G_1, \ldots, G_n\) and a subset \(R \subseteq \{<, \sqsubseteq, \sqsupseteq\}\), \(R \neq \emptyset\), the Maximum Common Structured Pattern (MCSP) problem asks to find a maximum-size common \(R\)-structured pattern of \(G\).

We will use the following terminology to describe special edge-disjoint linear graphs. A linear graph is called a sequence if it is \(\{<\}\)-comparable, a tower if it is \(\{\sqsubseteq\}\)-comparable, and a staircase if it is \(\{\sqsupseteq\}\)-comparable. We define the width (resp. height and depth) of a linear graph to be the size of the maximum cardinality sequence (resp. tower and staircase) subgraph of the graph. A \(\{<, \sqsubseteq\}\)-comparable linear graph with the additional property that any two maximal towers in it do not share an edge is called a sequence of towers. Similarly, a \(\{<, \sqsupseteq\}\)-comparable linear graph is a sequence of staircases if any two maximal staircases do not share an edge. A tower of staircases is a \(\{\sqsubseteq, \sqsupseteq\}\)-comparable linear graph where any pair of maximal staircases do not share an edge, and a staircase of towers is a \(\{\sqsupseteq, \sqsubseteq\}\)-comparable linear graph where any pair of maximal towers do not share an edge. A sequence of towers (resp. sequence of staircases, tower of staircases, and staircase of towers) is balanced if all of its maximal towers (resp. staircases, staircases, and towers) are of equal size. Figure 2 illustrates an example of the above types of linear graphs.

### 1.3 Our results

The results presented in this paper are as follows: We give tight hardness results for the MCSP problem for \(\{<, \sqsupseteq\}\)-structured patterns and \(\{\sqsubseteq, \sqsupseteq\}\)-structured patterns, extending the hardness result given in [27] for \(\{<, \sqsubseteq\}\)-structured patterns. Furthermore, we provide approximation algorithms that prove that the problem is approximable within ratios:

- \(2H(k)\) for \(\{<, \sqsupseteq\}\)-structured patterns,
- \(k^{1/2}\) for \(\{\sqsubseteq, \sqsupseteq\}\)-structured patterns, and
- \(O(\sqrt{k \log k})\) for \(\{<, \sqsubseteq, \sqsupseteq\}\)-structured patterns,

where \(k\) denotes the size of an optimal solution, and \(H(k) = \sum_{i=1}^{k} 1/i\) is the \(k\)-th harmonic number. The respective running times of these algorithms are \(O(nm^{2.5}\log^2 m)\), \(O(nm^{1.5})\), and \(O(nm^{2.5}\log^2 m)\), where \(n\) denotes the number
of input linear graphs, and $m$ denotes the maximum number of edges in any graph of the input. Along the way, we provide some combinatorial results concerning the different types of structured patterns that are of independent interest in their own right.

2 Simple and $\{<,\sqsubset\}$-Structured Patterns

We begin our study by considering the MCSP problem for simple structured patterns ($R$-structured pattern for a single relation $R \in \{<,\sqsubset,\sqsupset\}$), and for $\{<,\sqsubset\}$-structured patterns. We first discuss the analogy between the relations we defined for disjoint edges in a linear graph and well-studied relations defined for families of intervals. We show that known algorithms on interval families can be used to solve the MCSP problem for simple structured patterns in polynomial-time. Following this, we discuss results presented in [10,27] for the
MCSP problem for \{<, \sqsupseteq\}\-structured patterns. Most of the results in this section are known, and are presented here for sake of completeness.

For a given linear graph \(G\) of size \(m\), let \(\mathcal{I}(G) = \{[i,j] \mid (i,j) \in E(G)\}\) be the family of intervals obtained by considering each edge of \(G\) as an interval of the line, closed between both its endpoints. A pair of \(<\)-comparable edges in \(E(G)\) correspond to a pair of disjoint intervals in \(\mathcal{I}(G)\), a pair of \(\sqsupseteq\)-comparable edges correspond to a pair of nesting intervals, and a pair of \(\sqsubsetneq\)-comparable edges correspond to a pair of overlapping intervals. Note that this correspondence is bi-directional only if \(G\) is edge-disjoint, since a pair of edges sharing a vertex can correspond to a pair of nesting or overlapping intervals. Nevertheless, we can always modify \(\mathcal{I}(G)\) in such a way that all intervals have unique endpoints, and so that any pair of intervals who were sharing an endpoint now become non-nesting (resp. non-overlapping). A maximum pairwise disjoint subset of intervals can be computed in linear time using standard dynamic-programming, assuming the interval family is given in a sorted manner [20] (which we can provide in linear time in our case using bucket sorting). A maximum pairwise nesting subset can be computed in \(O(m \log \log m)\) in an interval family of \(m\) intervals (see for example the algorithm in [34]), and a maximum pairwise overlapping subset in \(O(m^{1.5})\) time [32].

**Lemma 2 ([20], [34] and [32]):** Let \(G\) be a linear graph of size \(m\). Then there exists

1. an \(O(m)\) algorithm for finding the largest \(<\>-comparable subgraph of \(G\),
2. an \(O(m \log \log m)\) algorithm for finding the largest \(\sqsupseteq\)-comparable subgraph of \(G\), and
3. an \(O(m^{1.5})\) algorithm for finding the largest \(\sqsubsetneq\)-comparable subgraph of \(G\).

**Corollary 3** The MCSP problem for \(<\>-structured patterns (resp. \(\sqsupseteq\)-structured patterns and \(\sqsubsetneq\)-structured patterns) is solvable in \(O(nm)\) (resp. \(O(nm \log \log m)\) and \(O(nm^{1.5})\)) time, where \(n = |G|\) and \(m = \max_{G \in \mathcal{G}} |E(G)|\).

We next consider \(<, \sqsupseteq\>-structured patterns. The MCSP problem for this type of patterns was considered in [10,27], in the context of multiple RNA structural alignment.

**Theorem 4 ([27]):** The MCSP problem for \(<, \sqsupseteq\>-structured patterns is NP-hard even if each input linear graph is a sequence of towers of height at most 2.

Note, however, that the MCSP problem is polynomial-time solvable in case the number of input linear graphs is a constant [27]. The MCSP problem for \(<, \sqsupseteq\>-structured patterns was proved to be approximable with ratio \(O(\log^2 k)\) [10], where \(k\) is the size of the optimal solution. The approxima-
tion ratio was later improved to $\log k + 1$ [27].

**Theorem 5** ([27]) The MCSP problem for $\{<, \sqsubseteq\}$-structured patterns is approximable within ratio $O(\log k)$ in $O(nm^2)$ time, where $k$ is the size of an optimal solution, $n = |G|$, and $m$ is the maximum size of any graph in $G$.

### 3 $\{<, \sqsupseteq\}$-Structured Patterns

We now turn to considering the MCSP problem for $\{<, \sqsupseteq\}$-structured patterns. We begin by proving a tight hardness result for the problem. Following this, we present an approximation algorithm for the problem which achieves an approximation ratio of $2H(k)$ in $O(nm^2 \cdot 5\log_2 m)$ time, where $k$ is the size of an optimal solution, $H(k) = \sum_{i=1}^{k} 1/i$, $n = |G|$, and $m$ is the maximum size of any graph in $G$.

**Theorem 6** The MCSP problem for $\{<, \sqsupseteq\}$-structured patterns is NP-hard even if each input linear graph is a sequence of staircases of depth at most 2.

**PROOF.** We reduce from the MCSP problem for $\{<, \sqsubseteq\}$-structured patterns, which is NP-hard even if each input graph is a sequence of towers of height at most 2 (Theorem 4). Let an arbitrary instance of MCSP for $\{<, \sqsubseteq\}$-structured patterns be given by a family $G = \{G_1, G_2, \ldots, G_n\}$ of $n$ sequences of towers with height at most 2. The corresponding instance of the MCSP problem for $\{<, \sqsupseteq\}$-structured patterns consists of a family $G' = \{G'_1, G'_2, \ldots, G'_n\}$ of $n$ sequences of staircases of depth at most 2, where for each $i$, $1 \leq i \leq n$, the graph $G'_i$ is obtained from $G_i$ by transforming each pair of $\sqsubseteq$-comparable edges into a pair of $\sqsupseteq$-comparable edges. Below is an example of such a transformation.

![Example Transformation](image)

Clearly, our construction can be carried out in polynomial-time. Furthermore, there exists a common $\{<, \sqsubseteq\}$-structured pattern of size $k$ in $G$ if and only if there exists a common $\{<, \sqsupseteq\}$-structured pattern of size $k$ in $G'$. □

A recent result [28] implies that the MCSP problem for $\{<, \sqsupseteq\}$-structured patterns is hard even if $G$ consists of only two linear graphs. However, the
input linear graphs used in [28] are of unlimited structure, unlike in the lemma above. For the case where \(|G| = 1\), the problem is still open [5,9,33].

We now show that the MCSP problem for \(<, \emptyset\>\)-structured is approximable within ratio \(2\mathcal{H}(k)\). The first ingredient of our proof is to observe that any \(<, \emptyset\>\)-structured pattern contains a sequence of staircases of substantial size.

**Lemma 7** Let \(G\) be a \(<, \emptyset\>\)-comparable linear graph. There exists a partition \(E(G) = E_{\text{RED}} \cup E_{\text{BLUE}}\) such that both \(G[E_{\text{RED}}]\) and \(G[E_{\text{BLUE}}]\) are sequences of staircases.

**Proof.** The proof is constructive and uses the following edge-coloring algorithm: first, sort the edges of \(G\) by ascending left vertex. Let \(E(G) = (e_1, e_2, \ldots, e_m)\) be the edges of \(G\) according to that order. Next, start coloring the edges one at a time, proceeding from \(e_1\) to \(e_m\), in the following manner: start by initializing \(e_1\) to be the current leading edge, and \(\text{RED}\) to be the current color. Color the current leading edge with the current color, and proceed to color all following edges with the current color, as long as they are crossing with the current leading edge. As soon as an edge cannot be colored with the current color (this edge and the current leading edge are not crossing), change the current color (\(\text{RED}\) to \(\text{BLUE}\) or \(\text{BLUE}\) to \(\text{RED}\)), set this edge as the current leading edge, and continue in a similar fashion.

Let us say that two edges \(e_i\) and \(e_j\) are colored during the same run if there does not exist a leading edge \(e_\ell\) with \(i < \ell < j\). The following facts are immediate consequences of the algorithm: (i) each edge is colored either \(\text{RED}\) or \(\text{BLUE}\), (ii) any two edges in the same run are colored by the same color and are \(<, \emptyset\>\)-comparable, and (iii) any two edges colored in different runs with the same color are \(<, \emptyset\>\)-comparable. Let \(E_{\text{RED}}\) and \(E_{\text{BLUE}}\) be the set of edges colored \(\text{RED}\) and \(\text{BLUE}\) respectively. By the above, it follows that \(E_{\text{RED}} \cup E_{\text{BLUE}}\) is a partition of \(E(G)\), and \(G[E_{\text{RED}}]\) and \(G[E_{\text{BLUE}}]\) are both sequences of staircases. \(\square\)

The second ingredient of our approach consists in showing that any sequence of staircases contains a balanced subgraph of substantial size.

**Lemma 8** Let \(G\) be a sequence of staircases of size \(k\). Then \(G\) contains a balanced sequence of staircases with at least \(\frac{k}{\mathcal{H}(k)}\) edges.

**Proof.** Let \(G_1, G_2, \ldots, G_p\) be the (maximal) staircases of \(G\). For each \(1 \leq i \leq k\), let \(n_i\) denote the number of staircases of \(G\) of size at least \(i\). Then \(k = \sum_{i=1}^{k} n_i\). Let \(G'\) be a largest balanced subgraph of \(G\), and set \(k' = |E(G')|\). By maximality of \(k'\), we have \(i \cdot n_i \leq k'\) for each \(1 \leq i \leq k\). Therefore, \(k = \sum_{i=1}^{k} n_i \leq \sum_{i=1}^{k} \frac{k'}{i} = k' \sum_{i=1}^{k} \frac{1}{i} = k' \mathcal{H}(k)\), and the lemma is proved. \(\square\)
Bal-Seq-Staircase($G, d, w$)

Input: A linear graph $G$ of size $m$, and two positive integers $d$ and $w$.

Result: true if and only if $G$ contains a balanced sequence of staircases of width $w$ and depth $d$.

(1) $E' \leftarrow \emptyset$.
(2) for $i = 1, 2, \ldots, m - 1$ do
   (a) Let $j$ be the smallest integer such that $G[i \ldots j]$ contains as a subgraph a staircase of size $d$ (set $j = \infty$ if no such integer exists).
   (b) if $j \neq \infty$ then $E' \leftarrow E' \cup \{(i, j)\}$.
(3) Compute $H$, the maximum $\{<\}$-comparable subgraph of the linear graph $G' = (V(G), E')$.
(4) if $|E(H)| \geq w$ then return true else return false.

Figure 3. Algorithm Bal-Seq-Staircase for finding a balanced sequence of staircases of width $w$ and depth $d$ in a linear graph $G$. For a linear graph $G \in \mathcal{G}$, and two integers $i$ and $j$ with $1 \leq i < j \leq |V(G)|$, we use $G[i, \ldots, j]$ to denote the subgraph of $G$ obtained by deleting all vertices labeled $k$ with $k < i$ or $j < k$.

Note that $\mathcal{H}(k)$ is bounded by $\ln k + O(1)$. Combining Lemmas 7 and 8, we obtain the following.

Corollary 9 Any $\{<, \not\leq\}$-comparable linear graph of size $k$ contains as a subgraph a balanced sequence of staircases of size at least $\frac{k}{2 \mathcal{H}(k)}$.

What is left is to show that, given a set of linear graphs, one can find in polynomial-time the size of the largest balanced sequence of staircases that occurs in each input linear graph. Devoted to this particular purpose is Algorithm Bal-Seq-Staircase (Figure 3).

Lemma 10 Algorithm Bal-Seq-Staircase($G, w, d$) runs in $O(m^{2.5} \log m)$ time and returns true if and only if $G$ contains a balanced sequence of staircases of width $w$ and depth $d$.

PROOF. Correctness of Algorithm Bal-Seq-Staircase is easily verifiable. We therefore concentrate on evaluating its time complexity. Steps 2 and 3 are the only non-trivial steps of the algorithm. Checking whether subgraph $G[i \ldots j]$ contains a staircase of size $d$ can be done in $O(m^{1.5})$ time (Lemma 2). Hence, using a simple binary search strategy, one can find for any $i$ the smallest $j$ such that $G[i \ldots j]$ contains a staircase of size $d$ in $O(m^{1.5} \log m)$ time. It follows that Step 2 in total requires $O(m^{2.5} \log m)$ time. Since Step 3 can be computed in $O(m)$ time (Lemma 2), the whole algorithm requires $O(m^{2.5} \log m)$ time. □

Theorem 11 The MCSP problem for $\{<, \not\leq\}$-structured patterns is approx-
imable within ratio $2\mathcal{H}(k)$ in $O(nm^{2.5}\log^2 m)$ time, where $k$ is the size of an optimal solution, $n = |G|$, and $m$ is the maximum size of any linear graph in $G$.

**PROOF.** The algorithm is as follows: let $P$ be a sequence of all pairs $(w, d) \in \{1, \ldots, m\} \times \{1, \ldots, m\}$ ordered in such a way such that $(w, d)$ precedes $(w', d')$ in this ordering if $wd < w'd'$. The algorithm binary searches through all these pairs, to find the largest balanced sequence of staircases that occurs in each graph in $G$. According to Corollary 9, this algorithm is indeed $2\mathcal{H}(k)$-approximate. Furthermore, due to Lemma 10, checking whether a balanced sequence of staircases of width $w$ and depth $d$ occurs in a linear graph of size $m$ can be done in $O(m^{2.5}\log m)$ time. Thus the complexity of the entire algorithm is $O(nm^{2.5}\log^2 m)$ time. □

4 \{\sqsubset, \sqsupset\}\text{-Structured Patterns}

We now consider \{\sqsubset, \sqsupset\}\text{-structured patterns. We begin by proving a hardness result, analogous to Theorem 6, which states that the MCSP problem for \{\sqsubset, \sqsupset\}\text{-structured patterns is NP-hard even if the input consists of towers of staircases of depth at most 2. However, contrarily to the general approach we used for \{\langle, \rangle\}\text{-structured patterns, we cannot use towers of staircases to obtain a good approximation ratio. Indeed, we show that there exists a \{\sqsubset, \sqsupset\}\text{-comparable linear graph of size $k$ which does not contain a tower of staircases of size $\varepsilon\sqrt{k}$ for some constant $\varepsilon$. On the other hand, such a graph must contain either a tower or a staircase with at least $\sqrt{k}$ edges.}

**Theorem 12** The MCSP problem for \{\sqsubset, \sqsupset\}\text{-structured patterns is NP-hard even if each input linear graph is a tower of staircases of depth at most 2.}

**PROOF.** As in proof of Theorem 6, we reduce from the MCSP problem for \{\langle, \rangle\}\text{-structured patterns which is NP-hard even if the input consists only of sequences of staircases, each of height at most 2 (Theorem 4). Let an arbitrary instance of the MCSP problem for \{\langle, \rangle\}\text{-structured patterns be given by a set $G = \{G_1, G_2, \ldots, G_n\}$ of $n$ sequences of towers with height at most 2. The corresponding instance of the MCSP problem for \{\langle, \rangle\}\text{-structured patterns consists of a set $G' = \{G'_1, G'_2, \ldots, G'_n\}$ of $n$ towers of staircases of depth at most 2, where for each $i = 1, 2, \ldots, n$, the graph $G'_i$ is obtained from $G_i$ by taking all the vertices which are right endpoints, reversing the order amongst them, and then placing them to the left of all left endpoints. Below is an example of such a transformation.
It is easily seen that any linear graph $G'_i$ is a tower of staircases of depth at most 2. Furthermore, any pair of edges in $G_i$ are $\{<\}$-comparable (resp. $\{\sqsubset\}$-comparable) if and only if their corresponding edges in $G'_i$ are $\{\sqsubset\}$-comparable (resp. $\{\triangleleft\}$-comparable). Then it follows that there exists a common $\{<,\sqsubset\}$-structured pattern of size $k$ in $G$ if and only if there exists a common $\{\sqsubset,\triangleleft\}$-structured pattern of size $k$ in $G'$. □

Note that the same theorem applies for staircases of towers, since if $G$ is a tower of staircases then $G'$ (as defined in the proof above) is a staircase of towers.

We next consider approximating the MCSP problem for $\{\sqsubset,\triangleleft\}$-structured patterns. First, let us observe the one-to-one correspondence between $\{\sqsubset,\triangleleft\}$-structured patterns and permutations. Let $G$ be a $\{\sqsubset,\triangleleft\}$-comparable linear graph of size $k$. Then the vertices in $G$ which are left endpoints of edges are labeled $\{1, \ldots, k\}$ and the right endpoints are labeled $\{k + 1, \ldots, 2k\}$. The permutation $\pi_G$ corresponding to $G$ is defined by $\pi_G(j - k) = i$ if and only if $(i, j) \in E(G)$. Clearly, all $\{\sqsubset,\triangleleft\}$-comparable linear graphs have corresponding permutations, and vice versa. It follows from this bijective correspondence that the number of different $\{\sqsubset,\triangleleft\}$-comparable linear graphs of size $k$ is exactly $k!$. Moreover, notice that increasing subsequences in $\pi_G$ correspond to $\{\triangleleft\}$-comparable subgraphs of $G$, while decreasing subsequences correspond to $\{\sqsubset\}$-comparable subgraphs (see Figure 4). More generally, a permutation $\sigma \in S_k$ is said to be a pattern (or to occur) within a permutation $\pi \in S_n$ if $\pi$ has a subsequence that is order-isomorphic to $\sigma$ [6], i.e., there exist indices $1 \leq i_1 \leq i_2 \leq \ldots \leq i_k \leq n$ such that, for $1 \leq x \leq y \leq k$, $\pi_{i_x} < \pi_{i_y}$ if and only if $\sigma_x < \sigma_y$. If $\pi$ does not contain $\sigma$, we say that $\pi$ avoids $\sigma$, or that it is $\sigma$-avoiding. For example, $\pi = (2 4 5 3 1)$ contains $(1 3 2)$ because the subsequence $(\pi_1 \pi_3 \pi_4) = (2 5 3)$ has the same relative order as $(1 3 2)$. However, $\pi = (4 2 3 5 1)$ is $(1 3 2)$-avoiding. In the light of the above bijection it is now clear that a linear graph $H$ does not occur in another linear graph $G$ if and only if $\pi_G$ avoids $\pi_H$.

The well-known Erdős-Szekeres Theorem [15] states that any permutation on $1, \ldots, k$ contains either an increasing or a decreasing subsequence of size at least $\sqrt{k}$. Hence, using the algorithms in Lemma 2 for finding the maximum
Figure 4. A \( \{\sqsubset, \sqsubseteq\} \)-structured pattern \( G \) and the corresponding permutation \( \pi_G = (594763218) \). Also illustrated is the bijective correspondence between decreasing subsequences (resp. increasing subsequences) of \( \pi_G \) and \( \{\sqsubset\} \)-structured (resp. \( \{\emptyset\} \)-structured) patterns of \( G \).

common \( \{\sqsubset\} \)-structured and \( \{\emptyset\} \)-structured patterns, we obtain the following theorem.

**Theorem 13**  The MCSP problem for model \( \mathcal{M} = \{\sqsubset, \emptyset\} \) is approximable within ratio \( \sqrt{k} \) in \( O(nm^{1.5}) \) time, where \( k \) is the size of an optimal solution \( n = |G| \), and \( m = \max_{G \in \mathcal{G}} |E(G)| \).

For \( k \in \mathbb{N} \) and \( \mathcal{R} \subseteq \{<, \sqsubseteq\} \), let \( \mathcal{G}_k \) be a set of \( \mathcal{R} \)-comparable linear graphs of size \( k \). Notice that the approximation algorithms we have offered and analyzed within this paper are all based on the key idea of identifying a family \( \mathcal{G}_k, k \in \mathbb{N} \), of \( \mathcal{R} \)-comparable linear graphs with \( |\mathcal{G}_k| \leq \text{poly}(k) \) and such that every \( \mathcal{R} \)-comparable linear graph of size at least \( f(k) \) contains at least one member of \( \mathcal{G}_k \). Clearly, this leads to a poly-time \( \frac{k}{f(k)} \)-approximation algorithm whenever the function \( \text{poly}(\cdot) \) is polynomial. We now want to first address the inherent limitations of this approach.

For \( k \in \mathbb{N} \), let \( \Pi_k \subseteq S_k \) be a set of \( |\Pi_k| \) permutations on \( k \) elements. From what said above, each permutation in \( \Pi_k \) can be equivalently regarded as a
comparable linear graph. Alon [1] recently communicated us a proof of essentially the following lemma.

**Lemma 14 ([1])** For every family of permutations $\Pi_k \subseteq S_k$, $k \in \mathbb{N}$ and $|\Pi_k| \leq 2^k$, there exists a permutation $\pi \in S_K$, $K = \Omega(k^2)$, which avoids all permutations in $\Pi_k$.

**PROOF.** The probability that a random (uniform) permutation in $S_k$ is one of the permutations in $\Pi_k$ is at most $2^k/k!$. Consider a random (uniform) permutation $\tilde{\pi}_K \in S_K$ and notice that there are at most $\binom{K}{k}$ distinct permutations in $S_k$ which are subpermutations of $\tilde{\pi}_K$. Moreover, each of these subpermutations of $\tilde{\pi}_K$ is a random (uniform) permutation in $S_k$, though these subpermutations are not mutually independent.

Therefore, by the linearity of expectation (which does not require independence to hold), the expected number of permutations in $\Pi_k$ which are subpermutations of a random permutation $\tilde{\pi}_K$ on $K = \varepsilon k^2$ elements is strictly less than 1. Thus there must exist a permutation $\pi_K \in S_K$ which avoids all permutations in $\Pi_k$, and so the lemma is proved. \qed

Note that the above lemma shows the tightness of the positive approximability result offered in this section, at least within the general approach pursued within this paper. In particular, it shows that there exists a $\{\sqsubseteq, \emptyset\}$-comparable
linear graph of size $K = \Omega(k^2)$ which does not contain any $\{\sqsubseteq, \sqsupseteq\}$-comparable linear graph out of a family of at most $2^k$ such graphs. Hence, even resorting on more involved or interesting families of $\{\sqsubseteq, \sqsupseteq\}$-comparable linear graphs to be used to probe our input graphs, no approximation guarantee better than $O(\sqrt{k})$ for maximum common $\{\sqsubseteq, \sqsupseteq\}$-structured patterns can be possibly achieved.

5 General Structured Patterns

In this section we consider the MCSP problem for general patterns, i.e., $\{\prec, \sqsubseteq, \sqsupseteq\}$-structured patterns. First, since $\{\prec, \sqsubseteq, \sqsupseteq\}$-structured patterns generalize all other types of patterns, all hardness results presented in previous sections apply for general structured patterns as well. We present three approximation algorithms with increasing time complexities but decreasing approximation ratios. The first one achieves an approximation ratio of $O(k^{2/3})$ in $O(nm^{1.5})$ time while the second one achieves an approximation ratio of $O(\sqrt{k \log^2 k})$ in $O(nm^2)$ time. The third one is an $O(\sqrt{k \log k})$-approximation algorithm which runs in $O(nm^{2.5} \log^2 m)$ time. All algorithms rely on sufficiently large sub-patterns that occur in any $\{\prec, \sqsubseteq, \sqsupseteq\}$-structured pattern, and the fact that finding maximum common structured patterns of these types is polynomial-time solvable.

We first observe that both relations $\prec$ and $\sqsubseteq$ induce partial orders on the edges of a given linear graph. Recall now that a chain (resp. anti-chain) in a partial order is a subset of pairwise comparable (resp. incomparable) elements. Dilworth’s Theorem [12], which is a generalization of the Erdős-Szekeres Theorem [15], states that in any partial order, the size of the maximum chain equals the size of the minimum anti-chain partitioning. Therefore, in any partial order on $k$ elements, the size of the maximum chain multiplied by the size of the maximum anti-chain is at least $k$. The following lemma states this property in our terms.

**Lemma 15** Let $H$ be a $\{\prec, \sqsubseteq, \sqsupseteq\}$-comparable linear graph of size $k$, width $w(H)$, and height $h(H)$. Also, let $hd(H)$ and $wd(H)$ be the sizes of the maximum $\{\sqsubseteq, \sqsupseteq\}$-comparable and $\{\prec, \sqsupseteq\}$-comparable subsets of $E(H)$. Then $k \leq w(H) \cdot hd(H)$ and $k \leq h(H) \cdot wd(H)$.

An immediate consequence of Lemma 15 is as follows.

**Lemma 16** Let $H$ be a $\{\prec, \sqsubseteq, \sqsupseteq\}$-comparable linear graph of size $k$. Then $H$ contains a simple structured pattern of size at least $k^{1/3}$.
**Proof.** If the width of $H$ is at least $k^{1/3}$ we are done. Otherwise, according to Lemma 15, $H$ has a subgraph $H'$ with at least $k^{2/3}$ edges which is $\{\sqsubset, \emptyset\}$-comparable. Applying Lemma 15 on $H'$, we obtain that $H'$ has a subgraph with at least $(k^{2/3})^{1/2} = k^{1/3}$ edges which is either $\{\sqsubset\}$-comparable or $\{\emptyset\}$-comparable. □

Combining the above lemma with the fact that a maximum common simple structured pattern of $G$ can be found in $O(nm^{1.5})$ time (Corollary 3), we obtain our first approximation algorithm for general structured patterns.

**Theorem 17** The MCSP problem for $\{<,\sqsubset,\emptyset\}$-structured patterns is approximable within ratio $O(k^{2/3})$ in $O(nm^{1.5})$ time, where $k$ is the size of an optimal solution, $n = |G|$, and $m = \max G \in G |E(G)|$.

It is easily seen, however, that Lemma 16 is tight. One way to obtain an extremal example of this is as follows: take $k^{1/3}$ balanced towers of staircases, each one of depth $k^{1/3}$ and height $k^{1/3}$, and concatenate them one next to the other into one supergraph of size $k$, reassigning labels accordingly.

![k^{1/3} balanced towers of staircases, each of width k^{1/3} and depth k^{1/3}](image)

**Lemma 18** Let $k$ be an integer such that $k^{1/3}$ is also integer. Then there exists a $\{<,\sqsubset,\emptyset\}$-comparable linear graph of size $k$ that does not contain a simple structured pattern of size $\varepsilon k^{1/3}$ for any $\varepsilon > 1$.

Dilworth’s Theorem does not apply on the crossing relation since it is not transitive. However, an analogous result proven in [25] (see also [26]) implies that for any $\{<,\sqsubset,\emptyset\}$-comparable linear graph $H$, $|E(H)| = O(d \cdot wh \log wh)$, where $d$ and $wh$ are sizes of the maximum $\{\emptyset\}$-comparable and $\{<,\sqsubset\}$-comparable subsets of $E(H)$, respectively. This yields the following analogous of Lemma 15.

**Lemma 19** Let $H$ be a $\{<,\sqsubset,\emptyset\}$-comparable linear graph of size $k$. Then $H$ contains a subgraph of size $\Omega(\sqrt{k / \log k})$ which is either $\{<,\sqsubset\}$-comparable or $\{\emptyset\}$-comparable.

Using Lemma 19, the algorithm for finding a maximum $\{\emptyset\}$-structured pattern given in Corollary 3, and the $O(\log k)$-approximation algorithm for $\{<,\sqsubset\}$-structured patterns given in Theorem 5, we obtain our second approximation algorithm.
Theorem 20 The MCSP problem for \(\{<, \sqsubset, \sqsupseteq\}\)-structured patterns is approximable within ratio \(O(\sqrt{k \log^2 k})\) in \(O(nm^2)\) time, where \(k\) is the size of an optimal solution, \(n = |G|\), and \(m = \max_{G \in \mathcal{G}} |E(G)|\).

For our third algorithm, we show that any \(\{<, \sqsubset, \sqsupseteq\}\)-comparable linear graph contains a subgraph of sufficient size that is either a tower or a balanced sequence of staircases.

Lemma 21 Let \(H\) be a \(\{<, \sqsubset, \sqsupseteq\}\)-comparable linear graph of size \(k\). Then \(H\) contains either a tower or a balanced sequence of staircases of size \(\Omega(\sqrt{k/\log k})\).

**PROOF.** Let \(k_1\) and \(k_2\) denote the sizes of the maximum \(\{\sqsubset\}\)-comparable and \(\{<, \sqsupseteq\}\)-comparable subgraphs of \(H\), respectively.

According to Lemma 15, we have \(k \leq k_1k_2\). If \(k_1 = \Omega(\sqrt{k/\log k})\), we are done. Otherwise, \(k_1 = o(\sqrt{k/\log k})\) and so \(k_2 = \Omega(\sqrt{k \log k})\). From Corollary 9 it follows that \(H\) contains a balanced sequence of staircases of size

\[
\frac{k_2}{2\mathcal{H}(k_2)} = \Omega\left(\frac{\sqrt{k \log k}}{\log \sqrt{k \log k}}\right) = \Omega(\sqrt{k/\log k}),
\]

and the lemma follows. \(\square\)

Applying Lemma 21 and the algorithms for finding the maximum common tower and balanced sequence of staircases in \(G\) given in Corollary 3 and Theorem 11, respectively, we obtain the following theorem.

Theorem 22 The MCSP problem for \(\{<, \sqsubset, \sqsupseteq\}\)-structured patterns is approximable within ratio \(O(\sqrt{k \log k})\) in \(O(nm^{2.5} \log^2 m)\) time, where \(k\) is the size of an optimal solution, \(n = |G|\), and \(m = \max_{G \in \mathcal{G}} |E(G)|\).

We next consider subgraphs of \(\{<, \sqsubset, \sqsupseteq\}\)-comparable linear graphs that are comparable by pairs of relations, i.e., by \(R \subseteq \{<, \sqsubset, \sqsupseteq\}\) with \(|R| = 2\). We show that any \(\{<, \sqsubset, \sqsupseteq\}\)-comparable linear graph of size \(k\) contains such a subgraph of size at least \(k^{2/3}\), and that this lower bound is relatively tight. Unfortunately, this result cannot be applied for approximation purposes (approximating the MCSP problem for \(\{\sqsubset, \sqsupseteq\}\)-patterns remains the bottleneck). Nevertheless, we present this result on account of independent interest.

Lemma 23 Let \(H\) be a \(\{<, \sqsubset, \sqsupseteq\}\)-comparable graph of size \(k\). Then \(H\) has a subgraph of size \(\varepsilon k^{2/3}\), where \(\varepsilon = \frac{\sqrt{7} - 1}{8}\), which is either \(\{<, \sqsubset\}\)-comparable, \(\{<, \sqsupseteq\}\)-comparable, or \(\{\sqsubset, \sqsupseteq\}\)-comparable.
PROOF. For simplicity, assume that both $k^{1/3}$ and $k^{2/3}$ are integers. Since $H$ is $\{<, \sqsubset, \emptyset\}$-comparable, we may assume that $V(H) = \{1, 2, \ldots, 2k\}$. For each $x = 1, 2, \ldots, 2k - 1$, let $E_x \subseteq E(H)$ be the set of edges $(i, j)$ with $i \leq x < j$. Note that for all $x = 1, 2, \ldots, 2k - 1$, $H[E_x]$, the subgraph of $H$ obtained by deleting all edges $E \setminus E_x$, is a $\{\sqsubset, \emptyset\}$-comparable subgraph of $H$. Therefore, if $|E_x| > \varepsilon k^{2/3}$ for some $x = 1, 2, \ldots, 2k - 1$, we are done.

Otherwise, for $y = 1, 2, \ldots, k^{1/3}$, let $p_y$ be the smallest integer in $\{1, 2, \ldots, 2k\}$ such that there are at least $k^{2/3}$ edges $(i, j) \in E(H)$ with $p_{y-1} \leq i < p_y$, where $p_0 = 1$. Also, for each $y = 1, 2, \ldots, k^{1/3}$, let $H_y$ be the subgraph of $H$ made of edges $(i, j)$ with $p_{y-1} \leq i < j < p_y$. Since $|E_x| \leq \varepsilon k^{2/3}$ for all $x = 1, \ldots, 2k - 1$, then $H_y$ contains at least $k^{2/3} - \varepsilon k^{2/3} = (1 - \varepsilon)k^{2/3}$ edges. According to Lemma 15, for each $y = 1, 2, \ldots, k^{1/3}$, $H_y$ contains either (i) a $\{<, \emptyset\}$-comparable subgraph $H'_y$ or (ii) a $\{\sqsubset\}$-comparable subgraph $H''_y$ of size at least $((1 - \varepsilon)k^{2/3})^{1/2} = (1 - \varepsilon)^{1/2}k^{1/3}$. Let $Y$ be the set of those $y$’s for which case (i) occurs. Then $H' = \bigcup_{y \in Y} H'_y$ is a $\{<, \emptyset\}$-comparable subgraph of $H$ and $H'' = \bigcup_{y \in Y} H''_y$ is a $\{<, \sqsubset\}$-comparable subgraph of $H$. Furthermore, at least one of $H'$ and $H''$ has at least $2^{-(1 - \varepsilon)^{1/2}k^{1/3}}k^{1/3} = 2^{-1} (1 - \varepsilon)^{1/2}k^{2/3}$ edges. But $2^{-1}(1 - \varepsilon)^{1/2} = \varepsilon$ for $\varepsilon = \frac{\sqrt{17} - 1}{8} \approx 0.39$, which proves the lemma. □

We believe the bound of Lemma 23 to be not the best possible. However, combining Lemmas 15 and 18, we show that the above lemma is relatively tight.

Lemma 24 Let $k$ be an integer such that $k^{1/3}$ is an integer. Then there exists a $\{<, \sqsubset, \emptyset\}$-comparable linear graph of size $k$ that contains neither a $\{<, \sqsubset\}$-comparable subgraph, nor a $\{<, \emptyset\}$-comparable subgraph, nor a $\{\sqsubset, \emptyset\}$-comparable subgraph of size at least $\varepsilon k^{2/3}$ for any $\varepsilon > 1$.

PROOF. Suppose, aiming at a contradiction, that the lemma is false. Then any $\{<, \sqsubset, \emptyset\}$-comparable linear graph of size $k$ contains a subgraph of size at least $\varepsilon k^{2/3}$, $\varepsilon > 1$, which is either $\{<, \sqsubset\}$-comparable, $\{<, \emptyset\}$-comparable, or $\{\sqsubset, \emptyset\}$-comparable. Therefore, by Lemma 15, any $\{<, \sqsubset, \emptyset\}$-comparable linear graph of size $k$ contains a subgraph with at least $\sqrt[3]{\varepsilon k^{2/3}} = \varepsilon^{1/2}k^{1/3}$ edges which is either $\{<\}$-comparable, $\{\sqsubset\}$-comparable, or $\{\emptyset\}$-comparable. According to Lemma 18, this is the desired contradiction. □

6 Discussion and Open Problems

In this paper, we introduced the MCSP problem as a general framework for many structure-comparison and structure-prediction problems, that occur
mainly in computational molecular biology. Our framework followed the approach in [33] by analyzing all types of $\mathcal{R}$-structured patterns, $\mathcal{R} \subseteq \{<, \sqsubseteq, \bowtie\}$. We gave tight hardness results for finding maximum common $\{<, \bowtie\}$-structured patterns and maximum common $\{<, \emptyset\}$-structured patterns. We also proved that the MCSP problem is approximable within ratios: (i) $2^\mathcal{H}(k)$ for $\{<, \bowtie\}$-structured patterns, (ii) $k^{3/2}$ for $\{\sqsubseteq, \emptyset\}$-structured patterns, and (iii) $O(\sqrt{k \log k})$ for $\{<, \sqsubseteq, \emptyset\}$-structured patterns.

There are many questions left open by our study. Below we list some of them. According to Lemma 23, we could substantially improve in terms of approximation ratio on all the algorithms suggested for general structured patterns, if we had a better approximation algorithm for $\{\sqsubseteq, \bowtie\}$-structured patterns. Is there an approximation algorithm which achieves a better ratio than the simple $\sqrt{k}$ algorithm? On the same note, can lower bounds on the approximation factor of the MCSP problem for $\{<, \bowtie\}$-structured patterns or $\{\sqsubseteq, \emptyset\}$-structured patterns be proven? How about $\{<, \sqsubseteq\}$-structured patterns or $\{<, \emptyset\}$-structured patterns? Besides, the last decade has seen an increasing development in the theory of parameterized complexity [13]. Can the MCSP problem be tackled better by applying this theory? Last but not least, the MCSP problem for $\{<, \bowtie\}$-structured patterns is still open in case $|G| = 1$, i.e., the case where the input consists of one linear graph $G$ and one wishes to find the largest $\{<, \bowtie\}$-comparable subgraph of $G$ (see [5,9,33]). Is there a polynomial-time algorithm for this problem?

References


