Theory Comput Syst (2017) 61:263–282
DOI 10.1007/s00224-015-9663-z

The Connectivity of Boolean Satisfiability: Dichotomies for Formulas and Circuits

Konrad W. Schwerdtfeger[1]

Published online: 28 October 2015
© Springer Science+Business Media New York 2015

Abstract For Boolean satisfiability problems, the structure of the solution space is characterized by the solution graph, where the vertices are the solutions, and two solutions are connected iff they differ in exactly one variable. In 2006, Gopalan et al. studied connectivity properties of the solution graph and related complexity issues for CSPs. They proved dichotomies for the diameter of connected components and for the complexity of the st-connectivity question, and conjectured a trichotomy for the connectivity question. Recently, we were able to establish the trichotomy. Here, we consider connectivity issues of satisfiability problems defined by Boolean circuits and propositional formulas that use gates, resp. connectives, from a fixed set of Boolean functions. We obtain dichotomies for the diameter and the two connectivity problems: on one side, the diameter is linear in the number of variables, and both problems are in P, while on the other side, the diameter can be exponential, and the problems are PSPACE-complete. For partially quantified formulas, we show an analogous dichotomy. A motivation is the relevance to reconfiguration problems and satisfiability algorithms.

Keywords Boolean satisfiability · Boolean circuits · Post's lattice · PSPACE-completeness · Dichotomy theorems · Graph connectivity

✉ Konrad W. Schwerdtfeger
k.w.s@gmx.net

[1] Institut für Theoretische Informatik, Leibniz Universität Hannover, Hannover, Germany

1 Introduction

The Boolean satisfiability problem (SAT), as well as many related questions like equivalence, counting, enumeration, and numerous versions of optimization, are of great importance in both theory and applications of computer science. In this article, we focus on the solution-space structure: We consider the *solution graph*, where the vertices are the solutions, and two solutions are connected iff they differ in exactly one variable. For this implicitly defined graph, we then study the connectivity and st-connectivity problems, and the diameter of connected components. Figures 1 and 2 give an impression of how solution graphs may look like.

While the standard satisfiability problem is defined for propositional formulas, which can be seen as one special form of descriptions for Boolean relations, satisfiability and related problems have also been considered for many alternative descriptions, e.g. Boolean constraint satisfactions problems (*CSP*s), Boolean circuits, binary decision diagrams, and Boolean neural networks. For the usual formulas with the connectives $\wedge$, $\vee$ and $\neg$, there are several common variants. A special form are formulas in conjunctive normal form (CNF-formulas). A generalization of CNF-formulas are CNF($\mathcal{S}$)-formulas, which are conjunctions of constraints on the variables taken from a finite template set $\mathcal{S}$.

Here we consider another type of generalization: Arbitrarily nested formulas built with connectives from some finite set of Boolean functions B (where the arity may be greater than two), known as $B - formulas$. Also we study B-*circuits*, where analogously the allowed gates implement the functions from B. As a further extension we consider partially quantified B-formulas.

A direct application of st-connectivity in solution graphs are *reconfiguration problems*, that arise when we wish to find a step-by-step transformation between two feasible solutions of a problem, such that all intermediate results are also feasible. Recently, the reconfiguration versions of many problems such as INDEPENDENT-SET, VERTEX-COVER, SET-COVER GRAPH-k-COLORING, SHORTEST-PATH have been studied, and complexity results obtained (see e.g. [12, 13]). Also of relevance are the connectivity properties to the problem of *structure identification*, where one is given a relation explicitly and seeks a short representation of some

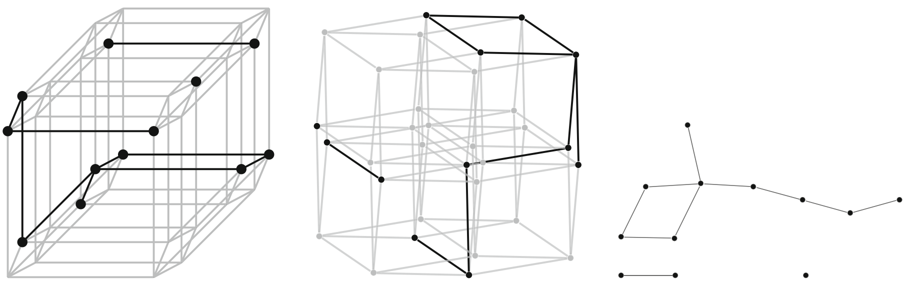

Fig. 1 Depictions of the subgraph of the 5-dimensional hypercube graph induced by a typical random Boolean relation with 12 elements. Left: highlighted on a orthographic hypercube projection. Center: highlighted on a "Spectral Embedding" of the hypercube graph by MATHEMATICA. Right: the sole subgraph, arranged by MATHEMATICA

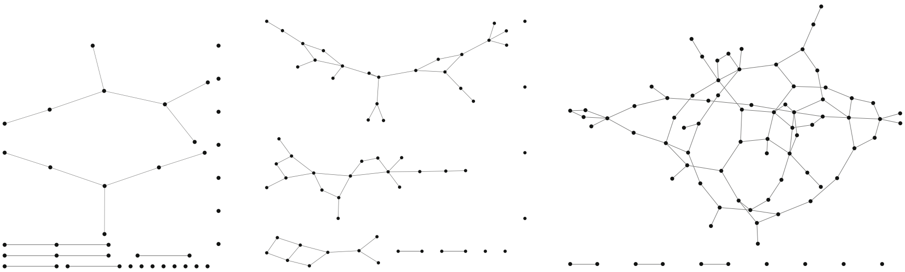

Fig. 2 Subgraphs of the 8-dimensional hypercube graph induced by typical random relations with 40, 60 and 80 elements, arranged by MATHEMATICA

kind (see e.g. [6]); this problem is important especially for learning in artificial intelligence.

A better understanding of the solution space structure also promises advancement of SAT algorithms: It has been discovered that the solution space connectivity is strongly correlated to the performance of standard satisfiability algorithms like WalkSAT and DPLL on random instances: As one approaches the *satisfiability threshold* (the ratio of constraints to variables at which random k-CNF-formulas become unsatisfiable for $k \geq 3$) from below, the solution space (with the connectivity defined as above) fractures, and the performance of the algorithms deteriorates [16, 17]. These insights mainly came from statistical physics, and lead to the development of the *survey propagation algorithm*, which has much better performance on random instances [16].

While current SAT solvers normally accept only CNF-formulas as input, one of the most important applications of satisfiability testing is verification and optimization in Electronic Design Automation (EDA), where the instances derive mostly from digital circuit descriptions [27]. Though many such instances can easily be encoded in CNF, the original structural information, such as signal ordering, gate orientation and logic paths, is lost, or at least obscured. Since exactly this information can be very helpful for solving these instances, considerable effort has been made recently to develop satisfiability solvers that work with the circuit description directly [27], which have far superior performance in EDA applications, or to restore the circuit structure from CNF [9]. This is a major motivation for our study.

Our perspective is mainly from complexity theory: We classify B-formulas and B-circuits by the worst-case complexity of the connectivity problems, analogously to Schaefer's dichotomy theorem for satisfiability of CSPs from 1978 [22], Lewis' dichotomy for satisfiability of B-formulas from 1979 [14], and Gopalan et al.'s classification for the connectivity problems of CSPs from 2006 [10]. Along the way, we will examine structural properties of the solution graph like its maximal diameter, and devise efficient algorithms for solving the connectivity problems.

We begin with a formal definition of some central concepts.

Definition 1 An n-ary *Boolean relation* is a subset of $\{0, 1\}^n$ $(n \geq 1)$. If ϕ is some description of an n-ary Boolean relation R, e.g. a propositional formula (where the variables are taken in lexicographic order), the *solution graph* $G(\phi)$ of ϕ is the subgraph of the n-dimensional hypercube graph induced by the vectors in R, i.e., the vertices of $G(\phi)$ are the vectors in R, and there is an edge between two vectors precisely if they differ in exactly one position.

We use $\boldsymbol{a}, \boldsymbol{b}, \ldots$ to denote vectors of Boolean values and $\boldsymbol{x}, \boldsymbol{y}, \ldots$ to denote vectors of variables, $\boldsymbol{a} = (a_1, a_2, \ldots)$ and $\boldsymbol{x} = (x_1, x_2, \ldots)$.

The *Hamming weight* $|\boldsymbol{a}|$ of a Boolean vector $\boldsymbol{a}$ is the number of 1's in $\boldsymbol{a}$. For two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, the *Hamming distance* $|\boldsymbol{a} - \boldsymbol{b}|$ is is the number of positions in which they differ.

If $\boldsymbol{a}$ and $\boldsymbol{b}$ are solutions of ϕ and lie in the same connected component of $G(\phi)$, we write $d_\phi(\boldsymbol{a}, \boldsymbol{b})$ to denote the shortest-path distance between $\boldsymbol{a}$ and $\boldsymbol{b}$.

The *diameter of a connected component* is the maximal shortest-path distance between any two vectors in that component. The *diameter of* $G(\phi)$ is the maximal diameter of any of its connected components.

2 Connectivity of CNF-Formulas

Research has focused on the structure of the solution space only quite recently: One of the earliest studies on solution-space connectivity was done for CNF($\mathcal{S}$)-formulas with constants (see the definition below), begun in 2006 by Gopalan et al. ([10, 11, 15, 24]).

In our proofs for B-formulas and B-circuits, we will use Gopalan et al.'s results for 3-CNF-formulas, so we have to introduce some related terminology.

Definition 2 A *CNF-formula* is a Boolean formula of the form $C_1 \wedge \cdots \wedge C_m$ $(1 \leq m < \infty)$, where each C_i is a *clause*, that is, a finite disjunction of *literals* (variables or negated variables). A *k-CNF-formula* $(k \geq 1)$ is a CNF-formula where each C_i has at most k literals.

For a finite set of Boolean relations $\mathcal{S}$, a *CNF($\mathcal{S}$)-formula* (with constants) over a set of variables V is a finite conjunction $C_1 \wedge \cdots \wedge C_m$, where each C_i is a *constraint application* (*constraint* for short), i.e., an expression of the form $R(\xi_1, \ldots, \xi_k)$, with a k-ary relation $R \in \mathcal{S}$, and each ξ_j is a variable in V or one of the constants 0, 1.

A *k-clause* is a disjunction of k variables or negated variables. For $0 \leq i \leq k$, let D_i be the set of all satisfying truth assignments of the k-clause whose first i literals are negated, and let $S_k = \{D_0, \ldots, D_k\}$. Thus, CNF($S_k$) is the collection of k-CNF-formulas.

Gopalan et al. studied the following two decision problems for CNF($\mathcal{S}$)-formulas:

- the *connectivity problem* CONN($\mathcal{S}$): given a CNF($\mathcal{S}$)-formula ϕ, is $G(\phi)$ connected? (if ϕ is unsatisfiable, then $G(\phi)$ is considered connected)
- the *st-connectivity problem* ST-CONN($\mathcal{S}$): given a CNF($\mathcal{S}$)-formula ϕ and two solutions $\boldsymbol{s}$ and $\boldsymbol{t}$, is there a path from $\boldsymbol{s}$ to $\boldsymbol{t}$ in $G(\phi)$?

Lemma 1 [10, Lemm 3.6] ST-CONN*(S_3) and* CONN*(S_3) are PSPACE-complete.*

Showing that the problems are in PSPACE is straightforward: Given a CNF(S_3)-formula ϕ and two solutions $\boldsymbol{s}$ and $\boldsymbol{t}$, we can guess a path of length at most 2^n between them and verify that each vertex along the path is indeed a solution. Hence ST-CONN(S_3) is in NPSPACE, which equals PSPACE by Savitch' s theorem. For CONN(S_3), by reusing space we can check for all pairs of vectors whether they are satisfying, and, if they both are, whether they are connected in $G(\phi)$.

The hardness-proof is quite intricate: it consists of a direct reduction from the computation of a space-bounded Turing machine M. The input-string w of M is mapped to a CNF(S_3)-formula ϕ and two satisfying assignments $\boldsymbol{s}$ and $\boldsymbol{t}$, corresponding to the initial and accepting configuration of a Turing machine M' constructed from M and w, s.t. $\boldsymbol{s}$ and $\boldsymbol{t}$ are connected in $G(\phi)$ iff M accepts w. Further, all satisfying assignments of ϕ are connected to either $\boldsymbol{s}$ or $\boldsymbol{t}$, so that $G(\phi)$ is connected iff M accepts w.

Lemma 2 [10, Lemm 3.7] *For $n \geq 2$, there is an n-ary Boolean function f with $f(1, \ldots, 1) = 1$ and a diameter of at least $2^{\lfloor \frac{n}{2} \rfloor}$.*

The proof of this lemma is by direct construction of such a formula.

3 Circuits, Formulas, and Post's Lattice

An n-ary *Boolean function* is a function $f : \{0, 1\}^n \to \{0, 1\}$. Let B be a finite set of Boolean functions.

A *B-circuit* $\mathcal{C}$ with input variables $x_1, \ldots, x_n$ is a directed acyclic graph, augmented as follows: Each node (here also called *gate*) with indegree 0 is labeled with an x_i or a 0-ary function from B, each node with indegree $k > 0$ is labeled with a k-ary function from B. The edges (here also called *wires*) pointing into a gate are ordered. One node is designated the output gate. Given values $a_1, \ldots, a_n \in \{0, 1\}$ to $x_1, \ldots, x_n$, $\mathcal{C}$ computes an n-ary function $f_{\mathcal{C}}$ as follows: A gate v labeled with a variable x_i returns a_i, a gate v labeled with a function f computes the value $f(b_1, \ldots, b_k)$, where $b_1, \ldots, b_k$ are the values computed by the predecessor gates of v, ordered according to the order of the wires. For a more formal definition see [26].

A *B-formula* is defined inductively: A variable x is a B-formula. If $\phi_1, \ldots, \phi_m$ are B-formulas, and f is an n-ary function from B, then $f(\phi_1, \ldots, \phi_n)$ is a B-formula. In turn, any B-formula defines a Boolean function in the obvious way, and we will identify B-formulas and the function they define.

It is easy to see that the functions computable by a B-circuit, as well as the functions definable by a B-formula, are exactly those that can be obtained from B by *superposition*, together with all projections [2]. By superposition, we mean substitution (that is, composition of functions), permutation and identification of variables, and introduction of *fictive variables* (variables on which the value of the function does not depend). This class of functions is denoted by $[B]$. B is *closed* (or said to be a *clone*) if $[B] = B$. A *base* of a clone F is any set B with $[B] = F$.

Already in the early 1920s, Emil Post extensively studied Boolean functions [20]. He identified all clones, found a finite base for each of them, and detected their inclusion structure: The clones form a lattice, called *Post's lattice*, depicted in Fig. 3.

The following clones are defined by properties of the functions they contain, all other ones are intersections of these. Let f be an n-ary Boolean function.

- BF is the class of all Boolean functions.
- R_0 (R_1) is the class of all 0-reproducing (1-reproducing) functions,
  f is *c-reproducing*, if $f(c, \ldots, c) = c$, where $c \in \{0, 1\}$.
- M is is the class of all monotone functions,
  f is *monotone*, if $a_1 \le b_1, \ldots, a_n \le b_n$ implies $f(a_1, \ldots, a_n) \le f(b_1, \ldots, b_n)$.
- D is the class of all self-dual functions,
  f is *self-dual*, if $f(x_1, \ldots, x_n) = \overline{f(\overline{x_1}, \ldots, \overline{x_n})}$.
- L is the class of all affine (on *linear*) functions,
  f is *affine*, if $f(x_1, \ldots, x_n) = x_{i_1} \oplus \cdots \oplus x_{i_m} \oplus c$ with $i_1, \ldots, i_m \in \{1, \ldots, n\}$ and $c \in \{0, 1\}$.
- S_0 (S_1) is the class of all 0-separating (1-separating) functions,
  f is *c-separating*, if there exists an $i \in \{1, \ldots, n\}$ s.t. $a_i = c$ for all $\boldsymbol{a} \in f^{-1}(c)$, where $c \in \{0, 1\}$.
- S_0^m $\left(\mathsf{S}_1^m\right)$ is the class of all functions that are 0-separating (1-separating) of degree m,
  f is *c-separating of degree m*, if for all $U \subseteq f^{-1}(c)$ of size $|U| = m$ there exists an $i \in \{1, \ldots, n\}$ s.t. $a_i = c$ for all $\boldsymbol{a} \in U$ ($c \in \{0, 1\}, m \ge 2$).

The definitions and bases of all classes are given in Table 1. For an introduction to Post's lattice and further references see e.g. [2].

The complexity of numerous problems for B-circuits and B-formulas has been classified by the types of functions allowed in B with help of Post's lattice (see e.g. [21, 23]), starting with satisfiability: Analogously to Schaefer's dichotomy for CNF($\mathcal{S}$)-formulasfrom 1978*,* Harry R. Lewis shortly thereafter found a dichotomy for B-formulas [14]: If $[B]$ contains the function $x \wedge \overline{y}$, Sat is NP-complete, else it is in P.

While for B-circuits the complexity of every decision problem solely depends on $[B]$ (up to AC^0 isomorphisms), for B-formulas this need not be the case (though it usually is, as for satisfiability and our connectivity problems, as we will see): The transformation of a B-formula into a B'-formula might require an exponential increase in the formula size even if $[B] = [B']$, as the B'-representation of some function from B may need to use some input variable more than once [18]. For example, let $h(x, y) = x \wedge \overline{y}$; then (x ∧ y) $\in [\{h\}]$ since $x \wedge y = h(x, h(x, y))$, but it is easy to see that there is no shorter $\{h\}$-representation of $x \wedge y$.

4 Computational and Structural Dichotomies for Connectivity

Now we consider the connectivity problems for B-formulas and B-circuits:

- $\mathsf{BF\text{-}Conn}(B)$: Given a B-formula ϕ, is $G(\phi)$ connected?

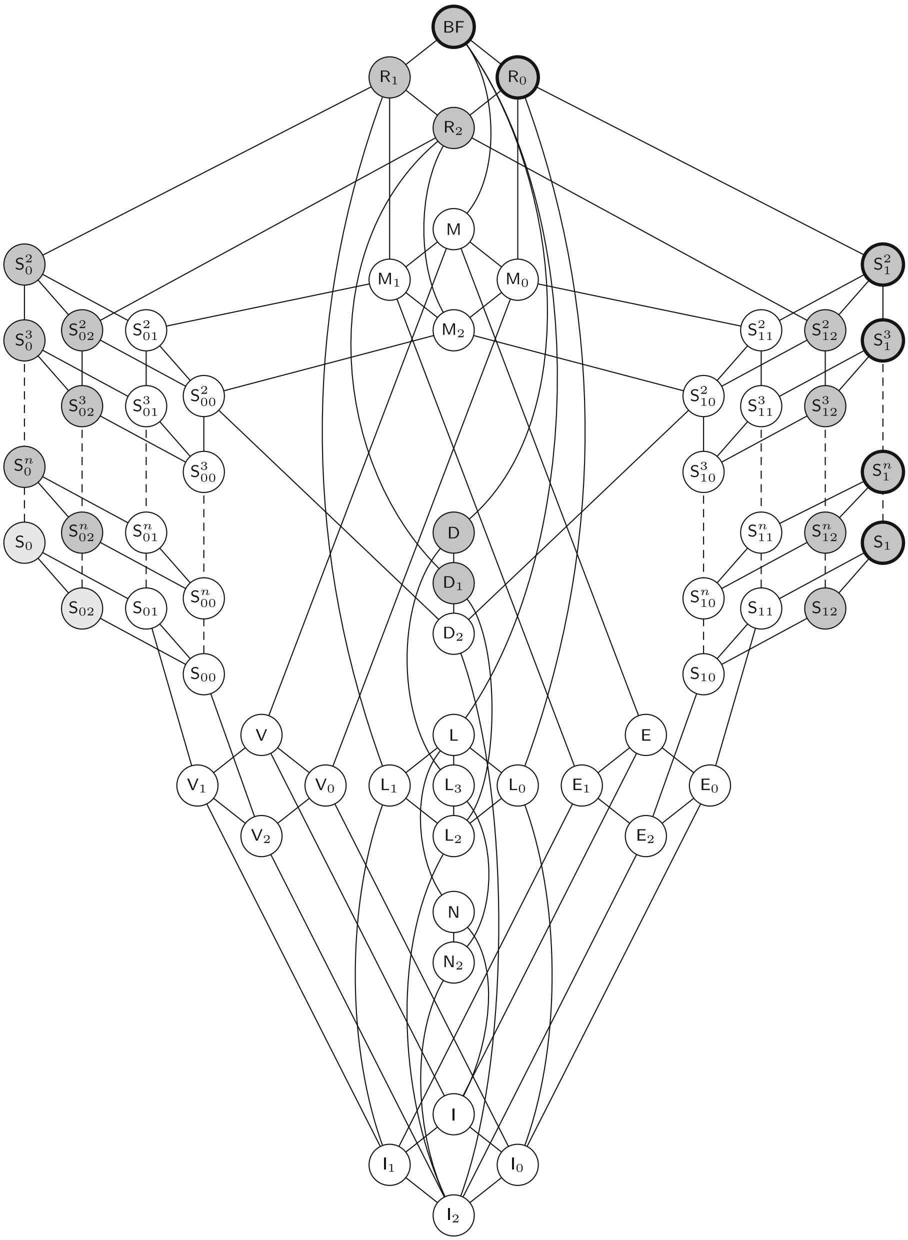

Fig. 3 Graphical representation of Post's lattice. The classes on the hard side of the dichotomy for the connectivity problems and the diameter are shaded gray; the light gray shaded ones are only on the hard side for formulas with quantifiers. For comparison, the classes for which SAT (without quantifiers) is NP-complete are circled bold

Table 1 List of all closed classes of Boolean functions with definitions and bases

BF	All Boolean functions	$\{x \wedge y, \neg x\}$
R_0	$\{f \in BF \mid f$ is 0-reproducing$\}$	$\{x \wedge y, x \oplus y\}$
R_1	$\{f \in BF \mid f$ is 1-reproducing$\}$	$\{x \vee y, x \leftrightarrow y\}$
R_2	$R_0 \cap R_1$	$\{x \vee y, x \wedge (y \leftrightarrow z)\}$
M	$\{f \in BF \mid f$ is monotone$\}$	$\{x \wedge y, x \vee y, 0, 1\}$
M_0	$M \cap R_0$	$\{x \wedge y, x \vee y, 0\}$
M_1	$M \cap R_1$	$\{x \wedge y, x \vee y, 1\}$
M_2	$M \cap R_2$	$\{x \wedge y, x \vee y\}$
S_0	$\{f \in BF \mid f$ is 0-separating$\}$	$\{x \rightarrow y\}$
S_0^n	$\{f \in BF \mid f$ is 0-separating of degree n$\}$	$\left\{x \rightarrow y, \text{dual}\left(T_n^{n+1}\right)\right\}$
S_1	$\{f \in BF \mid f$ is 1-separating$\}$	$\{x \nrightarrow y\}$
S_1^n	$\{f \in BF \mid f$ is 1-separating of degree n$\}$	$\left\{x \nrightarrow y, T_n^{n+1}\right\}$
S_{02}^n	$S_0^n \cap R_2$	$\left\{x \vee (y \wedge \neg z), \text{dual}\left(T_n^{n+1}\right)\right\}$
S_{02}	$S_0 \cap R_2$	$\{x \vee (y \wedge \neg z)\}$
S_{01}^n	$S_0^n \cap M$	$\left\{\text{dual}\left(T_n^{n+1}\right), 1\right\}$
S_{01}	$S_0 \cap M$	$\{x \vee (y \wedge z), 1\}$
S_{00}^n	$S0^n \cap R_2 \cap M$	$\left\{x \vee (y \wedge z), \text{dual}\left(T_n^{n+1}\right)\right\}$
S00	$S0 \cap R_2 \cap M$	$\{x \vee (y \wedge z)\}$
$S12^n$	$S1^n \cap R_2$	$\left\{x \wedge (y \vee \neg z), T_n^{n+1}\right\}$
S12	$S1 \cap R_2$	$\{x \wedge (y \vee \neg z)\}$
$S11^n$	$S1^n \cap M$	$\{T_n^{n+1}, 0\}$
S11	$S1 \cap M$	$\{x \wedge (y \vee z), 0\}$
$S10^n$	$S1^n \cap R_2 \cap M$	$\left\{x \wedge (y \vee z), T_n^{n+1}\right\}$
S10	$S1 \cap R_2 \cap M$	$\{x \wedge (y \vee z)\}$
D	$\{f \in BF \mid f$ is self-dual$\}$	$\{\text{maj}(x, \neg y, \neg z)\}$
D_1	$D \cap R_2$	$\{\text{maj}(x, y, \neg z)\}$
D_2	$D \cap M$	$\{\text{maj}(x, y, z)\}$
L	$\{f \in BF \mid f$ is linear$\}$	$\{x \oplus y, 1\}$
L_0	$L \cap R_0$	$\{x \oplus y\}$
L_1	$L \cap R_1$	$\{x \leftrightarrow y\}$
L_2	$L \cap R_2$	$\{x \oplus y \oplus z\}$
L_3	$L \cap D$	$\{x \oplus y \oplus z \oplus 1\}$
E	$\{f \in BF \mid f$ is constant or a conjunction$\}$	$\{x \wedge y, 0, 1\}$
E_0	$E \cap R_0$	$\{x \wedge y, 0\}$
E_1	$E \cap R_1$	$\{x \wedge y, 1\}$
E_2	$E \cap R_2$	$\{x \wedge y\}$
V	$\{f \in BF \mid f$ is constant or a disjunction$\}$	$\{x \vee y, 0, 1\}$
V_0	$V \cap R_0$	$\{x \vee y, 0\}$
V_1	$V \cap R_1$	$\{x \vee y, 1\}$
V_2	$V \cap R_2$	$\{x \vee y\}$
N	$\{f \in BF \mid f$ is essentially unary$\}$	$\{\neg x, 0, 1\}$
N_2	$N \cap D$	$\{\neg x\}$

Table 1 (continued)

I	$\{f \in \mathsf{BF} \mid f$ is constant or a projection$\}$	$\{x, 0, 1\}$
I_0	$\mathsf{I} \cap \mathsf{R}_0$	$\{x, 0\}$
I_1	$\mathsf{I} \cap \mathsf{R}_1$	$\{x, 1\}$
I_2	$\mathsf{I} \cap \mathsf{R}_2$	$\{x\}$

(T_k^n denotes the threshold function, $T_k^n(x_1, \ldots, x_n) = 1 \iff \sum_{i=1}^n x_i \geq k$, and dual$(f)(x_1, \ldots, x_n) = \overline{f(\overline{x_1}, \ldots, \overline{x_n})}$)

– st-BF-Conn(B): Given a B-formula ϕ and two solutions $\boldsymbol{s}$ and $\boldsymbol{t}$, is there a path from $\boldsymbol{s}$ to $\boldsymbol{t}$ in $G(\phi)$?

The corresponding problems for circuits are denoted Circ-Conn(B) resp. ST-CIRC-CONN(B).

Theorem 1 *Let B be a finite set of Boolean functions.*

1. *If $B \subseteq \mathsf{M}$, $B \subseteq \mathsf{L}$, or $B \subseteq \mathsf{S}_0$, then*
   - *(a)* ST-CIRC-CONN*(B) and* CIRC-CONN*(B) are in* P*,*
   - *(b)* ST-BF-CONN*(B) and* BF-CONN*(B) are in* P*,*
   - *(c) the diameter of every function $f \in [B]$ is linear in the number of variables of f.*
2. *Otherwise,*
   - (a) ST-CIRC-CONN*(B) and* CIRC-CONN*(B) are* PSPACE*-complete,*
   - (b) ST-BF-CONN*(B) and* BF-CONN*(B) are* PSPACE*-complete,*
   - (c) *there are functions $f \in [B]$ such that their diameter is exponential in the number of variables of f.*

The proof follows from the Lemmas in the next subsections. By the following proposition, we can relate the complexity of B-formulas and B-circuits.

Proposition 1 *Every B-formula ϕ can be transformed into an equivalent B-circuit $\mathcal{C}$ in polynomial time.*

Proof Any B-formula is equivalent to a special B-circuit where all function-gates have outdegree at most one: For every variable x of ϕ and for every occurrence of a function f in ϕ there is a gate in $\mathcal{C}$, labeled with x resp. f. It is clear how to connect the gates. □

4.1 The Easy Side of the Dichotomy

Lemma 3 *If $B \subseteq \mathsf{M}$, the solution graph of any n-ary function $f \in [B]$ is connected, and $d_f(\boldsymbol{a}, \boldsymbol{b}) = |\boldsymbol{a} - \boldsymbol{b}| \leq n$ for any two solutions $\boldsymbol{a}$ and $\boldsymbol{b}$.*

Proof Table 1 shows that f is monotone in this case. Thus, either $f = 0$, or $(1, \ldots, 1)$ must be a solution, and every other solution $\boldsymbol{a}$ is connected to $(1, \ldots, 1)$ in $G(\phi)$ since $(1, \ldots, 1)$ can be reached by flipping the variables assigned 0 in $\boldsymbol{a}$ one at a time to 1. Further, if $\boldsymbol{a}$ and $\boldsymbol{b}$ are solutions, $\boldsymbol{b}$ can be reached from $\boldsymbol{a}$ in $|\boldsymbol{a} - \boldsymbol{b}|$ steps by first flipping all variables that are assigned 0 in $\boldsymbol{a}$ and 1 in $\boldsymbol{b}$, and then flipping all variables that are assigned 1 in $\boldsymbol{a}$ and 0 in $\boldsymbol{b}$. □

Lemma 4 *If $B \subseteq \mathsf{S}_0$, the solution graph of any function $f \in [B]$ is connected, and $d_f(\boldsymbol{a}, \boldsymbol{b}) \leq |\boldsymbol{a} - \boldsymbol{b}| + 2$ for any two solutions $\boldsymbol{a}$ and $\boldsymbol{b}$.*

Proof Since f is 0-separating, there is an i such that $a_i = 0$ for every vector $\boldsymbol{a}$ with $f(\boldsymbol{a}) = 0$, thus every $\boldsymbol{b}$ with $b_i = 1$ is a solution. It follows that every solution $\boldsymbol{t}$ can be reached from any solution $\boldsymbol{s}$ in at most $|\boldsymbol{s} - \boldsymbol{t}| + 2$ steps by first flipping the i-th variable from 0 to 1 if necessary, then flipping all other variables in which $\boldsymbol{s}$ and $\boldsymbol{t}$ differ, and finally flipping back the i-th variable if necessary. □

Lemma 5 *If $B \subseteq \mathsf{L}$,*

1. ST-CIRC-CONN*(B) and* CIRC-CONN*(B) are in P,*
2. ST-BF-CONN*(B) and* BF-CONN*(B) are in P,*
3. *for any function $f \in [B]$, $d_f(\boldsymbol{a}, \boldsymbol{b}) = |\boldsymbol{a} - \boldsymbol{b}|$ for any two solutions $\boldsymbol{a}$ and $\boldsymbol{b}$ that lie in the same connected component of $G(\phi)$.*

Proof Since every function $f \in \mathsf{L}$ is linear, $f(x_1, \ldots, x_n) = x_{i_1} \oplus \ldots \oplus x_{i_m} \oplus c$, and any two solutions $\boldsymbol{s}$ and $\boldsymbol{t}$ are connected iff they differ only in fictive variables: If $\boldsymbol{s}$ and $\boldsymbol{t}$ differ in at least one non-fictive variable (i.e., an $x_i \in \{x_{i_1}, \ldots, x_{i_m}\}$), to reach $\mathbf{t}$ from $\mathbf{s}$, x_i must be flipped eventually, but for every solution $\boldsymbol{a}$, any vector $\boldsymbol{b}$ that differs from $\boldsymbol{a}$ in exactly one non-fictive variable is no solution. If $\boldsymbol{s}$ and $\boldsymbol{t}$ differ only in fictive variables, $\boldsymbol{t}$ can be reached from $\boldsymbol{s}$ in $|\boldsymbol{s} - \boldsymbol{t}|$ steps by flipping one by one the variables in which they differ.

Since $\{x \oplus y, 1\}$ is a base of L, every B-circuit $\mathcal{C}$ can be transformed in polynomial time into an equivalent $\{x \oplus y, 1\}$-circuit $\mathcal{C}'$ by replacing each gate of $\mathcal{C}$ with an equivalent $\{x \oplus y, 1\}$-circuit. Now one can decide in polynomial time whether a variable x_i is fictive by checking for $\mathcal{C}'$ whether the number of "backward paths" from the output gate to gates labeled with x_i is odd, so ST-CIRC-CONN(B) is in P.

$G(\mathcal{C})$ is connected iff at most one variable is non-fictive, thus CIRC-CONN(B) is in P.

By Proposition 1, ST-BF-CONN(B) and BF-CONN(B) are in P also. □

This completes the proof of the easy side of the dichotomy.

4.2 The Hard Side of the Dichotomy

Proposition 2 st-Circ-Conn*(B) and* Circ-Conn*(B), as well as* st-BF-Conn*(B) and* BF-Conn*(B), are in* PSPACE *for any finite set B of Boolean functions.*

Proof This follows as in Lemma 3.6 of [10] (see Lemma 1). □

An inspection of Post's lattice shows that if $B \nsubseteq \mathsf{M}$, $B \nsubseteq \mathsf{L}$, and $B \nsubseteq \mathsf{S}_0$, then $[B] \supseteq \mathsf{S}_{12}$, $[B] \supseteq \mathsf{D}_1$, or $[B] \supseteq \mathsf{S}_{02}^k\ \forall k \geq 2$, so we have to prove PSPACE-completeness and show the existence of B-formulas with an exponential diameter in these cases.

In the proofs, we will use the following notation: We write $\boldsymbol{x} = \boldsymbol{c}$ or $\boldsymbol{x} = c_1 \cdots c_n$ for $(x_1 = c_1) \wedge \cdots \wedge (x_n = c_n)$, where $\boldsymbol{c} = (c_1, \ldots, c_n)$ is a vector of constants; e.g., $\boldsymbol{x} = \boldsymbol{0}$ means $\overline{x}_1 \wedge \cdots \wedge \overline{x}_n$, and $\boldsymbol{x} = 101$ means $x_1 \wedge \overline{x}_2 \wedge x_3$. Further, we use $\boldsymbol{x} \in \{\boldsymbol{a}, \boldsymbol{b}, \ldots\}$ for $(\boldsymbol{x} = \boldsymbol{a}) \vee (\boldsymbol{x} = \boldsymbol{b}) \vee \ldots$. Also, we write $\psi(\overline{\boldsymbol{x}})$ for $\psi(\overline{x}_1, \ldots, \overline{x}_n)$. If we have two vectors of Boolean values $\boldsymbol{a}$ and $\boldsymbol{b}$ of length n and m resp., we write $\boldsymbol{a} \cdot \boldsymbol{b}$ for their concatenation $(a_1, \ldots, a_n, b_1, \ldots b_m)$.

All hardness proofs are by reductions from the problems for 1-reproducing 3-CNF-formulas, which are PSPACE-complete by the following proposition.

Proposition 3 *For 1-reproducing 3-CNF-formulas, the problems* st-Conn *and* Conn *are* PSPACE*-complete.*

Proof In the PSPACE-hardness proof for CNF(S_3)-formulas (Lemma 3.6 of [10], see Lemma 1), two satisfying assignments $\boldsymbol{s}$ and $\boldsymbol{t}$ to the constructed formula ϕ are known, so we can construct a connectivity-equivalent 1-reproducing 3-CNF-formula ψ, e.g. as $\psi(\boldsymbol{x}) = \phi(x_1 \oplus s_1 \oplus 1, \ldots, x_n \oplus s_n \oplus 1)$, and then check connectivity for ψ instead of ϕ. □

Lemma 6 *If* $[B] \supseteq \mathsf{S}_{12}$*,*

1. st-BF-Conn*(B) and* BF-Conn*(B) are* PSPACE*-complete,*
2. st-Circ-Conn*(B) and* Circ-Conn*(B) are* PSPACE*-complete,*
3. *for* $n \geq 3$*, there is an* n*-ary function* $f \in [B]$ *with diameter of at least* $2^{\left\lfloor \frac{n-1}{2} \right\rfloor}$.

Proof 1. We reduce the problems for 1-reproducing 3-CNF-formulas to the ones for B-formulas: We map a 1-reproducing 3-CNF-formula ϕ and two solutions $\boldsymbol{s}$ and $\boldsymbol{t}$ of ϕ to a B-formula ϕ' and two solutions $\boldsymbol{s'}$ and $\boldsymbol{t'}$ of ϕ' such that $\boldsymbol{s'}$ and $\boldsymbol{t'}$ are connected in $G(\phi')$ iff $\boldsymbol{s}$ and $\boldsymbol{t}$ are connected in $G(\phi)$, and such that $G(\phi')$ is connected iff $G(\phi)$ is connected.

While the construction of ϕ' is quite easy for this lemma, the construction for the next two lemmas is analogous but more intricate, so we proceed carefully in two steps, which we will adapt in the next two proofs: In the first step, we give a transformation T that transforms any 1-reproducing formula ψ into a connectivity-equivalent formula $T_\psi \in \mathsf{S}_{12}$ built from the standard connectives. Since $\mathsf{S}_{12} \subseteq [B]$, we can express T_ψ as a B-formula T^*_ψ. Now if we would apply T to ϕ directly, we would know that T_ϕ can be expressed as a B-formula. However, this could lead to an exponential increase in the formula size (see Section 3), so we have to show how to construct the B-formula in polynomial time. For this, in the second step, we construct a B-formula ϕ' directly from ϕ (by applying T to the clauses and the $\wedge$'s individually), and then show that ϕ' is equivalent to T_ϕ; thus we know that ϕ' is connectivity-equivalent to ϕ.

Step 1. From Table 1, we find that $\mathsf{S}_{12} = \mathsf{S}_1 \cap \mathsf{R}_2 = \mathsf{S}_1 \cap \mathsf{R}_0 \cap \mathsf{R}_1$, so we have to make sure that T_ψ is 1-seperating, 0-reproducing, and 1-reproducing. Let

$$T_\psi = \psi \wedge y,$$

where y is a new variable.

All solutions $\boldsymbol{a}$ of $T_\psi(\boldsymbol{x}, y)$ have $a_{n+1} = 1$, so T_ψ is 1-seperating and 0-reproducing; also, T_ψ is still 1-reproducing. Further, for any two solutions $\boldsymbol{s}$ and $\boldsymbol{t}$ of $\psi(\boldsymbol{x})$, $\boldsymbol{s}' = \boldsymbol{s} \cdot 1$ and $\boldsymbol{t}' = \boldsymbol{t} \cdot 1$ are solutions of $T_\psi(\boldsymbol{x}, y)$, and it is easy to see that they are connected in $G(T_\psi)$ iff $\boldsymbol{s}$ and $\boldsymbol{t}$ are connected in $G(\psi)$, and that $G(T_\psi)$ is connected iff $G(\psi)$ is connected.

Step 2. The idea is to parenthesize the conjunctions of ϕ such that we get a tree of $\wedge$'s of depth logarithmic in the size of ϕ, and then to replace each clause and each $\wedge$ with an equivalent B-formula. This can increase the formula size by only a polynomial in the original size even if the B-formula equivalent to $\wedge$ uses some input variable more than once.

Let $\phi = C_1 \wedge \cdots \wedge C_n$ be a 1-reproducing 3-CNF-formula. Since ϕ is 1-reproducing, every clause C_i of ϕ is itself 1-reproducing, and we can express T_{C_i} through a B-formula $T^*_{C_i}$. Also, we can express $T_{u\wedge v}$ through a B-formula $T^*_{u\wedge v}$ since $\wedge$ is 1-reproducing; we write $T_\wedge(\psi_1, \psi_2)$ for the formula obtained from $T_{u\wedge v}$ by substituting the formula ψ_1 for u and ψ_2 for v, and similarly write $T^*_\wedge(\psi_1, \psi_2)$ for the formula obtained from $T^*_{u\wedge v}$ in this way. We let $\phi' = \mathsf{Tr}(\phi)$, where Tr is the following recursive algorithm that takes a CNF-formula as input:

Algorithm $\mathsf{Tr}(\psi_1 \wedge \cdots \wedge \psi_m)$
If $m = 1$, return $T^*_{\psi_1}$.
Else return $T^*_\wedge\left(\mathsf{Tr}(\psi_1 \wedge \cdots \wedge \psi_{\lfloor m/2 \rfloor}), \mathsf{Tr}(\psi_{\lfloor m/2 \rfloor+1} \wedge \cdots \wedge \psi_m)\right)$.

Since the recursion terminates after a number of steps logarithmic in the number of clauses of ϕ, and every step increases the total formula size by only a constant factor, the algorithm runs in polynomial time. We show $\phi' \equiv T_\phi$ by induction on m. For

$m = 1$ this is clear. For the induction step, we have to show $T_\wedge^*(T_{\psi_1}, T_{\psi_2}) \equiv T_{\psi_1 \wedge \psi_2}$, but since $T_\wedge(\psi_1, \psi_2) \equiv T_\wedge^*(\psi_1, \psi_2)$, it suffices to show that $T_\wedge(T_{\psi_1}, T_{\psi_2}) \equiv T_{\psi_1 \wedge \psi_2}$:

$$T_\wedge(T_{\psi_1}, T_{\psi_2}) = (\psi_1 \wedge y) \wedge (\psi_2 \wedge y) \wedge y \equiv \psi_1 \wedge \psi_2 \wedge y = T_{\psi_1 \wedge \psi_2}.$$

2. This follows from 1. by Proposition 1.

3. By Lemma 2, there is an 1-reproducing $(n-1)$-ary function f with diameter of at least $2^{\left\lfloor \frac{n-1}{2} \right\rfloor}$. Let f be represented by a formula ϕ; then, T_ϕ represents an n-ary function of the same diameter in S_{12}. □

Lemma 7 *If* $[B] \supseteq \mathsf{D}_1$,

1. ST-BF-CONN*(B) and* BF-CONN*(B) are* PSPACE*-complete,*
2. ST-CIRC-CONN*(B) and* CIRC-CONN*(B) are* PSPACE*-complete,*
3. *for $n \geq 5$, there is an n-ary function $f \in [B]$ with diameter of at least* $2^{\left\lfloor \frac{n-3}{2} \right\rfloor}$.

Proof 1. As noted, we adapt the two steps from the previous proof.

Step 1. Since $\mathsf{D}_1 = \mathsf{D} \cap \mathsf{R}_0 \cap \mathsf{R}_1$, T_ψ must be self-dual, 0-reproducing, and 1-reproducing. For clarity, we first construct an intermediate formula $T_\psi^\sim \in \mathsf{D}_1$ whose solution graph has an additional component, then we eliminate that component.

For $\psi(\boldsymbol{x})$, let

$$T_\psi^\sim = (\psi(\boldsymbol{x}) \wedge (\boldsymbol{y} = \mathbf{1})) \vee \left(\overline{\psi(\overline{\boldsymbol{x}})} \wedge (\boldsymbol{y} = \mathbf{0})\right) \vee (\boldsymbol{y} \in \{100, 010, 001\}),$$

where $\boldsymbol{y} = (y_1, y_2, y_3)$ are three new variables.

$T_\psi^\sim$ is self-dual: for any solution ending with 111 (satisfying the first disjunct), the inverse vector is no solution; similarly, for any solution ending with 000 (satisfying the second disjunct), the inverse vector is no solution; finally, all vectors ending with 100, 010, or 001 are solutions and their inverses are no solutions. Also, $T_\psi^\sim$ is still 1-reproducing, and it is 0-reproducing (for the second disjunct note that $\overline{\psi(\overline{0 \cdots 0})} \equiv \overline{\psi(1 \cdots 1)} \equiv 0$).

Further, every solution $\boldsymbol{a}$ of ψ corresponds to a solution $\boldsymbol{a} \cdot 111$ of $T_\psi^\sim$, and for any two solutions $\boldsymbol{s}$ and $\boldsymbol{t}$ of ψ, $\boldsymbol{s}' = \boldsymbol{s} \cdot 111$ and $\boldsymbol{t}' = \boldsymbol{t} \cdot 111$ are connected in $G\left(T_\psi^\sim\right)$ iff $\boldsymbol{s}$ and $\boldsymbol{t}$ are connected in $G(\psi)$: The "if" is clear, for the "only if" note that since there are no solutions of $T_\psi^\sim$ ending with 110, 101, or 011, every solution of $T_\psi^\sim$ not ending with 111 differs in at least two variables from the solutions that do.

Observe that exactly one connected component is added in $G\left(T_\psi^\sim\right)$ to the components corresponding to those of $G(\psi)$: It consists of all solutions ending with 000, 100, 010, or 001 (any two vectors ending with 000 are connected e.g. via those ending with 100). It follows that $G\left(T_\psi^\sim\right)$ is always unconnected. To fix this, we modify $T_\psi^\sim$ to T_ψ by adding $1 \cdots 1 \cdot 110$ as a solution, thereby connecting $1 \cdots 1 \cdot 111$ (which is always a solution since $T_\psi^\sim$ is 1-reproducing) with $1 \cdots 1 \cdot 100$, and thereby with the additional component of T_ψ. To keep the function self-dual, we must in turn remove $0 \cdots 0 \cdot 001$, which does not alter the connectivity. Formally,

$$\begin{aligned} T_\psi &= \left(T_\psi^\sim \vee ((\boldsymbol{x} = \mathbf{1}) \wedge (\boldsymbol{y} = 110))\right) \wedge \neg\, ((\boldsymbol{x} = \mathbf{0}) \wedge (\boldsymbol{y} = 001)) \qquad (1)\\ &= (\psi(\boldsymbol{x}) \wedge (\boldsymbol{y} = \mathbf{1})) \vee \left(\overline{\psi(\overline{\boldsymbol{x}})} \wedge (\boldsymbol{y} = \mathbf{0})\right)\\ &\quad \vee (\boldsymbol{y} \in \{100, 010, 001\} \wedge \neg((\boldsymbol{x} = \mathbf{0}) \wedge (\boldsymbol{y} = 001)))\\ &\quad \vee((\boldsymbol{x} = \mathbf{1}) \wedge (\boldsymbol{y} = 110)). \end{aligned}$$

Now $G(T_\psi)$ is connected iff $G(\psi)$ is connected.

Step 2. Again, we use the algorithm Tr from the previous proof to transform any 1-reproducing 3-CNF-formula ϕ into a B-formula ϕ' equivalent to T_ϕ, but with the definition (1) of T (Fig. 4). Again, we have to show $T_\wedge(T_{\psi_1}, T_{\psi_2}) \equiv T_{\psi_1 \wedge \psi_2}$. Here,

$$\begin{aligned} T_\wedge(T_{\psi_1}, T_{\psi_2}) = &\left(T_{\psi_1} \wedge T_{\psi_2} \wedge (\boldsymbol{y} = \mathbf{1})\right) \vee \left(\overline{\overline{T_{\psi_1}} \wedge \overline{T_{\psi_2}}} \wedge (\boldsymbol{y} = \mathbf{0})\right)\\ &\vee \left(\boldsymbol{y} \in \{100, 010, 001\} \wedge \neg \left(\overline{T_{\psi_1}} \wedge \overline{T_{\psi_2}} \wedge (\boldsymbol{y} = 001)\right)\right)\\ &\vee \left(T_{\psi_1} \wedge T_{\psi_2} \wedge (\boldsymbol{y} = 110)\right). \end{aligned}$$

We consider the parts of the formula in turn: For any formula ξ we have $T_\xi(\boldsymbol{x}_\xi) \wedge (\boldsymbol{y} = \mathbf{1}) \equiv \xi(\boldsymbol{x}_\xi) \wedge (\boldsymbol{y} = \mathbf{1})$ and $T_\xi(\boldsymbol{x}_\xi) \wedge (\boldsymbol{y} = \mathbf{0}) \equiv \overline{\psi(\overline{\boldsymbol{x}_\xi})} \wedge (\boldsymbol{y} = \mathbf{0})$, where $\boldsymbol{x}_\xi$ denotes the variables of ξ. Using $\overline{\overline{T_{\psi_1}(\boldsymbol{x}_{\psi_1})} \wedge \overline{T_{\psi_2}(\boldsymbol{x}_{\psi_2})}} \wedge (\boldsymbol{y} = \mathbf{0}) = \left(T_{\psi_1}(\boldsymbol{x}_{\psi_1}) \vee T_{\psi_2}(\boldsymbol{x}_{\psi_2})\right) \wedge (\boldsymbol{y} = \mathbf{0})$, the first line becomes

$$\left(\psi_1(\boldsymbol{x}_{\psi_1}) \wedge \psi_2(\boldsymbol{x}_{\psi_2}) \wedge (\boldsymbol{y} = \mathbf{1})\right) \vee \left(\left(\overline{\psi_1(\overline{\boldsymbol{x}_{\psi_1}}) \wedge \psi_2(\overline{\boldsymbol{x}_{\psi_2}})}\right) \wedge (\boldsymbol{y} = \mathbf{0})\right).$$

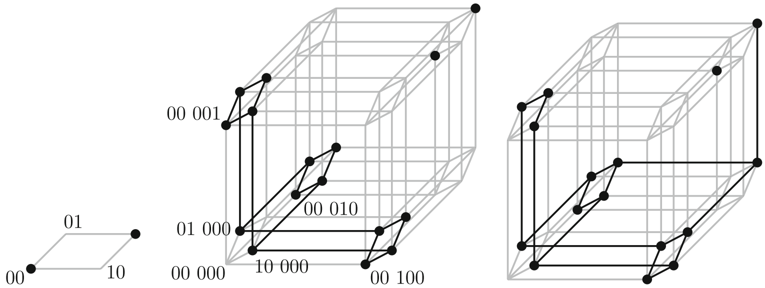

Fig. 4 An example for the transformation. Left: $\psi = (x_1 \vee \overline{x_2}) \wedge (\overline{x_1} \vee x_2)$, center: $T_\psi^\sim$, right: T_ψ. The "axis vertices" are labeled in the first two graphs

For the second line, we observe

$$\begin{aligned}\overline{T_\psi(\boldsymbol{x}_\psi)} \equiv & \left(\overline{\psi(\boldsymbol{x}_\psi)} \vee \neg(\boldsymbol{y}=\mathbf{1})\right) \wedge \left(\psi(\overline{\boldsymbol{x}}_{\boldsymbol{\psi}}) \vee \neg(\boldsymbol{y}=\mathbf{0})\right)\\ & \wedge \left(\boldsymbol{y} \notin \{100,010,001\} \vee \left((\boldsymbol{x}_\psi=\mathbf{0}) \wedge (\boldsymbol{y}=001)\right)\right)\\ & \wedge(\neg(\boldsymbol{x}_\psi=\mathbf{1}) \vee \overline{(\boldsymbol{y}=110)}),\end{aligned}$$

thus $\overline{T_\psi(\boldsymbol{x}_\psi)} \wedge (\boldsymbol{y}=001) \equiv (\boldsymbol{x}_\psi=\mathbf{0}) \wedge (\boldsymbol{y}=001)$, and the second line becomes

$$\vee \left(\boldsymbol{y} \in \{100,010,001\} \wedge \neg\left((\boldsymbol{x}_{\psi_1}=\mathbf{0}) \wedge (\boldsymbol{x}_{\psi_2}=\mathbf{0}) \wedge (\boldsymbol{y}=001)\right)\right).$$

Since $T_\psi(\boldsymbol{x}_\psi) \wedge (\boldsymbol{y}=110) \equiv (\boldsymbol{x}_\psi=\mathbf{1}) \wedge (\boldsymbol{y}=110)$ for any ψ, the third line becomes

$$\vee \left((\boldsymbol{x}_{\psi_1}=\mathbf{1}) \wedge (\boldsymbol{x}_{\psi_2}=\mathbf{1}) \wedge (\boldsymbol{y}=110)\right).$$

Now $T_\wedge\left(T_{\psi_1}, T_{\psi_2}\right)$ equals

$$\begin{aligned}T_{\psi_1\wedge\psi_2} = & \left(\psi_1(\boldsymbol{x}_{\psi_1}) \wedge \psi_2(\boldsymbol{x}_{\psi_2}) \wedge (\boldsymbol{y}=\mathbf{1})\right) \vee \left(\overline{\psi_1(\overline{\boldsymbol{x}_{\psi_1}}) \wedge \psi_2(\overline{\boldsymbol{x}_{\psi_2}})} \wedge (\boldsymbol{y}=\mathbf{0})\right)\\ & \vee \left(\boldsymbol{y} \in \{100,010,001\} \wedge \neg\left((\boldsymbol{x}_{\psi_1}=\mathbf{0}) \wedge (\boldsymbol{x}_{\psi_2}=\mathbf{0}) \wedge (\boldsymbol{y}=001)\right)\right)\\ & \vee \left((\boldsymbol{x}_{\psi_1}=\mathbf{1}) \wedge (\boldsymbol{x}_{\psi_2}=\mathbf{1}) \wedge (\boldsymbol{y}=110)\right).\end{aligned}$$

2. This follows from 1. by Proposition 1.

3. By Lemma 2 there is an 1-reproducing $(n-3)$-ary function f with diameter of at least $2^{\left\lfloor\frac{n-3}{2}\right\rfloor}$. Let f be represented by a formula ϕ; then, T_ϕ represents an n-ary function of the same diameter in D_1. □

Lemma 8 *If* $[B] \supseteq \mathsf{S}_{02}^k$ *for any* $k \geq 2$,

1. ST-BF-CONN*(B) and* BF-CONN*(B) are* PSPACE*-complete,*
2. ST-CIRC-CONN*(B) and* CIRC-CONN*(B) are* PSPACE*-complete,*
3. *for* $n \geq k+4$, *there is an n-ary function* $f \in [B]$ *with diameter of at least* $2^{\left\lfloor\frac{n-k-2}{2}\right\rfloor}$.

Proof 1. *Step 1.* Since $\mathsf{S}_{02}^k = \mathsf{S}_0^k \cap \mathsf{R}_0 \cap \mathsf{R}_1$, T_ψ must be 0-separating of degree k, 0-reproducing, and 1-reproducing. As in the previous proof, we construct an intermediate formula $T_\psi^\sim$. For $\psi(\boldsymbol{x})$, let

$$T_\psi^\sim = (\psi \wedge y \wedge (\boldsymbol{z}=\mathbf{0})) \vee (|\boldsymbol{z}|>1),$$

where y and $\boldsymbol{z}=(z_1,\ldots,z_{k+1})$ are new variables.

$T_\psi^\sim(\boldsymbol{x}, y, \boldsymbol{z})$ is 0-separating of degree k, since all vectors that are no solutions of $T_\psi^\sim$ have $|\boldsymbol{z}| \leq 1$, i.e. $\boldsymbol{z} \in \{0\cdots0, 10\cdots0, 010\cdots0, \ldots, 0\cdots01\} \subset \{0,1\}^{k+1}$, and thus any k of them have at least one common variable assigned 0. Also, $T_\psi^\sim$ is 0-reproducing and still 1-reproducing.

Further, for any two solutions $\boldsymbol{s}$ and $\boldsymbol{t}$ of $\psi(\boldsymbol{x})$, $\boldsymbol{s}'=\boldsymbol{s}\cdot1\cdot0\cdots0$ and $\boldsymbol{t}'=\boldsymbol{t}\cdot1\cdot0\cdots0$ are solutions of $T_\psi^\sim(\boldsymbol{x}, y, \boldsymbol{z})$ and are connected in $G\left(T_\psi^\sim\right)$ iff $\boldsymbol{s}$ and $\boldsymbol{t}$ are connected in $G(\psi)$.

But again, we have produced an additional connected component (consisting of all solutions with $|z| > 1$). To connect it to a component corresponding to one of ψ, we add $1 \cdots 1 \cdot 1 \cdot 10 \cdots 0$ as a solution,

$$T_\psi = (\psi \wedge y \wedge (z = \mathbf{0})) \vee (|z| > 1) \vee ((\boldsymbol{x} = \mathbf{1}) \wedge y \wedge (z = 10 \cdots 0)).$$

Now $G(T_\psi)$ is connected iff $G(\psi)$ is connected.

Step 2. Again we show that the algorithm Tr works in this case. Here,

$$\begin{aligned} T_\wedge(T_{\psi_1}, T_{\psi_2}) = &\left(T_{\psi_1}(\boldsymbol{x}_{\psi_1}) \wedge T_{\psi_2}(\boldsymbol{x}_{\psi_2}) \wedge y \wedge (z = \mathbf{0})\right) \vee (|z| > 1) \\ &\vee \left(T_{\psi_1}(\boldsymbol{x}_{\psi_1}) \wedge T_{\psi_2}(\boldsymbol{x}_{\psi_2}) \wedge y \wedge (z = 10 \cdots 0)\right). \end{aligned}$$

Since $T_\psi(\boldsymbol{x}_\psi) \wedge y \wedge (z = \mathbf{0}) \equiv \psi(\boldsymbol{x}_\psi) \wedge y \wedge (z = \mathbf{0})$ and $T_\psi(\boldsymbol{x}_\psi) \wedge y \wedge (z = 10 \cdots 0) \equiv (\boldsymbol{x}_\psi = 1) \wedge y \wedge (z = 10 \cdots 0)$ for any ψ, this is equivalent to

$$\begin{aligned} T_{\psi_1 \wedge \psi_2} = &\left(\psi_1(\boldsymbol{x}_{\psi_1}) \wedge \psi_2(\boldsymbol{x}_{\psi_2}) \wedge y \wedge (z = \mathbf{0})\right) \vee (|z| > 1) \\ &\vee \left(\boldsymbol{x}_{\psi_1} \wedge \boldsymbol{x}_{\psi_2} \wedge y \wedge (z = 10 \cdots 0)\right). \end{aligned}$$

2. This follows from 1. by Proposition 1.

3. By Lemma 2 there is an 1-reproducing $(n-k-2)$-ary function f with diameter of at least $2^{\left\lfloor \frac{n-k-2}{2} \right\rfloor}$. Let f be represented by a formula ϕ; then, T_ϕ represents an n-ary function of the same diameter in S_{02}^k. □

This completes the proof of Theorem 1.

5 The Connectivity of Quantified Formulas

Definition 3 A *quantified B-formula* ϕ (in prenex normal form) is an expression of the form

$$Q_1 y_1 \cdots Q_m y_m \varphi(y_1, \ldots, y_m, x_1, \ldots, x_n),$$

where φ is a B-formula, and $Q_1, \ldots, Q_m \in \{\exists, \forall\}$ are quantifiers. The solution graph $G(\phi)$ only involves the free variables$x_1, \ldots, x_n$.

For quantified B-formulas, we define the connectivity problems

- QBF-CONN(B): Given a quantified B-formula ϕ, is $G(\phi)$ connected?
- ST-QBF-CONN(B): Given a quantified B-formula ϕ and two solutions $\boldsymbol{s}$ and $\boldsymbol{t}$, is there a path from $\boldsymbol{s}$ to $\boldsymbol{t}$ in $G(\phi)$?

Theorem 2 *Let B be a finite set of Boolean functions.*

1. *If $B \subseteq \mathsf{M}$ or $B \subseteq \mathsf{L}$, then*
   - *(a)* ST-QBF-CONN*(B) and* QBF-CONN*(B) are in P,*
   - *(b) the diameter of every quantified B-formula is linear in the number of free variables.*

2. *Otherwise,*

   (a) ST-QBF-CONN*(B) and* QBF-CONN*(B) are* PSPACE*-complete,*
   (b) there are quantified B-formulas with at most one quantifier such that their diameter is exponential in the number of free variables.

Proof 1. For $B \subseteq \mathsf{M}$, any quantified B-formula ϕ represents a monotone function: Using $\exists y\psi(y,\boldsymbol{x}) = \psi(0,\boldsymbol{x}) \vee \psi(1,\boldsymbol{x})$ and $\forall y\psi(y,\boldsymbol{x}) = \psi(0,\boldsymbol{x}) \wedge \psi(1,\boldsymbol{x})$ recursively, we can transform ϕ into an equivalent M-formula since $\wedge$ and $\vee$ are monotone. Thus as in Lemma 3, ST-QBF-CONN(B) and QBF-CONN(B) are trivial, and $d_f(\boldsymbol{a},\boldsymbol{b}) = |\boldsymbol{a}-\boldsymbol{b}|$ for any two solutions $\boldsymbol{a}$ and $\boldsymbol{b}$.

For a quantified B-formula $\phi = Q_1y_1 \cdots Q_my_m\varphi$ with $B \subseteq \mathsf{L}$, we first remove the quantifications over all fictive variables of φ (and eliminate the fictive variables if necessary). If quantifiers remain, ϕ is either tautological (if the rightmost quantifier is $\exists$) or unsatisfiable (if the rightmost quantifier is $\forall$), so the problems are trivial, and $d_f(\boldsymbol{a},\boldsymbol{b}) = |\boldsymbol{a}-\boldsymbol{b}|$ for any two solutions $\boldsymbol{a}$ and $\boldsymbol{b}$. Otherwise, we have a quantifier-free formula and the statements follow from Lemma 5.

2. Again as in Lemma 1, it follows that ST-QBF-CONN(B) and QBF-CONN(B) are in PSPACE, since the evaluation problem for quantified B-formulas is in PSPACE [22].

An inspection of Post's lattice shows that if $B \not\subseteq \mathsf{M}$ and $B \not\subseteq \mathsf{L}$, then $[B] \supseteq \mathsf{S}_{12}$, $[B] \supseteq \mathsf{D}_1$, or $[B] \supseteq \mathsf{S}_{02}$, so we have to prove PSPACE-completeness and show the existence of B-formulas with an exponential diameter in these cases.

For $[B] \supseteq \mathsf{S}_{12}$ and $[B] \supseteq \mathsf{D}_1$, the statements for the PSPACE-hardness and the diameter obviously carry over from Theorem 1.

For $B \supseteq \mathsf{S}_{02}$, we give a reduction from the problems for (unquantified) 3-CNF-formulas; we proceeded again similar as in the proof of Lemma 6. We give a transformation T_ψ s.t. $T_\psi \in \mathsf{S}_{02}$ for all formulas ψ. Since $\mathsf{S}_{02} = \mathsf{S}_0 \cap \mathsf{R}_0 \cap \mathsf{R}_1$, T_ψ must be self-dual, 0-reproducing, and 1-reproducing. For $\psi(\boldsymbol{x})$ let

$$T_\psi = (\psi \wedge y) \vee z,$$

with the two new variables y and z.

T_ψ is 0-separating since all vectors that are no solutions have $z = 0$. Also, T_ψ is 0-reproducing and 1-reproducing. Again, we use the algorithm Tr from the proof of Lemma 6 to transform any 3-CNF-formula ϕ into a B-formula φ' equivalent to T_ϕ. Again, we show

$$\begin{aligned} T_\wedge(T_{\psi_1}, T_{\psi_2}) &= (((\psi_1 \wedge y) \vee z) \wedge ((\psi_2 \wedge y) \vee z) \wedge y) \vee z \\ &\equiv ((\psi_1 \wedge y) \wedge (\psi_2 \wedge y) \wedge y) \vee z \\ &\equiv (\psi_1 \wedge \psi_2 \wedge y) \vee z = T_{\psi_1 \wedge \psi_2}. \end{aligned}$$

Now let

$$\phi' = \forall z \varphi'.$$

Then, for any two solutions $\boldsymbol{s}$ and $\boldsymbol{t}$ of $\phi(\boldsymbol{x})$, $\boldsymbol{s}' = \boldsymbol{s} \cdot 1$ and $\boldsymbol{t}' = \boldsymbol{t} \cdot 1$ are solutions of $\phi'(\boldsymbol{x}, y)$, and they are connected in $G(\phi')$ iff $\boldsymbol{s}$ and $\boldsymbol{t}$ are connected in $G(\phi)$, and $G(\phi')$ is connected iff $G(\phi)$ is connected.

The proof of Lemma 2 shows that there is an $(n-1)$-ary function f with diameter of at least $2^{\left\lfloor \frac{n-1}{2} \right\rfloor}$. Let f be represented by a formula ϕ; then ϕ' as defined above is a quantified B-formula with n free variables and one quantifier with the same diameter. □

Remark 1 An analog to Theorem 2 also holds for quantified circuits as defined in [21, Section 7].

6 Future Directions

While for st-connectivity and connectivity of B-formulas and B-circuits we now have a quite complete picture, there is a multitude of interesting variations in different directions with open problems.

As mentioned in the abstract, for CNF($\mathcal{S}$)-formulas with constants, we have a complete classification for both connectivity problems and the diameter also [24]. However, for CNF($\mathcal{S}$)-formulas without constants, the complexity of the connectivity problem is still open in some cases [25].

Besides CNF($\mathcal{S}$)-formulas, B-formulas and B-circuits, there are further variants of Boolean satisfiability, and investigating connectivity in these settings might be worthwhile as well. For example, disjunctive normal forms with special connectivity properties were studied by Ekin et al. already in 1997 for their "important role in problems appearing in various areas including in particular discrete optimization, machine learning, automated reasoning, etc." [7].

Other connectivity-related problems already mentioned by Gopalan et al. are counting the number of components and approximating the diameter. Recently, Mouawad et al. investigated the question of finding the shortest path between two solutions [19], which is of special interest to reconfiguration problems.

Furthermore, our definition of connectivity is not the only sensible one: One could regard two solutions connected whenever their Hamming distance is at most d, for any fixed $d \geq 1$; this was already considered related to random satisfiability, see [1]. This generalization seems meaningful as well as challenging.

Finally, a most interesting subject are CSPs over larger domains; in 1993, Feder and Vardi conjectured a dichotomy for the satisfiability problem over arbitrary finite domains [8], and while the conjecture was proved for domains of size three in 2002 by Bulatov [4], it remains open to date for the general case. Close investigation of the solution space might lead to valuable insights here.

For k-colorability, which is a special case of the general CSP over a k-element set, the connectivity problems and the diameter were already studied by Bonsma and Cereceda [3], and Cereceda, van den Heuvel, and Johnson [5]. They showed that for $k = 3$ the diameter is at most quadratic in the number of vertices and the st-connectivity problem is in P, while for $k \geq 4$, the diameter can be exponential and st-connectivity is PSPACE-complete in general.

References

1. Achlioptas, D., Ricci-Tersenghi, F.: On the solution-space geometry of random constraint satisfaction problems. In: Proceedings of the Thirty-Eighth Annual ACM Symposium on Theory of Computing, pp. 130–139. ACM (2006)
2. Böhler, E., Creignou, N., Reith, S., Vollmer, H.: Playing with boolean blocks, part i: Posts lattice with applications to complexity theory. In: SIGACT News (2003)
3. Bonsma, P., Cereceda, L.: Finding paths between graph colourings: Pspace-completeness and superpolynomial distances. Theor. Comput. Sci. **410**(50), 5215–5226 (2009)
4. Bulatov, A.A.: A dichotomy theorem for constraints on a three-element set. In: Proceedings of The 43rd Annual IEEE Symposium on Foundations of Computer Science, 2002, pp. 649–658. IEEE (2002)
5. Cereceda, L., Van den Heuvel, J., Johnson, M.: Finding paths between 3-colorings. J. Graph Theory **67**(1), 69–82 (2011)
6. Creignou, N., Kolaitis, P., Zanuttini, B.: Structure identification of boolean relations and plain bases for co-clones. J. Comput. Syst. Sci. **74**(7), 1103–1115 (2008)
7. Ekin, O., Hammer, P.L., Kogan, A.: On connected boolean functions. Discr. Appl. Math. **96**, 337–362 (1999)
8. Feder, T., Vardi, M.Y.: The computational structure of monotone monadic snp and constraint satisfaction: A study through datalog and group theory. SIAM J. Comput. **28** (1), 57–104 (1998)
9. Fu, Z., Malik, S.: Extracting logic circuit structure from conjunctive normal form descriptions. In: 20th International Conference on VLSI Design, 2007. Held Jointly with 6th International Conference on Embedded Systems, pp. 37–42. IEEE (2007)
10. Gopalan, P., Kolaitis, P.G., Maneva, E., Papadimitriou, C.H.: The connectivity of boolean satisfiability: Computational and structural dichotomies. SIAM J. Comput. **38**(6), 2330–2355 (2009). doi:10.1137/07070440X
11. Gopalan, P., Kolaitis, P.G., Maneva, E.N., Papadimitriou, C.H.: The connectivity of boolean satisfiability: Computational and structural dichotomies. ICALP'06, 346–357 (2006). doi:10.1007/11786986_31
12. Ito, T., Demaine, E.D., Harvey, N.J.A., Papadimitriou, C.H., Sideri, M., Uehara, R., Uno, Y.: On the complexity of reconfiguration problems. Theor. Comput. Sci. **412**(12–14), 1054–1065 (2011). doi:10.1016/j.tcs.2010.12.005
13. Kamiński, M., Medvedev, P., Milanič, M.: Shortest paths between shortest paths and independent sets. In: Combinatorial Algorithms, pp. 56–67. Springer (2011)
14. Lewis, H.R.: Satisfiability problems for propositional calculi. Mathematical Systems Theory **13**(1), 45–53 (1979)
15. Makino, K., Tamaki, S., Yamamoto, M.: On the boolean connectivity problem for horn relations. In: Proceedings of the 10th International Conference on Theory and Applications of Satisfiability Testing, SAT'07, pp. 187–200 (2007)
16. Maneva, E., Mossel, E., Wainwright, M.J.: A new look at survey propagation and its generalizations. J. ACM (JACM) **54**(4), 17 (2007)
17. Mézard, M., Mora, T., Zecchina, R.: Clustering of solutions in the random satisfiability problem. Phys. Rev. Lett. **94**(19), 197,205 (2005)
18. Michael, T.: On the applicability of post's lattice. Inf. Process. Lett. **112**(10), 386–391 (2012)
19. Mouawad, A.E., Nishimura, N., Pathak, V., Raman, V.: Shortest reconfiguration paths in the solution space of Boolean formulas. arXiv:1404.3801 (2014)
20. Post, E.L.: The Two-Valued Iterative Systems of Mathematical Logic.(AM-5), vol. 5. Princeton University Press (1941)
21. Reith, S., Wagner, K.W.: The complexity of problems defined by Boolean circuits (2000)
22. Schaefer, T.J.: The complexity of satisfiability problems. STOC '78, 216–226 (1978). doi:10.1145/800133.804350
23. Schnoor, H.: Algebraic techniques for satisfiability problems. Ph.D. thesis, Universität Hannover (2007)
24. Schwerdtfeger, K.W.: A computational trichotomy for connectivity of boolean satisfiability. Journal on Satisfiability, Boolean Modeling and Computation **8**, 173–195 (2013)

25. Schwerdtfeger, K.W.: The connectivity of boolean satisfiability: No-constants and quantified variants. arXiv:1403.6165 (2014)
26. Vollmer, H.: Introduction to Circuit Complexity: A Uniform Approach. Springer, New York (1999)
27. Wu, C.A., Lin, T.H., Lee, C.C., Huang, C.Y.R.: Qutesat: A robust circuit-based sat solver for complex circuit structure. In: Proceedings of the Conference on Design, Automation and Test in Europe, pp. 1313–1318. EDA Consortium (2007)