Decidability of string graphs ☆

Marcus Schaefer[a,*] and Daniel Štefankovič[b,c]

[a] *Department of Computer Science, DePaul University, 243 South Wabash, Chicago, IL 60604, USA*
[b] *Department of Computer Science, University of Chicago, 1100 East 58th Street, Chicago, IL 60637, USA*
[c] *Department of Computer Science, Comenius University, Bratislava, Slovakia*

Abstract

We show that string graphs can be recognized in nondeterministic exponential time by giving an exponential upper bound on the number of intersections for a minimal drawing realizing a string graph in the plane. This upper bound confirms a conjecture by Kratochvíl and Matoušek (J. Combin. Theory. Ser. B 53 (1991).) and settles the long-standing open problem of the decidability of string graph recognition (Bell System Tech. J. 45 (1996) 1639; Open Problem at Fifth Hungarian Collogium on Combinatories, 1976). Finally we show how to apply the result to solve another old open problem: deciding the existence of Euler diagrams, a fundamental problem of topological inference (Proceedings of the 14th International Joint Conference on Artificial Intelligence, 1995, p. 901). The general theory of Euler diagrams turns out to be as hard as second-order arithmetic.

1. Introduction

Is it possible that some A is B, some B is C, but no A is C? Easily, you say, and your mind conjures up a diagram that Euler (and Leibniz, and Sturm before him) would have used to illustrate this situation (see Fig. 1).

However, it is not always possible to illustrate a situation that is logically consistent by a Euler diagram in the plane: we can turn the complete graph on five vertices into an example with 15 regions, one for each vertex and edge, that cannot be drawn in the plane [Sin66]. Given a set of specifications, can we effectively determine whether there is a Euler diagram or not?

☆ An earlier version of this paper appeared as DePaul Technical Report TR00-005, September 2000.

* Corresponding author.
E-mail addresses: mschaefer@cs.depaul.edu (M. Schaefer), stefanko@cs.uchicago.edu (D. Štefankovič).

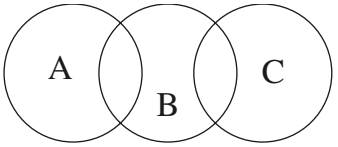

Fig. 1. Some A is B, some B is C, but no A is C.

Diagrammatic reasoning is concerned with the representability of logical relations in the plane and other spaces. This area has drawn attention from different research groups including Artificial Intelligence and Geometrical Information Systems [All83,GPP95], Spatial Databases [PSV99], Integrated Circuits [Sin66], and Logic [GPP95,LP97]. One of the major open problems in this area is the decidability of the existence of a representation for a given, logically consistent, formula. Even the special case illustrated in Fig. 1 in which we specify for a collection of regions whether they should intersect or not has been open since the 1960s.

This case is captured by the combinatorial notion of string graphs. String graphs are intersection graphs of curves in the plane with a vertex for each curve, and an edge representing an intersection between two curves. The notion of string graphs was hinted at in a 1959 paper by Benzer [Ben59] on genetic structures, and isolated in 1966 by Sinden [Sin66] who stated the main problem thus:

> It is specified which pairs of a collection of curves (or connected regions) in the plane cross and which pairs do not cross. When are such specifications consistent?

Sinden was working on the layout problem of integrated circuits (thin film RC circuits to be precise), and the string graph problem arose naturally in this context, since the technology for creating the circuits made it possible for some pairs of conductors to cross. On the theoretical side he observed that all planar graphs are string graphs, and also gave the example described earlier of a graph which is not a string graph.

Ron Graham, in 1976, introduced the problem to the combinatorial community [Gra76], and in the same year Ehrlich et al. [EET76] showed that computing the chromatic number of string graphs is **NP**-hard.

Several special cases of the string graph problem were found to be effectively solvable. Benzer's paper, for example, suggested the notion of interval graphs, that is, intersection graphs of subintervals of the real line. Interval graphs, as well as circular-arc graphs, and series-parallel string graphs were shown to be recognizable in polynomial time [McC01,MNT88], but the general recognition problem remained open. It was not even known whether the problem is decidable. In 1991 Kratochvíl [Kra91b] proved that recognizing string graphs is **NP**-hard, showing that a characterization cannot be polynomial time computable (unless $\mathbf{P} = \mathbf{NP}$). At the same time Kratochvíl and Matoušek [KM91] proved the surprising result that some string graphs require an exponential number of intersections to be realized in the plane. They conjectured an exponential upper bound on the number of intersections.

In this paper we show that this conjecture is indeed true, placing the recognition problem of string graphs in **NEXP**. János Pach and Géza Tóth independently obtained a proof of the decidability of the string graph recognition problem by showing that certain configurations of cells do not occur in a minimal realization [PT01].

Because the exponential upper bound on intersections is matched by an exponential lower bound, we cannot hope to improve the complexity analysis by a sharper upper bound. However, it turns out that although there can be an exponential number of intersections, these intersections are very structured and allow a compressed representation of polynomial size. Using recent results on the solvability of word equations Schaefer et al. build on the results of the present paper to show that string graphs can be recognized in **NP** [SSS01] which is a tight result, since the problem is known to be **NP**-hard [Kra91b].

As one of the motiviations of the study of string graphs we mentioned the realizability Euler diagrams. In Euler diagrams we distinguish only two possible relations between regions: they intersect, or they do not intersect. Topological inference allows a much more refined set of predicates to specify the mutual relation of any pair of regions giving us the very expressive topological expressions. In Section 4 we show that the realizability of topological expressions can be reduced to the string graph problem, and is therefore in **NP** as well (using the result from [SSS01] that string graph recognition is in **NP**).

In Section 5 we look at the complexity of the theory of diagrams, in which we allow quantified expressions. Euler diagrams are the special case in which we only consider existential expressions. The theory is undecidable, as a matter of fact, it is as hard to decide as truth in second-order arithmetic.

2. Preliminaries

A *homeomorphism* is a bijective continuous mapping whose inverse is also continuous. A *curve*, or *string*, is a set homeomorphic to $[0, 1]$ which implies that strings, or curves, do not self-intersect and are compact. Given a collection of curves $(C_i)_{i\in I}$ in the plane, the corresponding intersection graph is $(I, \{\{i,j\} : C_i \text{ and } C_j \text{ intersect}\})$. The *size* of a collection of curves is the number of intersection points. We assume that no three curves intersect in the same point. A graph isomorphic to the intersection graph of a collection of curves in the plane is called a *string graph*. Let $c_s(G)$ be the size of a smallest (in the sense of size defined above) set of curves whose intersection graph is isomorphic to G, and define $c_s(m) = \max\{c_s(G) : G \text{ has } m \text{ edges}\}$.[1] It is not immediately obvious that $c_s(G)$ is a finite number if G is a string graph. Conceivably, an infinite number of intersections might be required to realize a string graph. As was observed earlier by Kratochvíl et al. [KGK86], this is not the case: we can assume that the curves intersect only finitely often. For completeness we include a proof of this result.

Lemma 2.1 (Kratochvíl et al. [KGK86]). *A string graph can be realized by a family of polygonal arcs with a finite number of intersections.*

Proof. Assume we have a string graph realized by a family of curves $(C_i)_{i\in I}$. Let $C = \cup_{i\in I} C_i$. Note that C is a compact set. For a point p in C, and an open neighborhood O of p in the plane we say that O *respects* the family of curves, if O only intersects curves that contain p. Every point has

[1] The functions defined in this section are based on similar functions in the papers by Kratochvíl and Matoušek [KM91,Kra98].

at least one such neighborhood (in fact infinitely many), since all curves are compact. (If a compact set intersects all open neighborhoods of a point, it has to contain that point.) Let $\mathcal{O}$ be the collection of all O such that O is an open neighborhood of a point $p \in C$ such that O respects $(C_i)_{i \in I}$. Then $\mathcal{O}$ is an open cover of C, and, by the Heine–Borel Theorem (which we can apply since C is compact), it contains a finite cover $\mathcal{O}'$ of C. Therefore, each curve C_i is contained in a finite collection $\mathcal{O}'_i \subseteq \mathcal{O}'$ of open sets none of whom contain any points of curves C_j that C_i does not intersect. Hence we can replace C_i with a simple polygonal arc P_i in $\mathcal{O}'_i$, while guaranteeing that P_i intersects P_j if C_i and C_j intersect. The later is possible since there must be a set $O \in \mathcal{O}'$ which contains a common intersection point of C_i and C_j. □

Given a graph $G = (V, E)$ and a set $R \subseteq \binom{E}{2} = \{\{e, f\} : e, f \in E\}$ on E, we call a drawing D of G in the plane a *weak realization* of (G, R) if only pairs of edges which are in R are allowed to intersect in D. We call (G, R) *weakly realizable* if it has a weak realization. Note that in a weak realization the pairs of edges specified in R do not *have* to intersect. We say that D is a *realization* of G, and call G *realizable*, if *exactly* the pairs of edges in R intersect in D.[2] Let us define $c_w(G, R)$ as the smallest number of intersections in a weak realization of (G, R), $c_w(G) = \max\{c_w(G, R) : (G, R) \text{ has a weak realization}\}$, and $c_w(m) = \max\{c_w(G) : G \text{ has } m \text{ edges}\}$. Similarly define $c_r(G, R)$, $c_r(G)$, and $c_r(m)$ for realizations. The quantity $cr(G) = c_w(G, \binom{E}{2})$ is known as the *crossing number* of the graph G, and was shown to have an **NP**-complete decision problem by Garey and Johnson in the early 1980s [GJ83]. The other extreme case, $c_w(G, \emptyset)$, is equivalent to planarity testing, and therefore in **P**.

The following relationships between the functions we defined are well known:

(i) $c_w(m) \leqslant c_r(m)$,
(ii) $c_r(m) \leqslant 4c_s(m^2 + 4m)$, and
(iii) $c_s(m) \leqslant 4c_w(2m) + 2m$.

The first inequality follows from $c_w(G, R) \leqslant \max\{c_r(G, R') : R' \subseteq R, \text{ and } (G, R) \text{ has a realization}\}$. Kratochvíl gave a reduction of realizability to string graphs [Kra91b, Proposition 1]: suppose we are given (G, R) where $G = (V, E)$. Let $V' = V \cup E \cup \{(u, e) : u \in e \in E\}$, and $E' = R \cup \{\{u, (u, e)\}, u \in e \in E\} \cup \{\{e, (u, e)\}, u \in e \in E\}$. Then (G, R) is realizable, if and only if $G' = (V', E')$ is a string graph. This reduction implies the second inequality. Finally, the string graph problem can be reduced in polynomial time to the weak realizability problem [MNT88,Kra91b] as follows. Given a graph $G = (V, E)$, let $G' = (V \cup E, \{\{u, e\} : u \in e \in E\})$, and $R = \{\{\{u, e\}, \{v, f\}\} : \{u, v\} \in E\}$. Then G is a string graph if and only if (G', R) is weakly realizable. This reduction implies the third inequality.

Kratochvíl and Matoušek [KM91] showed that $c_w(m) \geqslant 2^{cm}$ for some positive constant c. Our main result shows that $c_w(m) \leqslant m2^m$ (Kratochvíl and Matoušek conjectured an upper bound of 2^{m^k}). This implies the decidability of string graphs, which was a long-standing open problem in the field.

[2] Kratochvíl [Kra98,Kra91a,Kra91b] calls (G, R) an *abstract topological graph*, and uses the word *feasible* for weakly realizable.

3. Bounding the Number of Intersections

If we assign each curve in a collection of curves a unique number, we can look at the intersections of the curves along a particular curve as a word over an alphabet (here we use the fact that the number of intersections is finite).

The basic idea behind the exponential upper bound is to show that if such a word is too long, it contains a substructure which can be redrawn using a smaller number of intersections. This allows us to bound the number of intersections along every curve in a drawing of smallest size, thereby bounding the size of the whole drawing.

Lemma 3.1. *Every word of length* $\geqslant 2^n$ *over an alphabet of size n contains a non-trivial subword in which every character occurs an even number of times.*

Proof. Let $\Sigma = \{1, \ldots, n\}$, and $w \in \Sigma^*, |w| \geqslant 2^n$. To every $i \in \{0, \ldots, 2^n\}$ assign a vector v_i in $\mathbb{Z}_2^n$ whose jth coordinate is the parity of the number of occurrences of the symbol j in the prefix of w of length i. (In particular, v_0 is the all-zero vector.) Since there are $2^n + 1$ indices, but only 2^n vectors in $\mathbb{Z}_2^n$, there are $0 \leqslant i < j \leqslant 2^n$ such that $v_i = v_j$. The non-trivial subword of w starting in position $i + 1$ and ending in position j fulfills the conditions of the lemma. □

Theorem 3.2. *Let G be a graph with m edges,* $R \subseteq \binom{E}{2}$ *such that* (G, R) *is weakly realizable, and let D be a weak realization of* (G, R) *with the minimal number of intersections. Then for any edge* $e \in G$ *there are less than* 2^m *intersections on the curve realizing e in D.*

Proof. Suppose not. Let D be a weak realization of (G, R) with the minimal number of intersections and let e be an edge of G which has more than $2^m - 1$ intersections in D. Lemma 2.1 tells us that the number of intersections in the realization is finite. By Lemma 3.1 we can choose a non-trivial segment of this edge which is intersected an even number of times by any other edge. Draw a window around this segment containing no other intersections of D. This is possible, since the number of intersections in the drawing D is finite. Let $2n_f$ $(n_f \in \mathbb{N})$ be the number of intersections of edge f with e in that window. For each edge f assign numbers $1, 2, \ldots, 4n_f$ to intersections with the window in the order they appear on f (choose an arbitrary orientation of f). For an example see Fig. 2.

We can assume (by an application of the Jordan–Schoenflies theorem [MT01]) that the window is a circle, that e within the window is a straight line passing through the center, and that for every f intersections $2i - 1$ and $2i$ are mirror images of each other (with e as the mirror), $i \in \{1, \ldots, 2n_f\}$.

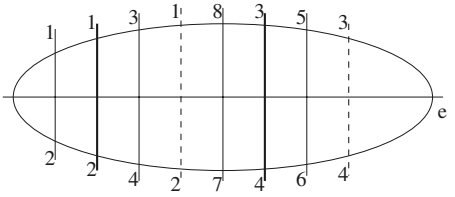

Fig. 2. Segment of e with surrounding window.

(The intersections with e are the only intersections of D within the window, since this is how we chose the window.)

Remove everything inside of the window with the exception of e. For each edge f there is a connection between intersection $4i-2$ and $4i-1$ lying completely outside the window, $i \in \{1, \ldots, n_f\}$. Use circular inversion along the circle to bring all of these connections inside the window. Now mirror everything inside the window along e.

This yields for every edge f a connection between $4i-3$ and $4i$, $i \in \{1, \ldots, n_f\}$, inside the window. We can now build a new version of f: start at intersection 1 (which is connected to one of the endpoints of f), and continue to 4 (inside the window), from 4 to 5 (outside the window), 5 to 8 (inside the window), and so on up to $4n_f$ which is the last intersection of f with the window before it terminates at its other endpoint. Hence this new version of f still connects its two original endpoints (here we needed that f intersects e an even number of times).

Note that we have reduced the number of intersections of f with the window from $4n_f$ to $2n_f$. Every intersection between curves inside the circle corresponds to an intersection outside and hence this new realization respects R. We might have lost intersections between curves, but this is acceptable, since we only require D to be a weak realization.

The number of intersections along e might have increased, since a connection from outside the window brought inside by circular inversion might intersect e arbitrarily often. However, we do know that, overall, we halved the number of intersections between the intersections and the boundary of the window. We can therefore move the part of e inside the circle to coincide with one of the two arcs into which e separates the boundary of the window. We choose the arc which results in the smaller number of intersections with e. Since each edge f causes at most $2n_f$ intersections with the window, this means that the number of intersections on e within the area of the window has been halved, and hence the total number of intersections of the drawing has been decreased, contradicting the assumption that D was of minimal size. □

Corollary 3.3. *String graph recognition is in* **NEXP**.

Proof. Theorem 3.2, and the fact that $c_s(m) \leqslant 4c_w(2m) + 2m$ (see the preliminaries) shows that if G is a string graph, there is a collection of curves of size $M = 2^{O(m)}$ whose intersection graph is isomorphic to G. We can consider the drawing of the collection of curves as a planar graph (each intersection point becoming a vertex) with at most M vertices. By a result of Schnyder [Sch90], and de Fraysseix et al. [dFPP90] there is a drawing of this graph on an $M \times M$ grid. Hence in **NEXP** we can guess a graph on such a grid and verify whether its intersection graph is isomorphic to G. □

The same argument shows that we can decide the (weak) realizability of a topological graph (G, R) in **NEXP**.

4. Topological Inference

We mentioned earlier that settling the problem of recognizing string graphs solves an old open problem from topological inference about regions in the plane [CGP98,Sin66]. A *region* is a subset

of the plane homeomorphic to the closed unit disc. Let ∂A denote the boundary of the region A. If we specify for a collection of regions $(A_i)_{i\in I}$ which pairs may intersect and which may not, the question is whether we can draw these regions to meet the specification? Since the existence of such a drawing is not affected if we change the universe of discourse from regions to curves, the problem is equivalent to the string graph problem, and therefore solvable in **NEXP**.[3]

Topological inference works over a larger set of predicates than overlap and disjoint. Egenhofer determined all eight possible relationships of two simply connected regions based on whether the intersection of their interior, boundary and exterior is empty or not [Ege91]. The relations are disjoint, equal, inside, contains, cover, covered, meet, and overlap. See Fig. 3 for definitions. For two simply connected regions A and B exactly one of these predicates will be true.

We call a Boolean combination of the topological predicates a *topological expression*. A topological expression is *explicit*, if it specifies the relationship between any pair of variables, meaning it is of the form $\bigwedge_{A,B\in I} \mathscr{P}_{A,B}(A,B)$, where I is the set of variables, and $\mathscr{P}_{A,B}$ is one of the eight basic predicates (for each $A,B\in I$). We can always assume that the expression does not contain the predicates contains or cover, because we can substitute them by inside and covered. Quantifying topological expressions we obtain *topological formulas*. Determining the truth of these (where the universe is the set of all regions in the plane) is the goal of *topological inference* [GPP95]. Of main interest are the purely existential formulas, since they express the existence of diagrammatic representations of logical relationships (Euler diagrams). In this case we also speak of the *realizability* of a topological expression.

In some special cases the realizability of a topological expressions is known to be in **P**. Planar map graphs were introduced in [CGP98]. A *k-planar map graph* is the intersection graph of a set of regions with disjoint interiors such that at most k regions meet in a point. Planar graphs are exactly the 3-planar map graphs. A graph is called a *planar map graph* if it is a k-planar map graph for some k. The problem of recognizing planar map graphs is equivalent to the realizability of explicit topological expressions containing only relations meet and disjoint. In [CGP98] a polynomial time algorithm for recognizing 4-planar map graphs was given. Thorup [Tho98] found a polynomial time algorithm for recognizing planar map graphs in general. Other special cases of the problem have been classified [GPP95,FH02], but the complexity of the general realizability problem of topological expressions has remained open.

In this section we will show how the decidability of the existential theory of topological expressions follows from the decidability of string graphs. More precisely we show that the realizability of topological expressions can be decided in **NEXP**. The next section complements this result by showing that the general theory is undecidable.

Talking about a realization of meet(A,B), or covered(A,B) we call any point belonging to $\partial A\cap\partial B$ a *contact point* of A and B. In the other cases points belonging to the intersection of ∂A and ∂B we simply call *intersection points*.

The *Hausdorff distance* dist(A,B) of two regions is defined as

$$\operatorname{dist}(A,B)=\max\{\sup_{x\in A}\inf_{y\in B} d(x,y),\sup_{y\in B}\inf_{x\in A} d(x,y)\},$$

[3] Because regions are homeomorphic to the unit disc, they are simply connected. The topological inference problem changes dramatically if we only require regions to be connected. As Kratochvíl [Kra91a, Section 2] points out, in that case *any* specification can be realized.

disjoint(A, B)	the boundaries and interiors of A and B do not intersect.
equal(A, B)	A and B have the same interior.
inside(A, B) ($\equiv$ contains(B, A))	the interior and boundary of A are contained in the interior of B.
covered(A, B) ($\equiv$ cover(B, A))	the interior of A is properly contained in the interior of B, and the boundaries intersect.
meet(A, B)	the interior of A is contained in the exterior of B, and the boundaries intersect.
overlap(A, B)	the interior of A intersects the exterior of B, and vice versa.

Fig. 3. The eight relationships between regions (Egenhofer).

where $d(x, y)$ is the Euclidean distance of two points in the plane. The Hausdorff distance is a metric for regions, i.e. it is symmetric, satisfies the triangle inequality and $\text{dist}(A, B) = 0$ iff $A = B$. We let

$$d(A, B) = \inf_{x \in A} \inf_{y \in B} d(x, y).$$

Note that for closed, non-empty sets $d(A, B) > 0$ iff $A \cap B = \emptyset$. For arbitrary sets d is not a metric.

We will now show how to redraw a realization of an explicit topological expression to bound the number of contact points in the drawing. Note that for any explicit expression there is always an equivalent explicit expression not containing equal.

Lemma 4.1. *Let φ be an explicit topological expression not containing* equal. *If there is a drawing realizing φ, then there is a drawing realizing φ in which the number of contact points on each boundary is bounded by the square of the number of variables in φ.*

Proof. Fix a drawing of φ. Let $A_1, \ldots, A_{|I|}$ be the family of variables occurring in φ. We can assume that the variables are sorted such that for $i \leqslant j$ there is no covered(A_i, A_j) in φ. If such an ordering does not exist, then φ has no realization (since covered(A_i, A_j) means that A_i is properly contained in A_j). For each meet(A, B) and covered(A, B) in φ we choose a witness point $p_{A,B} \in \partial A \cap \partial B$. Let

$$\varepsilon_1 = \min_{i,j \in I} \{d(\partial A_i, \partial A_j) : \varphi \text{ contains disjoint}(A_i, A_j) \text{ or inside}(A_i, A_j)\}$$

$$\varepsilon_2 = \min_{i,j \in I} \{\text{dist}(A_i \cap \partial A_j, \partial A_i) : \varphi \text{ contains overlap}(A_i, A_j)\}$$

Note that $\varepsilon_1 > 0$, since boundaries are closed and disjoint. Also $\varepsilon_2 > 0$, since there is a point in $A_i \cap \partial A_j$ which is inside A_i. Let $\varepsilon = \min\{\varepsilon_1, \varepsilon_2\}/2$. If B is a region with $\text{dist}(B, A_i) \leqslant \varepsilon$ then

$$\begin{aligned} &\text{inside}(A_i, A_j) \;\Rightarrow\; \text{inside}(B, A_j) \quad \text{inside}(A_j, A_i) \;\Rightarrow\; \text{inside}(A_j, B) \\ &\text{disjoint}(A_i, A_j) \;\Rightarrow\; \text{disjoint}(B, A_j) \quad \text{overlap}(A_i, A_j) \;\Rightarrow\; \text{overlap}(B, A_j) \end{aligned} \tag{1}$$

This means changing the regions slightly (up to a Hausdorff distance of ε) does not change the relationships inside, disjoint, and overlap. Unfortunately the same is not true for meet and covered. We will redraw the regions one by one, removing unnecessary contact points while preserving the meet and covered relationships.

Suppose then that for $A_1, \ldots, A_{i-1}$ the only contact points on their boundaries are witness points. We will show how to redraw A_i to make this true for $A_1, \ldots, A_i$ while preserving the condition that $A_1, \ldots, A_{|I|}$ realize φ.

Let $\psi : D \mapsto A_i$ be the homeomorphism of the closed unit disc to A_i. Using the Jordan–Schoenflies theorem [MT01] we extend ψ to a homeomorphism of the whole plane to itself which we call ψ again. Since ψ is uniformly continuous (if we consider it as being defined on the compactification of the plane), there exists η such that if $(1-\eta)D \subseteq E \subseteq D$ then $\operatorname{dist}(\psi(E), A_i) < \varepsilon$. Let F be the union of $\psi^{-1}(A_j)$ for which $\operatorname{covered}(A_i, A_j)$ occurs in φ. By the way we ordered the variables this can only happen for $j < i$. Since we assumed that $A_1, \ldots, A_{i-1}$ only contain witness points as contact points on their boundary, we conclude that $F \cap \partial D$ contains only witness points. Choose E such that $F \cup (1-\eta)D \subseteq E \subseteq D$ and $E \cap \partial D = \{p_{A_i,A_k} : k \in I\}$, that is E intersects ∂D exactly in the witness points and covers F. Replace A_i by $\psi(E)$. By the implications in (1) all inside, disjoints and overlaps are preserved. Because E contains all witness points for region A_i, all covered and meet relations are satisfied after this step, and only the witness points are contact points of A_i. Since contact points of A_j, $j < i$ did not change, this conditions remain true after redrawing all regions. Finally note that we used a quadratic number of contact points (potentially one for each pair of variables). □

Before we prove the decidability result for topological inference we need to introduce a refined variant of realizability. Let (G, R, S) be such that $R, S \subseteq \binom{E}{2}$, and $R \cap S = \emptyset$. We call (G, R, S) *realizable* if G can be drawn in the plane, such that only the pairs of edges in $R \cup S$ intersect, and all the pairs of edges in S *do* intersect. It is easy to see that this variant can also be decided in **NEXP**, since the same exponential upper bounds on the intersection number applies.

Theorem 4.2. *The realizability of a topological expression can be decided in* **NEXP**.

Theorem 4.2 follows from the following lemma which allows us (in **NP**) to translate the realizability of a topological expression to the realizability of some (G, R, S). Since that problem can be solved in **NEXP**, the realizability of topological expressions can be decided in **NEXP**.

Lemma 4.3. *The realizability of a topological expression* **NP***-reduces to the realizability problem of the form* (G, R, S); *that is, for every topological expression* φ *we can in* **NP** *compute triples* (G, R, S) *such that* φ *is realizable, if and only if one of the* (G, R, S) *is realizable.*

Proof. Given a topological expression φ over variables $(A_i)_{i \in I}$ we have to rephrase the problem as a realizability problem of the form (G, R, S). We begin by simplifying φ.

If the topological expression φ can be realized, then there is an explicit topological expression ψ which can be realized, and ψ implies φ. In **NP** we can guess an explicit topological expression ψ over the variables $(A_i)_{i \in I}$, and verify in polynomial time that ψ implies φ (since ψ specifies the

relationship between every pair of variables, this verification corresponds to evaluating the truth of a formula for a given assignment). Hence we can assume that φ is an explicit topological expression to begin with. Furthermore, in polynomial time, we remove the relation of equality from φ by renaming variables, and we substitute any occurrence of cover(B, A) with covered(A, B), and contains(B, A) by inside(A, B).

In summary, we can assume that φ is an explicit topological expression containing (positive) occurrences of the relations disjoint, meet, covered, overlap, inside only.

Suppose that a topological graph (G, R, S) satisfies:

(1) There are vertices z, z_1, z_2, z_3 in G connected to each other by edges which may not intersect any other edges.
(2) For each A_i there is a vertex c_i (center) and a circle graph B_i (boundary) with at least three vertices, and no two edges of B_i may intersect.
(3) Each vertex in B_i is connected to c_i, z_1, z_2, z_3; these edges are not allowed to intersect the boundary B_i, and no edge with endpoint c_i may intersect an edge with endpoint z_1, z_2, or z_3.
(4) The boundaries B_i, B_j may share vertices unless disjoint(A_i, A_j), or inside(A_i, A_j) is contained in φ.
(5) If φ contains meet(A_i, A_j) or cover(A_i, A_j) then B_i and B_j share at least one common vertex.
(6) Edges of B_i, B_j may intersect only if φ contains overlap(A_i, A_j).
(7) We say that a vertex v is an *in-R_i-witness* (*out-R_i-witness*) if it does not belong to B_i and is adjacent to c_i (z_1, z_2, and z_3, resp.) using an edge (edges, resp.) which are not allowed to intersect B_i. If disjoint(A_i, A_j) is in φ, then there is an out-R_i-witness on B_j, and an out-R_j-witness on B_i. If inside(A_i, A_j) then there is an in-R_j-witness on B_i. If meet(A_i, A_j), then there is an out-R_i-witness on B_j between any two vertices shared with B_i, and an out-R_j-witness on B_i between any two vertices shared with B_j. If covered(A_i, A_j) then there is an in-R_i-witness on the boundary B_j between any two vertices shared with B_i. If overlap(A_i, A_j) then there is an in-R_i witness and an out-R_i witness on the boundary B_j, and vice versa.

We claim that if any (G, R, S) fulfilling these conditions has a realization then φ can be realized as an Euler diagram. Consider a realization of (G, R, S). We can assume that z lies outside the triangle z_1, z_2, z_3. Hence by (1) all other vertices and edges lie inside the triangle. Because of (3) vertex c_i must lie inside of B_i (z_1, z_2, and z_3 being outside). Let region R_i be the interior of B_i together with its boundary. By (7) any in-R_i-witness lies inside R_i, and any out-R_i-witness lies in the exterior of R_i. For inside(A_i, A_j), *and* disjoint(A_i, A_j) boundaries may not intersect by (4) and (5) and therefore the in/out-witnesses guarantee the correct relationship between the corresponding regions. For overlap(A_i, A_j) we have in/out-witnesses of overlap. For meet(A_i, A_j) the interior of R_i cannot intersect R_j, and vice versa because of the out-witnesses; similarly for covered(A_i, A_j).

We will next show that if φ can be drawn as an Euler diagram then there is a (G, R, S) which satisfies the conditions above, and whose size is polynomial in $|\varphi|$. This completes the proof, since in **NP** we can guess such a (G, R, S).

If φ is realizable, then we can redraw a graph realizing it using Lemma 4.1 such that the number of contact points is at most $|I|^2$. Enclose the diagram within a new region Z. On ∂Z choose three points z_1, z_2, z_3, choose z outside Z and connect z to z_1, z_2, z_3 with edges outside Z. Choose c_i inside each R_i. Furthermore select at least three vertices on each ∂R_i, including all contact points,

and connect them to z_1, z_2, z_3 with edges inside $Z - R_i$ and to c_i with edges inside R_i (thus (3) is satisfied). Clearly (4) is satisfied. Since all contact points were chosen on each ∂R_i both (5) and (6) are satisfied, because we know that if two edges intersect then they intersect in an intersection point of their boundaries. Finally we can choose in/out witnesses for disjoint(R_i, R_j), inside(R_i, R_j), meet(R_i, R_j), covered(R_i, R_j), and overlap(R_i, R_j). Note that we chose at most $|I|^2$ witnesses and at most $|I|^4$ in/out witnesses. Hence (G, R, S) has size polynomial in $|\varphi|$. □

5. The theory of diagrams

We have shown that the existential theory of diagrams is decidable by reducing it to the existential theory of strings. In this section we show that the first-order theory of diagrams is undecidable.

Theorem 5.1. *The first-order theory of diagrams is Δ^1_ω-complete.*

Δ^1_ω is the ωth level of the analytical hierarchy, the level which captures the complexity of deciding truth in second-order arithmetic (see [Odi89] for details). Our proof proceeds in three steps, interpreting the first-order theory of strings in the first-order theory of diagrams, interpreting the first-order theory of diagrams in second-order arithmetic, and finally coding second-order arithmetic into the first-order theory of strings, showing that all three theories have the same complexity.

Proposition 5.2. *The first-order theory of strings can be computably interpreted in the first-order theory of diagrams. More precisely for every Σ_k formula about strings we can compute an equivalent Σ_{k+2} formula about diagrams.*

Proof. We model a curve s as a pair of regions (A, B) such that $A \cap B = s$. To this end we define predicates over diagrams, curve$(A, B) \in \Sigma_2$ which is true if and only if $A \cap B$ is a curve, and intersect$((A, B), (C, D)) \in \Sigma_2$ which is true if the two curves represented by (A, B) and (C, D) intersect. Assuming we have these predicates we can easily translate each sentence about curves into an equivalent sentence about diagrams (see for example the proof of the Reduction Theorem [Hod93]). We can also arrange the quantifiers in such a way that the translated sentence lies in Σ_{k+2}.

For the definition of curve and intersect we make use of a predicate union(C, A, B) which is true if and only if $C \subseteq A \cup B$.

$$
\begin{aligned}
\text{union}(C, A, B) \quad &\Leftrightarrow \quad C \subseteq A \cup B \\
&:\Leftrightarrow \quad (\forall D)[\text{contains}(C, D) \Rightarrow \neg(\text{disjoint}(D, A) \wedge \text{disjoint}(D, B)] \\
\text{curve}(A, B) \quad &\Leftrightarrow \quad A \cap B \text{ is a curve} \\
&:\Leftrightarrow \quad \text{meet}(A, B) \wedge \\
&\qquad (\exists C)[\text{cover}(C, A) \wedge \text{cover}(C, B) \wedge \text{union}(C, A, B)] \\
\text{intersect}((A, B), (C, D)) \quad &\Leftrightarrow \quad \text{the curves } A \cap B \text{ and } C \cap D \text{ intersect} \\
&:\Leftrightarrow \quad \text{curve}(A, B) \wedge \text{curve}(C, D) \wedge \\
&\qquad (\exists E, F)[\text{cover}(A, E) \wedge (\text{cover}(C, E) \vee \text{cover}(D, E)) \wedge \\
&\qquad (\text{cover}(A, F) \vee \text{cover}(B, F)) \wedge \text{cover}(D, F) \wedge \text{meet}(E, F)]
\end{aligned}
$$

In the verification that these predicates model the properties they are supposed to, we need the compactness of the regions. The predicate curve(A, B) expresses that there is a region C which is the union of the two meeting regions A and B. Since A and B are simply connected and do not intersect this implies that they meet in a curve. To see that intersect$((A, B), (C, D))$ expresses intersection of two curves note that E and F as in the predicate have to meet in a point that belongs to all four regions. □

Proposition 5.3. *The theory of diagrams can be interpreted in second-order arithmetic.*

Proof. We will only sketch the basic ideas needed to interpret diagrams in second-order arithmetic. A region is homeomorphic to the unit disk, hence it can be described by a countable set of real numbers dense in it. In second-order arithmetic we can talk about real numbers, and countable sets of real numbers. With this we can write a basic predicate that tests whether a real number is contained in the closure of a countable set of real numbers. We can also test whether a real number is contained within the interior of the closure of a countable set of real numbers, by asking for all reals in a neighborhood of the real number to be contained in the closure. With these two basic predicates we can express the eight predicates that express the possible relationships between regions as predicates over countable sets of reals. □

Proposition 5.4. *Second-order arithmetic can be interpreted in the theory of strings.*

Proof. We begin by defining several predicates from the basic predicate intersect(x, y). The idea is to model points in the plane as the unique intersection point of two curves.

$$
\begin{aligned}
a \subseteq b \quad &\Leftrightarrow \quad a \text{ is a subset of } b \\
&:\Leftrightarrow \quad (\forall x)[\text{intersect}(x, a) \Rightarrow \text{intersect}(x, b)] \\
\text{unique}(a, b) \quad &\Leftrightarrow \quad a \text{ and } b \text{ have a unique intersection point} \\
&:\Leftrightarrow \quad \text{intersect}(a, b) \wedge \\
&\qquad (\forall x \subseteq a)(\forall y \subseteq b)[\text{intersect}(x, b) \wedge \text{intersect}(y, a) \Rightarrow \text{intersect}(x, y)] \\
\text{cross}(a, b, c) \quad &\Leftrightarrow \quad a \text{ and } b \text{ have a unique intersection point, and it lies on } c \\
&:\Leftrightarrow \quad \text{unique}(a, b) \wedge (\forall x)[x \subseteq a \wedge \text{intersect}(x, b) \Rightarrow \text{intersect}(x, c)] \\
\text{intpt}(a, b, c, d) \quad &\Leftrightarrow \quad a \text{ and } b \text{ have a unique intersection point, and it lies on } c \text{ and } d \\
&:\Leftrightarrow \quad \text{cross}(a, b, c) \wedge \text{cross}(a, b, d) \\
\text{union}(a, b, c, d) \quad &\Leftrightarrow \quad a = b \cup c \cup d \text{ and } b \cap d = \emptyset \\
&:\Leftrightarrow \quad b, c, d \subseteq a \wedge \neg \text{intersect}(b, d) \wedge \\
&\qquad (\forall x)[\text{intersect}(x, a) \Rightarrow \text{intersect}(x, b) \vee \text{intersect}(x, c) \vee \text{intersect}(x, d)] \\
\text{interior}(a, b) \quad &\Leftrightarrow \quad a \text{ and } b \text{ only have interior intersection points} \\
&:\Leftrightarrow \quad \text{union}(a, a_1, a_2, a_3) \wedge \text{unique}(a_1, a_2) \wedge \text{unique}(a_2, a_3) \wedge \\
&\qquad \text{union}(b, b_1, b_2, b_3) \wedge \text{unique}(b_1, b_2) \wedge \text{unique}(b_2, b_3) \wedge \\
&\qquad \bigwedge_{i,j \in \{1,3\}} \neg \text{intersect}(a_i, b_j)
\end{aligned}
$$

$$
\begin{aligned}
\text{finint}(a,b) &\Leftrightarrow a\cap b \text{ is a finite set}\\
&:\Leftrightarrow (\forall x,y)[\text{intpt}(x,y,a,b)\Rightarrow(\exists a'\subseteq a)(\exists b'\subseteq b)[\text{intpt}(a',b',x,y)\wedge \text{interior}(a',b')]]\\
\text{allinton}(a,b,c) &\Leftrightarrow a\cap b\subseteq c\\
&:\Leftrightarrow (\forall u\subseteq a)[\text{intersect}(u,b)\Rightarrow \text{intersect}(u,c)]\\
\text{sameint}(a,b,c) &\Leftrightarrow a\cap b = a\cap c\\
&:\Leftrightarrow \text{allinton}(a,b,c)\wedge \text{allinton}(a,c,b)
\end{aligned}
$$

The compactness of the curves involved is necessary to guarantee the correctness of most of these definitions. The main predicate is $\text{finint}(a,b)$ which expresses that a and b have finitely many intersection points only. This is verified by requiring each intersection point to be contained within an open neighborhood on each curve (the interiors of a' and b'). Since a and b are compact, there can only be finitely many such points.

We also need a family of formulas that we will later use to express cardinality. Note that the length of these formulas depends on k, hence we cannot use them in the definition of finint.

$$
\begin{aligned}
\text{unique}_k(a,b) &\Leftrightarrow |a\cap b| = k\\
&:\Leftrightarrow (\exists a_1,\ldots,a_k\subseteq a)[\bigwedge_{1\leqslant i\leqslant k}\text{unique}(a_i,b)\wedge \bigwedge_{1\leqslant i<j\leqslant k}\neg\,\text{intersect}(a_i,a_j)\wedge\\
&\quad (\forall x\subseteq a)[\text{intersect}(x,b)\Rightarrow[\bigvee_{1\leqslant i\leqslant k}\text{intersect}(x,a_i)]]\\
\text{less}(a,b,c,d) &\Leftrightarrow |a\cap b|<|c\cap d|<\infty \text{ (assuming } a\cap c=\emptyset)\\
&:\Leftrightarrow (\exists x)[\text{finint}(x,a)\wedge \text{finint}(x,c)\wedge \text{sameint}(a,b,x)\wedge \text{sameint}(c,d,x)\wedge\\
&\quad (\forall u\subseteq x)[\text{unique}(u,a)\Rightarrow(\exists v)[u\subseteq v\subseteq x\wedge \text{unique}_2(v,c)\wedge \text{unique}(v,a)]]]
\end{aligned}
$$

If $\text{less}(a,b,c,d)$ and a and c are disjoint, then the number of intersection points of a and b is strictly less than the number of intersection points of c and d (and both are finite). This allows us to define equipollence.

$$
\begin{aligned}
\text{eqint}(a,b,c,d) &\Leftrightarrow |a\cap b| = |c\cap d|<\infty \text{ (assuming } a\cap c=\emptyset)\\
&:\Leftrightarrow \neg\,\text{less}(a,b,c,d)\wedge\neg\,\text{less}(c,d,a,b)
\end{aligned}
$$

For the following let us fix the strings $U=[0,1]$ and $N=\{(x,x\sin(1/x))\}\cup\{(0,0)\}$ which will be used as parameters. Furthermore choose pairwise disjoint sets $N_i\subseteq N$ such that N_i and U intersect in exactly i points (for all $i\in\mathbb{N}$).

We need to show how to translate unnested atomic statements of second-order arithmetic into the theory of strings. For each number variable x reserve a string variable s_x; we require $s_x\subseteq N$. For set variables X reserve string variables s_X. We require $\text{allinton}(U,s_X,N)$. Translate as follows:

$$
\begin{aligned}
i=x \quad &\text{as} \quad (\exists u,v)[\text{finint}(u,v)\wedge \text{eqint}(N_i,U,u,v)\wedge \text{eqint}(u,v,s_x,U)]\\
x+y=z \quad &\text{as} \quad (\exists u,v,w_1,w_2,w_3)[\text{finint}(u,v)\wedge \text{union}(v,w_1,w_2,w_3)\wedge\neg\,\text{intersect}(w_2,u)\wedge\\
&\qquad \text{eqint}(s_x,U,u,w_1)\wedge \text{eqint}(s_y,U,u,w_3)\wedge \text{eqint}(s_z,U,u,v)\\
i\in X \quad &\text{as} \quad \text{allinton}(N_i,U,s_X)
\end{aligned}
$$

These predicates are sufficient to build all of second-order arithmetic. Hence we have shown that second-order arithmetic can be interpreted in the theory of strings with parameters $U, N, N_1, \ldots$. While we cannot define the pair (U, N) we only need to be able to define a pair of strings that has exactly $\omega + 1$ intersections. This, however, is easy:

$$\begin{aligned}\text{universe}(u,v) \quad \Leftrightarrow \quad & \neg\ \text{finint}(u,v) \wedge (\exists x \subseteq u)[\text{unique}(x,v) \wedge \\ & (\forall y : x \subseteq y \subseteq u)[\neg\ \text{unique}(y,v)] \wedge (\forall y \subseteq u)[\neg\ \text{intersect}(y,x) \Rightarrow\ \text{finint}(y,v)]]\end{aligned}$$

The predicate universe(u, v) expresses that u and v intersect infinitely often in $\omega + 1$ many points. The accumulation point is the unique point in which x and v intersect. Every neighborhood of this point contains intersection points of u and v. Furthermore, there are no other accumulation points since all subcurves of u disjoint from x only have finitely many intersections with v. Curves u and v such as these can serve instead of the particular sets U and N used above, hence we can simply require that universe(U, N).

We then define N_i to be the smallest substring of v that has i intersections with u, is disjoint from $N_0, \ldots, N_{i-1}$, and such that there are no intersection points of u and v between N_{i-1} and N_i. □

6. Concluding remarks

While it is satisfying to know that string graphs can be effectively, if not efficiently, recognized, the gap between **NP** and **NEXP** is large, and a more precise classification is called for. As we mentioned earlier, the problem is settled in a paper by Schaefer et al. which shows that string graphs can be recognized in **NP** [SSS01]. By Lemma 4.3 this implies that the topological inference problem can also be decided in **NP**.

Kratochvíl [Kra98] suggested a different approach to obtaining an exponential upper bound. He conjectured that in any smallest weak realization of a (G, R) any edge which is crossed at least once is crossed exactly once by some other edge. He showed that his conjecture implies $c_w(m) \leqslant m(2^{m-1} - 1)/2$. The status of the conjecture remains open (a counterexample presented in an earlier version of this article turned out to be faulty).

Is the string graph recognition problem decidable on surfaces of higher genus? Our proof essentially relies on the inversion operation which will not be available to us (at least not in the straightforward manner we used it) in surfaces other than the 2-sphere. In [SSS01] we show that this problem can be solved using very recent results on trace monoids.

Related to the string graph problem is the problem of computing the crossing number of a graph (the smallest number of intersections necessary to draw the graph in the plane). This problem has long been known to be **NP**-complete. Martin Grohe [Gro01] recently showed it to be solvable in time $O(f(k)n^2)$, where k is the number of intersections, and $f(k) = O(2^{2^{p(k)}})$, for some polynomial p, implying that it is fixed parameter tractable, since for fixed k the complexity is quadratic. We obtain an interesting variant of the crossing number problem by asking for the smallest number of *pairs* of edges that need to intersect to draw the graph in the plane (where each such pair can intersect any number of times). Call this the *crossing pairs number* of a graph G [PT00]. Our proof technique implies that if there is a drawing of G in which at most k pairs of edges intersect, then there is a drawing of G with at most $2k2^{2k}$ intersections. We can then use

Grohe's result to conclude that the crossing pairs number of a graph G is fixed parameter tractable. With the techniques from [SSS01] we can also show that the problem lies in **NP**, however, we do not know whether the problem is **NP**-hard.

Acknowledgments

We thank Mikkel Thorup for helpful suggestions.

References

[All83] J.F. Allen, Maintaining knowledge about temporal intervals, Comm. ACM 26 (11) (1983) 832–843.

[Ben59] S. Benzer, On the topology of the genetic fine structure, Proc. Natl. Acad. Sci. 45 (1959) 1607–1620.

[CGP98] Z.-Z. Chen, M. Grigni, C. Papadimitriou, Planar map graphs, in: Proceedings of the 30th Annual ACM Symposium on Theory of Computing (STOC-98), ACM Press, New York, 1998, pp. 514–523.

[dFPP90] H. de Fraysseix, J. Pach, R. Pollack, How to draw a planar graph on a grid, Combinatorica 10 (1990) 41–51.

[EET76] G. Ehrlich, S. Even, R.E. Tarjan, Intersection graphs of curves in the plane, J. Combin. Theory 21 (1976) 8–20.

[Ege91] M.J. Egenhofer, Reasoning about binary topological relations, Lecture Notes Comput. Sci. 525 (1991) 143–160.

[FH02] J. Flower, J. Howse, Generating euler diagrams, in: Second International Conference on Theory and Application of Diagrams, Lecture Notes in Artificial Intelligence, Vol. 2317, Springer, Berlin, 2002, pp. 61–75.

[GJ83] M.R. Garey, D.S. Johnson, Crossing number is NP-complete, SIAM J. Algebraic Discrete Methods 4 (3) (1983) 312–316.

[Gra76] R.L. Graham, Problem 1, in: Open Problems at 5th Hungarian Colloquium on Combinatorics, Keszthely, Hungary, 1976.

[GPP95] M. Grigni, D. Papadias, C. Papadimitriou, Topological inference, in: Proceedings of the 14th International Joint Conference on Artificial Intelligence, Montreal, 1995, pp. 901–905.

[Gro01] M. Grohe, Computing crossing numbers in quadratic time, in: Proceedings of the 32nd ACM Symposium on Theory of Computing, Heraklion, Greece, July 6–8 2001, pp. 231–236.

[Hod93] W.A. Hodges, Model Theory, Cambridge University Press, Cambridge, 1993.

[Kra91a] J. Kratochvíl, String graphs. I. The number of critical nonstring graphs is infinite, J. Combin. Theory Series B 52 (1991) 53–66.

[Kra91b] J. Kratochvíl, String graphs. II. Recognizing string graphs is NP-hard, J. Combin. Theory Series B 52 (1991) 67–78.

[Kra98] J. Kratochvíl, Crossing number of abstract topological graphs, Lecture Notes Comput. Sci. 1547 (1998) 238–245.

[KM91] J. Kratochvíl, J. Matoušek, String graphs requiring exponential representations, J. Combin. Theory Series B 53 (1991) 1–4.

[KGK86] J. Kratochvíl, M. Goljan, P. Kučera, String graphs, Rozpravy Československ Akad. Věd Řada Mat. Přírod. Věd. 3 (96) (1986) 1–96.

[LP97] O.J. Lemon, I. Pratt, Spatial logic and the complexity of diagrammatic reasoning, Mach. Graphics and Vision 6 (1) (1997) 89–108.

[McC01] R.M. McConnell, Linear-time recognition of circular-arc graphs, in: Proceedings of the 42nd Symposium on Foundations of Computer Science, Las Vegas, October 14–17, 2001, pp. 386–394.

[MNT88] J. Matoušek, J. Nešetřil, R. Thomas, On polynomial time decidability of induced-minor-closed classes, Comment. Math. Univ. Carolin. 29 (4) (1988) 703–710.

[MT01] B. Mohar, C. Thomassen, Graphs on Surfaces, Johns Hopkins University Press, Baltimore, 2001.
[Odi89] P. Odifreddi, Classical Recursion Theory, North-Holland, Amsterdam, 1989.
[PT00] J. Pach, G. Tóth, Which crossing number is it anyway?, J. Combin. Theory Series B 80 (2000) 225–246.
[PT01] J. Pach, G. Tóth, Recognizing string graphs is decidable, in: P. Mutzel (Ed.), Graph Drawing 2001, Lecture Notes in Computer Science, Springer, Berlin, 2001, pp. 247–260.
[PSV99] C.H. Papadimitriou, D. Suciu, V. Vianu, Topological queries in spatial data-bases, J. Comput. System Sci. 58 (1999) 29–53.
[SSS01] M. Schaefer, E. Sedgwick, D. Štefankovič, Recognizing string graphs in NP, Technical Report TR01-011, DePaul University, October 2001.
[Sch90] W. Schnyder, Embedding planar graphs on the grid, in: D. Johnson (Ed.), Proceedings of the 1st Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '90), San Francisco, CA, USA, SIAM, Philadelphia, January 1990, pp. 138–148.
[Sin66] F.W. Sinden, Topology of thin film circuits, Bell System Tech. J. 45 (1966) 1639–1662.
[Tho98] M. Thorup, Map graphs in polynomial time, in: IEEE Symposium on Foundations of Computer Science, Palo Alto, CA, 1998, pp. 396–405.