The Computational Complexity of Monotonicity in Probabilistic Networks

Johan Kwisthout

Department of Information and Computing Sciences, Utrecht University

Technical Report UU-CS-2007-040

www.cs.uu.nl

ISSN: 0924-3275

The Computational Complexity of Monotonicity in Probabilistic Networks

Johan Kwisthout

2007

Abstract

Many computational problems related to probabilistic networks are complete for complexity classes that have few 'real world' complete problems. For example, the decision variant of the inference problem (PR) is PP-complete, the MAP-problem is NP^{PP}-complete and deciding whether a network is monotone in mode or distribution is co-NP^{PP}-complete. We take a closer look at monotonicity; more specific, the computational complexity of determining whether the values of the variables in a probabilistic network can be ordered, such that the network is monotone. We prove that this problem – which is trivially co-NP^{PP}-hard – is co-$\text{NP}^{\text{NP}^{\text{PP}}}$-hard in networks which allow a succinct representation of the conditional probabilities.

1 Introduction

Probabilistic networks [1] (also called Bayesian or belief networks) represent a joint probability distribution on a set of statistical variables. A probabilistic network is often described by a directed acyclic graph and a set of conditional probabilities. The nodes represent the statistical variables, the arcs (or lack of them) represent (in)dependencies induced by the joint probability distribution. Probabilistic networks are often used in decision support systems such as medical diagnosis systems (see e.g. [2] or [3]). Apart from their relevance in practical situations, they are interesting from a theoretical viewpoint as well.

Many problems related to probabilistic networks happen to be complete for complexity classes that have few 'real world' complete problems. For example, the decision variant of the inference problem PR (is the probability of a specific instantiation of a variable greater than p) is PP-complete [4], where the exact inference problem is #P-complete [5]. The problem of finding the most probable explanation (MPE), i.e., the most likely instantiation to all variables, has an NP-complete decision variant [6]. On the other hand, the PARTIAL MAP problem, determining whether there exists an instantiation to a subset of all variables (the so-called MAP variables), such that the maximum aposteriori probability of the other variables is greater than p, is NP^{PP}-complete [7]. Determining whether a network is monotone (in mode or in distribution) is co-NP^{PP}-complete [8].

Monotonicity is often studied in the context of probabilistic classification, where a network is constructed of evidence variables (like observable symptoms and test results), non-observable intermediate variables, and one or more classification variables. Informally, the conditional probability of a variable C given evidence variables E is monotone, if higher ordered instantiations to E always lead to higher values of C (isotone) or always lead to lower values of C (antitone). The question whether these relations are monotone is particularly relevant during the construction and verification of the network. Often, domain experts will declare that certain relations ought to be monotone, and the conditional probabilities in the network should then respect these assumptions. When a violation of monotonicity of a certain relation is found, the encoded probabilities should be reconsidered, by elicitating better estimations or using more data to learn from.

While complexity results are known for the MONOTONICITY problem when all variables have fixed orderings, no such results have been obtained yet for the related problem where no such fixed order is presumed. Nevertheless, while variables sometimes have a trivial ordering (e.g., *always* > *sometimes* > *never*), such an ordering might be arbitrary, and determining a 'good' ordering might reduce the part of the network where monotonicity is violated. This problem is interesting from a theoretical viewpoint as well. If we can determine whether adding this extra 'degree of freedom' to the MONOTONICITY problem 'lifts' the complexity of the problem into a broader class, we gain some insight in the properties and power of these types of complexity classes.

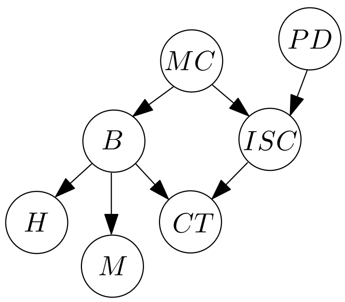
In the remainder of this paper, some preliminaries are introduced in Section 2, and various monotonicity problem variants and their computational complexity are discussed in Section 3. In Section 4, we present an (alternative) proof for the co-NP^{PP}-completeness of a restricted version of the MONOTONICITY problem as presented in [8]. This proof technique is then used in Section 5 to show that the MONOTONICITY problem with no fixed orderings, is co-$\text{NP}^{\text{NP}^{\text{PP}}}$-hard if we allow a succinct conditional probability representation of the variables. Finally, in Section 6 these results are discussed and the paper is concluded.

2 Preliminaries

Before formalizing the problems for which we want to determine their computational complexity, we first need to introduce some definitions and notations. Throughout this paper, we will refer to the *Brain tumor* network, shown in Figure 1, as a running example. This network, adapted from [9], captures some fictitious and incomplete medical knowledge related to metastatic cancer. The presence of metastatic cancer (modeled by the variable MC) typically induces the development of a brain tumor (B), and an increased level of serum calcium (ISC). The latter can also be caused by Paget's disease (PD). A brain tumor is likely to increase the severity of headaches (H) the patient will suffer. Long-term memory (M) is probably impaired, or even malfunctioning. Furthermore, it is likely that a CT-scan (CT) of the head will find a tumor if it is present, but it may also reveal other anomalies, like a fracture or a lesion, which might indicate an increased serum calcium. Note that in this network, MC, PD, B, and ISC are binary variables, while M, H and CT have multiple values.

2.1 Monotonicity in probabilistic networks

Let $\mathbf{B} = (G, \Gamma)$ be a probabilistic or Bayesian network where Γ, the set of conditional probability distributions, is composed of rational probabilities, and let Pr be its joint probability distribution. The conditional probability distributions in Γ can be explicit, i.e., represented with look-up tables, or succinct (implicitly defined), i.e., represented by a polynomial time computable function. If Γ consists only of explicit distributions then $\mathbf{B}$ will be denoted as an explicit network. Let $\Omega(V)$ denote the set of values that $V \in V(G)$ can take. Vertex A is



$\Pr(mc) =$	0.20
$\Pr(pd) =$	0.10
$\Pr(b \mid mc) =$	0.20
$\Pr(b \mid \overline{mc}) =$	0.05
$\Pr(M = \text{norm.} \mid b) =$	0.50
$\Pr(M = \text{imp.} \mid b) =$	0.40
$\Pr(M = \text{malf.} \mid b) =$	0.10
$\Pr(M = \text{norm.} \mid \bar{b}) =$	0.70
$\Pr(M = \text{imp.} \mid \bar{b}) =$	0.25
$\Pr(M = \text{malf.} \mid \bar{b}) =$	0.05
$\Pr(H = \text{sev.} \mid b) =$	0.70
$\Pr(H = \text{mod.} \mid b) =$	0.25
$\Pr(H = \text{abs.} \mid b) =$	0.05
$\Pr(H = \text{sev.} \mid \bar{b}) =$	0.30
$\Pr(H = \text{mod.} \mid \bar{b}) =$	0.20
$\Pr(H = \text{abs.} \mid \bar{b}) =$	0.50

$\Pr(isc \mid mc, pd) =$	0.95
$\Pr(isc \mid mc, \overline{pd}) =$	0.80
$\Pr(isc \mid \overline{mc}, pd) =$	0.70
$\Pr(isc \mid \overline{mc}, \overline{pd}) =$	0.20
$\Pr(\text{tum.} \mid b, isc) =$	0.80
$\Pr(\text{tum.} \mid b, \overline{isc}) =$	0.90
$\Pr(\text{tum.} \mid \bar{b}, isc) =$	0.05
$\Pr(\text{tum.} \mid \bar{b}, \overline{isc}) =$	0.10
$\Pr(\text{fract.} \mid b, isc) =$	0.18
$\Pr(\text{fract.} \mid b, \overline{isc}) =$	0.01
$\Pr(\text{fract.} \mid \bar{b}, isc) =$	0.55
$\Pr(\text{fract.} \mid \bar{b}, \overline{isc}) =$	0.40
$\Pr(\text{les.} \mid b, isc) =$	0.02
$\Pr(\text{les.} \mid b, \overline{isc}) =$	0.09
$\Pr(\text{les.} \mid \bar{b}, isc) =$	0.40
$\Pr(\text{les.} \mid \bar{b}, \overline{isc}) =$	0.50

Figure 1: The *Brain tumor* network

denoted as a predecessor of B if $(A, B) \in A(G)$. For a node B with predecessors $A_1, \ldots, A_n$, the *configuration template* $\mathbf{A}$ is defined as $\Omega(A_1) \times \ldots \times \Omega(A_n)$; a particular instantiation of $A_1, \ldots, A_n$ will be denoted as a *configuration* of $\mathbf{A}$.

Monotonicity can be defined as stochastical dominance (monotone in distribution) or in a modal sense (monotone in mode). For variable set E, with value assignments e and e' ($e \prec e'$) and variable set C, the network is isotone in distribution if $\Pr(C \mid e)$ is stochastically dominant over $\Pr(C \mid e')$. The network is isotone in mode if the most probable instantiation of C given assignment e is lower ordered than the most probable instantiation of C given assignment e'. In practice, C will normally be a single variable of interest (e.g., the main classifier or output variable in the network), and E will normally denote the set of observable variables. Without loss of generality, we will assume in the remainder that C is a singleton variable, rather than a set of variables. Monotonicity can be defined on a global scale, or locally (only relations between endpoints of arcs in the network are considered). The latter is relevant when constructing qualitative probabilistic networks (See for example [10]). In this paper, we discuss global monotonicity in distribution only; the reader can refer to [11] for a discussion of local monotonicity. We distinguish between weak and strong notions of global monotonicity.

Definition 1 (global monotonicity [8]). Let F_{Pr} be the cumulative distribution function for a node $V \in V(G)$, defined by $F_{Pr}(v) = Pr(V \leq v)$ for all $v \in \Omega(V)$. Let $C \in V(G)$ and let $E \subseteq V(G) \setminus \{C\}$, and let $\mathbf{E}$ be the configuration template of E. C is *strongly monotone in* E, if either

$$e \preceq e' \rightarrow F_{Pr}(c \mid e) \leq F_{Pr}(c \mid e') \quad \text{for all } c \in \Omega(C) \text{ and all } e, e' \in \mathbf{E} \quad \text{, or}$$
$$e \preceq e' \rightarrow F_{Pr}(c \mid e) \geq F_{Pr}(c \mid e') \quad \text{for all } c \in \Omega(C) \text{ and all } e, e' \in \mathbf{E}$$

C is *weakly monotone in* E, if C is strongly monotone in $\{E_i\}$, for all variables $E_i \in E$.

In our running example, we might want to know whether MC is monotone in M and H, i.e., whether the probabilities in the network are such, that more severe symptoms always make metastatic cancer more likely. It follows that MC is *weakly* monotone in $\{M, H\}$, since MC is isotone in H (more severe headaches make metastatic cancer *more* likely) and it is antitone in M (better functioning long term memory makes metastatic cancer *less* likely). For that reason, MC is not *strongly* monotone in $\{M, H\}$, given the ordering *severe* > *moderate* > *absent*) for H and *normal* > *impaired* > *malfunctioning*) for M.

Note that all networks that are strongly monotone in some set E are also weakly monotone, but not vice versa: whereas the strong variant assumes a partial order on all configurations of E, the weak variant allows independent isotone or antitone effects for all variables in E. Put in another way: we could make a weakly monotone network also strongly monotone by reversing the order of the values of some variables in E, such that all effects are antitone or all effects are isotone. In our example, we could reverse the order of the values of M to make MC also strongly monotone in $\{M, H\}$.

Typically, monotonicity is relevant in the construction phase of a network. The probabilities associated with the network are often elicited from estimations by experts, or learned from data. In both cases, the network properties induced by these probabilities might not reflect the actual situation. For example, medical experience related to brain tumors may dictate that more severe symptoms always increase the likeliness of metastatic cancer. When a domain expert insists that a certain relation ought to be monotone, the joint probability distribution should be such that this property is reflected in the network [12]. If monotonicity is violated, the probability distribution in the network can be revised in cooperation with the expert.

The above notions of monotonicity assume an implicit *ordering* on the values of the variables involved. Such an ordering is often trivial (e.g., $b > \bar{b}$ and *severe* > *moderate* > *absent*) but sometimes it is arbitrary, like an ordering of the anomalies that may be found using a CT-scan. Nevertheless, a certain ordering is necessary to determine whether the network is monotone, or to determine which parts of the network are violating monotonicity assumptions. Thus, for nodes where no *a priori* ordering is given, we want to order the values of these nodes in such a way that either the whole network becomes monotone, or – if no such ordering exists – the number of relations where monotonicity is violated is minimized. We define the notion of an *interpretation* of V to denote a certain ordering on the values of V. Note that the number of distinct interpretations of a node with k values equals $k!$, as each permutation of the k values is a distinct interpretation.

Definition 2 (interpretation). Let $\Omega(V)$ denote the set of values of $V \in V(G)$. An *interpretation* of V, denoted I_V, is defined as a total ordering on $\Omega(V)$. The *interpretation set* $\mathbf{I_V}$ is defined as the set of all possible interpretations of V.

We will often omit the subscript of an interpretation if no confusion of variables is possible; for arbitrary interpretations we will often use σ and τ.

Suppose that monotonicity between metastatic cancer MC and the observations CT and M is demanded in the *Brain tumor* network. The reader can infer that monotonicity is preserved if CT is ordered *tumor* > *fracture* > *lesion*, but is violated when CT is ordered *tumor* > *lesion* > *fracture*, for example. Thus, the relevant variables in the network for this monotonicity query can be ordered such that MC is monotone in $\{CT, M\}$. Note that MC is *not* monotone in $\{CT, M, H\}$ for *any* ordering of the variables, since

$\Pr(mc \mid H = abs., M = norm., CT = tum.) > \Pr(mc \mid H = abs., M = norm., CT = les.)$, but $\Pr(mc \mid H = abs., M = malf., CT = tum.) < \Pr(mc \mid H = abs., M = malf., CT = les.)$.

If monotonicity is violated in a network, it is important to minimize the number of probabilities that need to be reconsidered. Thus, when constructing networks, one should find an ordering that minimizes the offending context. In this paper,we will investigate the computational complexity of this problem.

2.2 Computational complexity

In the remainder, we assume that the reader is familiar with basic concepts of computational complexity theory, such as the classes P, NP and co-NP, hardness, completeness, oracles, and the polynomial hierarchy (PH). For a thorough introduction to these subjects, we refer to textbooks like [13] and [14].

In addition to these concepts, we use the *counting hierarchy* (CH) [15, 16]. The counting hierarchy closely resembles (in fact, *contains*) the polynomial hierarchy, but also involves the class PP (probabilistic polynomial time), i.e., the class of languages L accepted by a non-deterministic Turing Machine where the *majority* of the paths accept a string s if and only if $s \in L$. Recall that the polynomial hierarchy can be characterized by alternating existential and universal operators applied to P, where $\exists^P P$ equals $\Sigma_1^p = \text{NP}$, $\forall^P P$ equals $\Pi_1^p = \text{co-NP}$, while $\forall^P \exists^P \forall^P \ldots P$ equals Π_k^p and $\exists^P \forall^P \exists^P \ldots P$ equals Σ_k^p, where k denotes the number of alternating quantifiers.

A convenient way to relate the counting hierarchy to the polynomial hierarchy is by introducing an additional operator C, where C_0^p equals P, C_1^p equals PP, and in general $C_{k+1}^p = C \cdot C_k^p = (C_k^p)^{\text{PP}}$. Interesting complexity classes can be defined using these operators $\exists^P$, $\forall^P$ and C in various combinations. For example, $\exists^P CP$ equals the class NP^{PP}, $\forall^P CP$ equals co-$\text{NP}^{\text{PP}}$and $\exists^P \forall^P CP$ equals $\text{NP}^{\text{NP}^{\text{PP}}}$. Default complete problems for these kind of complexity classes are defined by Wagner [15] using quantified satisfiability variants. In this paper we consider in particular the complete problems MAJSAT, E-MAJSAT, A-MAJSAT, EA-MAJSAT, and AE-MAJSAT, which will be used in the hardness proofs. These problems are proven complete by Wagner [15] for the classes PP, NP^{PP}, co-NP^{PP}, $\text{NP}^{\text{NP}^{\text{PP}}}$ and co-$\text{NP}^{\text{NP}^{\text{PP}}}$, respectively. In all problems, we consider a Boolean formula ϕ with n variables X_i, with $1 \leq i \leq n$, and we introduce quantifiers to bound subsets of these variables.

MAJSAT
Instance: Let $\mathbf{X}$ denote the set of variables of ϕ.
Question: Does at least half of the instantiations of $\mathbf{X}$ satisfy ϕ?

E-MAJSAT
Instance: Let $1 \leq k \leq n$, let $\mathbf{X_E}$ denote the set of variables X_1 to X_k and let $\mathbf{X_M}$ denote the set of variables X_{k+1} to X_n.
Question: Is there an instantiation of $\mathbf{X_E}$, such that at least half of the instantiations of $\mathbf{X_M}$ satisfy ϕ?

A-MAJSAT
Instance: Let $1 \leq k \leq n$, let $\mathbf{X_A}$ denote the set of variables X_1 to X_k and let $\mathbf{X_M}$ denote the set of variables X_{k+1} to X_n.

Question: Does, for every possible instantiation of $\mathbf{X_A}$, at least half of the instantiations of $\mathbf{X_M}$ satisfy ϕ?

EA-MAJSAT
Instance: Let $1 \leq k \leq l \leq n$, let $\mathbf{X_E}$, $\mathbf{X_A}$, and $\mathbf{X_M}$ sets of variables X_1 to X_k, X_{k+1} to X_l, and X_{l+1} to X_n, respectively.
Question: Is there an instantiation of $\mathbf{X_E}$, such that, for every possible instantiation of $\mathbf{X_A}$, at least half of the instantiations of $\mathbf{X_M}$ satisfy ϕ?

AE-MAJSAT
Instance: Let $1 \leq k \leq l \leq n$, let $\mathbf{X_A}$, $\mathbf{X_E}$, and $\mathbf{X_M}$ sets of variables X_1 to X_k, X_{k+1} to X_l, and X_{l+1} to X_n, respectively.
Question: Is there, for all instantiations of $\mathbf{X_A}$, a possible instantiation of $\mathbf{X_E}$, such that at least half of the instantiations of $\mathbf{X_M}$ satisfy ϕ?

In the remainder, we denote the complement of a problem P as NOT-P, with 'yes' and 'no' answers reversed with respect to the original problem P. Note that, by definition, if P is in complexity class C, then NOT-P is in co-C, and, likewise, if NOT-P is in C, then P is in co-C.

2.3 Succinct conditional probability representation

Normally, probabilistic networks are specified by a directed acyclic graph (e.g., using an adjacency list or matrix) and, for each variable, a conditional probability table. This table specifies, for each value of the variable and for each configuration of its parents, the probability of the variable having that particular value given this configuration. Normally, such probability tables are explicitly specified by a look-up table (e.g., a matrix). In some cases, however, the values and the probabilities are such, that they can be encoded in a more efficient data structure. Take for example a variable C with parents $X_1 \ldots X_4$, who have values T and F with uniform probability. C has values $c_0 \ldots c_4$ and acts as a counting variable. The probability of C having value c_i is exactly the probability that i out of C's four parents have value T. For example, $\Pr(C = c_0) = \frac{1}{16}$ and $\Pr(C = c_2) = \frac{6}{16}$. Given the above description as an input, we can compute the probability of C having a particular value (given an instantiation to X) in polynomial time. In line with formalizations of succinct representations of graphs, e.g. [17, 15], we will assume that the probability of any value can be calculated using a family of boolean circuits, which characterizes polynomial time computations [18].

Note that this definition of succinct conditional probability representation does not enforce that a particular interpretation can be encoded or evaluated in polynomial time. A succinct encoding of the values of a variable, can evaluate to exponentially many values and their probabilities; thus, it need not be the case that a polynomial succinct encoding exists of all individual interpretations.

3 Monotonicity variants and their complexity

In this paper, we study the computational complexity of various variants of global monotonicity. The following problems are defined on a probabilistic network $\mathbf{B} = (G, \Gamma)$, where $G = (V, A)$ is a directed acyclic graph, $C \in V(G)$ and $E \subseteq V(G) \setminus \{C\}$.

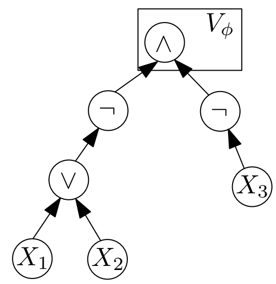


Figure 2: The probabilistic network corresponding to $\neg(x_1 \vee x_2) \wedge \neg x_3$

1. The STRONG GLOBAL MONOTONICITY problem is the problem of deciding whether C is strongly globally monotone in E, given an interpretation of all $V \in V(G)$. This problem has been proven co-NP$^{\text{PP}}$-complete [8] for explicit networks.

2. The WEAK GLOBAL MONOTONICITY problem is the problem of deciding whether C is weakly globally monotone in E, given an interpretation of all $V \in V(G)$.

3. The GLOBAL E-MONOTONICITY problem is the problem of deciding whether there exists an interpretation of all $V \in V(G)$, such that C is globally monotone in E.

Note that, if there exists an interpretation such that C is *weakly* monotone in E, there also exists an interpretation such that C is *strongly* monotone in E, making a distinction between weak and strong 'E'-variants redundant.

The WEAK GLOBAL MONOTONICITY and GLOBAL E-MONOTONICITY problems will be discussed in Sections 4 and 5. In these sections, we use a proof technique introduced by Park and Darwiche [7] to construct a probabilistic network $\mathbf{B}_\phi$ from a given Boolean formula ϕ with n variables. For all variables $X_i (1 \leq i \leq n)$ in this formula, we create a variable X_i in G, with possible values T and F and a uniform probability distribution. For each logical operator in ϕ, we create an additional variable in G, whose parents are the variables that correspond to the input of the operator, and whose conditional probability table is equal to the truth table of that operator. For example, the $\wedge$-operator would have a conditional probability of 1 if and only if both its parents have the value T, and 0 otherwise. Furthermore, we denote the top-level operator in ϕ with V_ϕ. In Figure 2 such a network is constructed for the formula $\neg(x_1 \vee x_2) \wedge \neg x_3$. Now, for any particular instantiation $\mathbf{x}$ of the set of all variables $\mathbf{X}$ in the formula we have that the probability of V_ϕ, given the corresponding configuration equals 1 if $\mathbf{x}$ satisfies ϕ, and 0 if $\mathbf{x}$ does not satisfy ϕ. Without any instantiation, the probability of V_ϕ is $\frac{\#_q}{2^n}$, where $\#_q$ is the number of satisfying instantiations of $\mathbf{X}$. Using such constructs, Park and Darwiche proved that the decision variant of the MAP problem is NP$^{\text{PP}}$-complete; we will use this construct as a starting point to prove completeness results for WEAK GLOBAL MONOTONICITY and GLOBAL E-MONOTONICITY in the following sections.

4 Weak Global Monotonicity

In this section, we present a proof for WEAK GLOBAL MONOTONICITY (with explicit representations) using the technique of Park and Darwiche [7]. Note that STRONG GLOBAL MONOTONICITY has been proven to be co-NP$^{\text{PP}}$-complete in [8] using a reduction from the

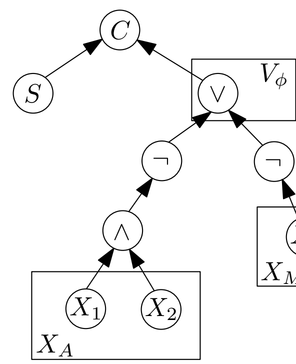


Figure 3: Construct for hardness proof Monotonicity

decision variant of the PARTIAL MAP-problem, and that co-NP$^{\text{PP}}$-hardness of the weak variant can be proven by restriction. We construct a reduction from A-MAJSAT, the relevant satisfiability variant discussed in Section 2, in order to facilitate our main result in the next section. First, we state the decision problem for which we prove co-NP$^{\text{PP}}$-hardness:

WEAK GLOBAL MONOTONICITY
Instance: Let $\mathbf{B} = (G, \Gamma)$ be a Bayesian network where Γ is composed of explicitly represented rational probabilities, and let Pr be its joint probability distribution. Let $C \subseteq V(G)$ and $E \subseteq V(G) \setminus \{C\}$.
Question: Is $\mathbf{B}$ weakly monotone in distribution in E?

Below, we will prove that any instance $(\phi, \mathbf{X_A}, \mathbf{X_M})$ of A-MAJSAT can be translated to a probabilistic network that is monotone, if and only if $(\phi, \mathbf{X_A}, \mathbf{X_M})$ is satisfiable. As an example, let us consider the formula $\phi = \neg(x_1 \wedge x_2) \vee \neg x_3$ (see Figure 3), and let $\mathbf{X_A} = \{x_1, x_2\}$ and $\mathbf{X_M} = \{x_3\}$. This is a 'yes'-instance of A-MAJSAT because, for every configuration of $\mathbf{X_A}$, at least half of the configurations of $\mathbf{X_M}$ satisfies ϕ. From ϕ we construct a network $\mathbf{B}_\phi$ as described in the previous section. Furthermore, a node C ('classifier') and a node S ('select') is added, with arcs (S, C) and (V_ϕ, C), where V_ϕ is the top node in $\mathbf{B}_\phi$. S has values T and F with uniform distribution, and C has conditional probabilities as denoted in Table 1. The proof of theorem 3 shows that in the thus constructed network, $\Pr(C \mid S \wedge \mathbf{X_A})$ is weakly monotone in distribution, if and only if the corresponding A-MAJSAT-instance $(\phi, \mathbf{X_A}, \mathbf{X_M})$ is satisfiable.

	c_1	c_2	c_3
$S = T \wedge V_\phi = T$	0.5	0.25	0.25
$S = T \wedge V_\phi = F$	0.5	0.25	0.25
$S = F \wedge V_\phi = T$	0.25	0.375	0.375
$S = F \wedge V_\phi = F$	0.375	0.5	0.125

Table 1: Conditional probability table for C

Theorem 3. WEAK GLOBAL MONOTONICITY *is co-*NP$^{\text{PP}}$*-complete.*

Proof. To prove membership of co-NP$^{\text{PP}}$, we consider NOT-WEAK GLOBAL MONOTONICITY and prove membership of NP$^{\text{PP}}$. In this complement problem we must decide whether there exist instantiations to the evidence variables E such that $\mathbf{B}$ is *not* monotone in distribution. This is clearly in NP$^{\text{PP}}$: we can non-deterministically choose instantiations $\mathbf{e_1} \preceq \mathbf{e_2}$ to E and values $c < c' \in \Omega(C)$, and verify that $F_{Pr}(c \mid e_1) \leq F_{Pr}(c' \mid e_1)$, but $F_{Pr}(c' \mid e_2) \leq F_{Pr}(c \mid e_2)$ since PR is PP-complete.

To prove co-NP$^{\text{PP}}$-hardness, we construct a transformation from the A-MAJSAT problem. Let $(\phi, \mathbf{X_A}, \mathbf{X_M})$ be an instance of this problem, and let $\mathbf{B}_\phi$ be the network constructed from ϕ as described above. Given a particular configuration $\mathbf{x}$ of all n variables in $\mathbf{X_A} \cup \mathbf{X_M}$, $\Pr(V_\phi \mid \mathbf{x})$ equals 1 if $\mathbf{x}$ is a satisfying configuration and 0 if it is not, hence, for any configuration $\mathbf{X_A}$, $V_\phi \geq \frac{1}{2}$ if at least half of the instantiations to $\mathbf{X_M}$ satisfy ϕ. Since C is conditioned on V_ϕ, it follows from Table 1 that if any configuration of $\mathbf{X_A}$ leads to $\Pr(V_\phi) < \frac{1}{2}$, then C is no longer monotone in $S \wedge \mathbf{X_A}$, since $F_{Pr}(c_1 \mid S = T) > F_{Pr}(c_1 \mid S = F)$, but $F_{Pr}(c_2 \mid S = T) < F_{Pr}(c_2 \mid S = F)$ as we can calculate[1] from the conditional probability table for C. On the other hand, if $(\phi, \mathbf{X_A}, \mathbf{X_M})$ is a satisfying instantiation of A-MAJSAT, then $\Pr(V_\phi) \geq \frac{1}{2}$ and thus $\mathbf{B}_\phi$ is weakly globally monotone. Thus, if we can decide whether $\mathbf{B}_\phi$ is weakly globally monotone in $S \cup \mathbf{X_A}$, we are able to decide $(\phi, \mathbf{X_A}, \mathbf{X_M})$. Therefore WEAK GLOBAL MONOTONICITY is co-NP$^{\text{PP}}$-hard. □

This completeness result (and the similar result for the strong variant) also holds when the probability distribution of the variables is succinctly represented, rather than explicitly represented using look-up tables.

Corollary 4. GLOBAL MONOTONICITY *with succinct probability representation is co-*NP$^{\text{PP}}$*-complete.*

Proof. co-NP$^{\text{PP}}$-hardness follows directly from Theorem 3. We will now prove membership of co-NP$^{\text{PP}}$. Since the values of ψ are ordered, we can - without loss of generality - assume that they are represented by natural numbers, with $<$ as a natural ordering. Let us assume that, for a particular variable ψ, its values and their corresponding probabilities are succinctly represented by a family of boolean circuits that produces $\Pr(\psi = k) = p_k$ for any value k and a configuration of its parents. Assume we are presented two sets of configurations $\mathbf{e_1}$ and $\mathbf{e_2}$ to a set of variables, including ψ, such that $\mathbf{e_1} \prec \mathbf{e_2}$, and $F_{Pr}(C \mid \mathbf{e_1}) > F_{Pr}(C \mid \mathbf{e_2})$. Since ψ is ordered, we can verify that $\mathbf{e_1} \prec \mathbf{e_2}$ in time, polynomial to the input length. Since we can also calculate $F_{Pr}(C \mid \mathbf{e_1})$ and $F_{Pr}(C \mid \mathbf{e_2})$ in polynomial time, given access to a PP oracle, $(\mathbf{e_1}, \mathbf{e_1})$ is a certificate of a counter example that can be verified in polynomial time using a PP oracle. Hence, even with succinct probability representation, GLOBAL MONOTONICITY $\in$ co-NP$^{\text{PP}}$. □

5 Global E-Monotonicity

We now use the proof technique from the previous section to prove that GLOBAL E-MONOTONICITY is co-NP$^{\text{NP}^{\text{PP}}}$-hard if we allow succinct representations for the conditional probability distributions. The canonical satisfiability variant complete for that complexity class is

[1] $F_{Pr}(c_1 \mid S = T) = \Pr(c_1 \mid V_\phi = T \wedge S = T) \cdot \Pr(V_\phi = T) + \Pr(c_1 \mid V_\phi = F \wedge S = T) \cdot \Pr(V_\phi = F) = (0.5+\epsilon) \cdot 0.5 + (0.5-\epsilon) \cdot 0.5 = 0.5$. Likewise, $F_{Pr}(c_1 \mid S = F) = 0.25 \cdot (0.5-\epsilon) + 0.375 \cdot (0.5+\epsilon) = 0.3125 + 0.125\epsilon$. On the other hand, $F_{Pr}(c_2 \mid S = T) = \Pr(c_1 \mid S = T) + \Pr(c_2 \mid S = T) = 0.5 + 0.25 < F_{Pr}(c_2 \mid S = F) = \Pr(c_1 \mid S = F) + \Pr(c_2 \mid S = F) = (0.3125 + 0.125\epsilon) + (0.4375 + 0.125\epsilon) = 0.75 + 0.25\epsilon$.

AE-MAJSAT. However, we will use the equivalent complexity class NOT-EA-MAJSAT for our reduction. Thus, instead of $\forall^P \exists^P \mathcal{C}$ we use the equivalent problem statement $\neg\exists^P \forall^P \mathcal{C}$. The reader can verify that this is an equivalent problem formulation if we negate the variables that are bounded by the $\mathcal{C}$ operator, and thus that both are complete problems for co-$\text{NP}^{\text{NP}^{\text{PP}}}$. Again, we start with a formal definition of the relevant decision problem:

> GLOBAL E-MONOTONICITY
> **Instance:** Let $\mathbf{B} = (G, \Gamma)$ be a Bayesian network where Γ is composed of rational probabilities, and let Pr be its joint probability distribution. Let $\Omega(V)$ denote the set of values that $V \in V(G)$ can take, and let $C \in V(G)$ and $E \subseteq V(G) \setminus \{C\}$.
> **Question:** Is there an interpretation I_V for all variables $V \in V(G)$, such that $\mathbf{B}$ is monotone in distribution in E?

We will see that any instance $(\phi, \mathbf{X_E}, \mathbf{X_A}, \mathbf{X_M})$ of NOT-EA-MAJSAT can be translated to a probabilistic network for which there exists an ordering of the values of its variables that makes the network monotone, if and only if $(\phi, \mathbf{X_E}, \mathbf{X_A}, \mathbf{X_M})$ is *not* satisfiable. As an example, let us consider the formula $\phi = \neg((x_1 \vee x_2) \wedge (x_3 \vee x_4)) \wedge x_5$ (Figure 4), let $\mathbf{X_E} = \{x_1, x_2\}$ and let $\mathbf{X_A} = \{x_3, x_4\}$ and $\mathbf{X_M} = \{x_5\}$. One can verify that this is indeed a 'yes'-instance of NOT-EA-MAJSAT: Whatever instantiation of the variables in $\mathbf{X_E}$ we choose, there always exists an instantiation to $\mathbf{X_A}$ (in particular, $x_3 = x_4 = F$) such that at least half of the instantiations of $\mathbf{X_M}$ satisfies ϕ. Thus, there does not exist an instantiation to $\mathbf{X_E}$, such that for *all* instantiations to $\mathbf{X_A}$ at least half of the instantiations of $\mathbf{X_M}$ does *not* satisfy ϕ.

Again, we denote V_ϕ as the top node in $\mathbf{B}_\phi$. We now add three additional variables, C with values c_1, c_2, c_3, D with values d_1, d_2, and a variable ψ. This variable is succincly defined and has (implicit) values w_0 to w_{2^m-1} ($m =\mid \mathbf{X_E} \mid$) that correspond to configurations $\mathbf{x_E}$ of $\mathbf{X_E}$. These values are ordered by the binary representation of each configuration $\mathbf{x_E}$, e.g., for an instantiation $\mathbf{x_E} = X_1 = F, \ldots, X_{m-1} = F, X_m = T$ the binary representation would be $0 \ldots 01$ and therefore this particular configuration would correspond with w_1. Likewise, all possible configurations of $\mathbf{X_E}$ are mapped to values w_i of ψ. Furthermore, there are arcs (V_ϕ, C), (ψ, C), (C, D), and from every variable in $\mathbf{X_E}$ to ψ. The conditional probability $\Pr(C \mid V_\phi \wedge \psi)$ is defined in Table 2, where ϵ is a sufficiently small number, e.g. $\epsilon \leq \frac{1}{2^{m+3}}$. The conditional probabilities $\Pr(\psi \mid \mathbf{X_E})$ and $\Pr(D \mid C)$ are defined in Table 3. Note, that the conditional probability distributions of both ψ and C are defined succinctly.

The conditional probabilities of D are chosen in such a way, that D is monotone in C if and only if $I_C = \{c_1 < c_2 < c_3\}$. In the example, the possible values of ψ are numbered as follows: $w_0 = \{X_1 = F, X_2 = F\}, w_1 = \{X_1 = F, X_2 = T\}, w_2 = \{X_1 = T, X_2 = F\}, w_3 = \{X_1 = T, X_2 = T\}$. For $i = 0 \ldots 3$, the conditional probability table $\Pr(C \mid V_\phi \wedge \psi = w_i)$ is defined as in Table 4. We have already seen that, for all configurations to $\mathbf{X_A}$, the configuration $X_3 = X_4 = F$ of $\mathbf{X_E}$ ensures that the majority of the possible configurations of $\mathbf{X_M}$ satisfies ϕ. Therefore, for all configurations of $\mathbf{X_A}$, there is at least one configuration of ψ (namely, $\psi = w_0$) such that $V_\phi \geq \frac{1}{2}$. Since C is conditioned on V_ϕ, we can calculate from the table that monotonicity is violated: $F_{Pr}(c_1 \mid \psi = w_0) = 0.625 - 0.5\epsilon > F_{Pr}(c_1 \mid \psi = w_1) = 0.4375$ but $F_{Pr}(c_2 \mid \psi = w_0) = 0.75 - 0.5\epsilon < F_{Pr}(c_2 \mid \psi = w_1) = 0.75$. Thus, independent of the manner the values of $\Omega(\psi)$ are ordered, there is always at least one violation of monotonicity for any interpretation in $\mathbf{I}_\psi$ if $V_\phi \geq \frac{1}{2}$. If, on the other hand, there does not exist such configuration to $\mathbf{X_E}$, then $V_\phi < \frac{1}{2}$ for all possible configurations to $\mathbf{X_E}$, and thus there is an ordering of

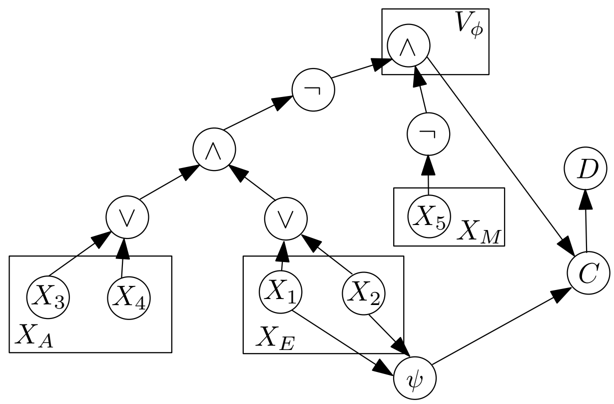


Figure 4: Construct for hardness proof E-Monotonicity

$$
\begin{array}{rcl}
\Pr(C = c_1 \,|\, V_\phi = T \wedge \psi = w_i) = & \frac{1}{2} - \frac{i}{2^{m+1}} - \epsilon & \text{if i} = 0 \\
 & \frac{1}{2} - \frac{i}{2^{m+1}} & \text{otherwise} \\
\Pr(C = c_2 \,|\, V_\phi = T \wedge \psi = w_i) = & \frac{i+1}{2^m} - \frac{1}{2^{m+1}} & \\
\Pr(C = c_3 \,|\, V_\phi = T \wedge \psi = w_i) = & \frac{1}{2} - \frac{i+1}{2^{m+1}} + \epsilon & \text{if i} = 0 \\
 & \frac{1}{2} - \frac{i+1}{2^{m+1}} & \text{otherwise} \\
\Pr(C = c_1 \,|\, V_\phi = F \wedge \psi = w_i) = & 1 - \frac{i+1}{2^m} & \\
\Pr(C = c_2 \,|\, V_\phi = F \wedge \psi = w_i) = & \frac{i+1}{2^{m+1}} & \\
\Pr(C = c_3 \,|\, V_\phi = F \wedge \psi = w_i) = & \frac{i+1}{2^{m+1}} &
\end{array}
$$

Table 2: Conditional probability for C

	d_1	d_2
c_1	0.20	0.80
c_2	0.40	0.60
c_3	0.60	0.40

$$
\begin{array}{rcl}
\Pr(\psi = w_i \,|\, \mathbf{x_E}) = & 1 & \text{if } w_i \text{ corresponds to } \mathbf{x_E} \\
 & 0 & \text{otherwise}
\end{array}
$$

Table 3: Conditional probabilities for $\Pr(D \,|\, C)$ and $\Pr(\psi \,|\, X)$

the interpretations in $\mathbf{I}_\psi$ such that $\Pr(C \,|\, \mathbf{X_A})$ is monotone. Note that we cannot assume an *a priori* ordering on the values of ψ in this situation: although all configurations of $\mathbf{X_E}$ lead to $V_\phi < \frac{1}{2}$, some may be closer to $\frac{1}{2}$ than others and thus, because of the conditioning on V_ϕ, lead to higher values in C. We claim that there is a possible interpretation I for all variables in $\mathbf{X_E} \cup \{\psi\}$ in the thus constructed network, such that the network is globally monotone, if and only if the corresponding NOT-EA-MAJSAT-instance is satisfiable.

Theorem 5. GLOBAL E-MONOTONICITY *is co-*$\text{NP}^{\text{NP}^{\text{PP}}}$*-hard for networks that allow succinct conditional probability representations.*

	c_1	c_2	c_3
$\psi = w_0 \wedge V_\phi = T$	$0.5 - \epsilon$	0.125	$0.375 + \epsilon$
$\psi = w_1 \wedge V_\phi = T$	0.375	0.375	0.25
$\psi = w_2 \wedge V_\phi = T$	0.25	0.625	0.125
$\psi = w_3 \wedge V_\phi = T$	0.125	0.875	0
$\psi = w_0 \wedge V_\phi = F$	0.75	0.125	0.125
$\psi = w_1 \wedge V_\phi = F$	0.5	0.25	0.25
$\psi = w_2 \wedge V_\phi = F$	0.25	0.375	0.375
$\psi = w_3 \wedge V_\phi = F$	0	0.5	0.5

Table 4: Conditional probability for C in the example

Proof. To prove co-$\text{NP}^{\text{NP}^{\text{PP}}}$-hardness, we construct a transformation from NOT-EA-MAJSAT. Let $(\phi, \mathbf{X_E}, \mathbf{X_A}, \mathbf{X_M})$ be an instance of this problem, and let $\mathbf{B}_\phi$ be the network constructed from ϕ as described above. If $(\phi, \mathbf{X_E}, \mathbf{X_A}, \mathbf{X_M})$ is *not* satisfiable, then there exists an instantiation to ψ, such that $\Pr(V_\psi) \geq \frac{1}{2}$ and thus – again, because of the conditioning of C on V_ψ – monotonicity is violated. But if this is the case, then there exist $w_i, w_j \in \psi$ and $c < c' \in C$ such that $F_{Pr}(c \,|\, \psi = w_i) \leq F_{Pr}(c' \,|\, \psi = w_i)$, but $F_{Pr}(c' \,|\, \psi = w_j) \leq F_{Pr}(c \,|\, \psi = w_j)$ independent of the ordering of the values of ψ. Note that the variable-and operator-nodes have binary values, making an ordering irrelevant[2], and the ordering on C and D is imposed by the conditional probability $\Pr(D \mid C)$. Thus, if we would be able to decide that there is an interpretation of the values of the variables of $\mathbf{B}_\phi$ such that $\mathbf{B}_\phi$ is globally monotone in distribution, we are able to decide $(\phi, \mathbf{X_E}, \mathbf{X_A}, \mathbf{X_M})$. On the other hand, given that the network is globally monotone, we know that there cannot be an instantiation to $\mathbf{X_E}$ such that $(\phi, \mathbf{X_E}, \mathbf{X_A}, \mathbf{X_M})$ is satisfied. Hence, GLOBAL E-MONOTONICITY is co-$\text{NP}^{\text{NP}^{\text{PP}}}$-hard. It may not be obvious that the above construction can be made in polynomial time. Note that, regardless how large $\mathbf{X_E}$ becomes, both the conditional probabilities $\Pr(\psi \,|\, \mathbf{X_E})$ and $\Pr(C \mid V_\phi \wedge \psi)$ can be succinctly represented using families of polynomial sized circuits. For $\Pr(\psi \,|\, \mathbf{X_E})$ these circuits are a trivial combination of AND-gates deciding whether w_i corresponds to $\mathbf{X_E}$. We argue that a similar family of circuits exists for $\Pr(C \,|\, V_\phi \wedge \psi)$ by noting that the probabilities are fractions whose denominator has the form 2^m, with $m = \mid X_E \mid \leq \mid X \mid$. Therefore, $\mathbf{B}_\phi$ can be constructed using succinct inputs in polynomial time. □

To prove *completeness* for co-$\text{NP}^{\text{NP}^{\text{PP}}}$, one needs to prove membership as well as hardness. Note that GLOBAL E-MONOTONICITY is in co-$\text{NP}^{\text{NP}^{\text{PP}}}$ for a restricted class of networks, namely networks with either explicit conditional probability representation, or with succinctly represented interpretations, i.e., networks in which every ordering can be calculated in polynomial time from the input. For a membership proof of that restricted class of networks we use NOT-WEAK GLOBAL MONOTONICITY as an NP^{PP}oracle. With the aid of this oracle, an interpretation for the values of the variables that violates monotonicity is an NP membership certificate for NOT-GLOBAL E-MONOTONICITY , thus by definition the problem is in co-$\text{NP}^{\text{NP}^{\text{PP}}}$. Nevertheless, *without* this restriction, the NP membership certificate (i.e., the interpretation of the variables) can grow exponentially large with respect to the succinct

[2] if $\mathbf{B}_\phi$ is isotone for $x < \bar{x}$, it is antitone for $\bar{x} < x$ and vice versa

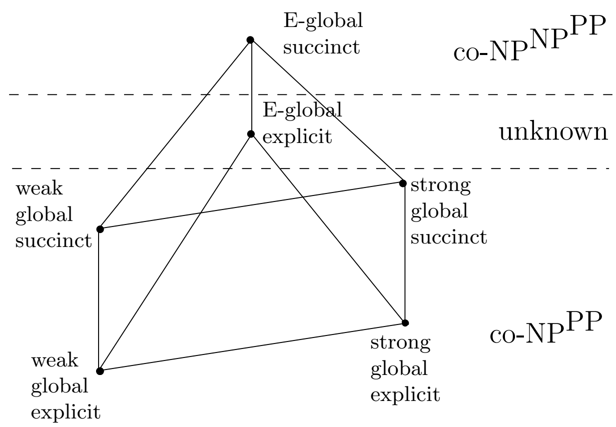


Figure 5: Known complexity results

conditional probability representation, even when using a factorial number notation (see e.g. page 65–66 of [19]) to represent permutations.

6 Conclusion

In this paper, several variants of the Global Monotonicity problem in probabilistic networks were introduced. In Figure 5, the known complexity results for strong and weak global monotonicity variants with explicit or succinct conditional probability distributions are presented. The main result is the hardness proof of Global E-Monotonicity with succinct probability representation. It is established that this problem is co-NP$^{\mathrm{NP}^{\mathrm{PP}}}$-hard. Furthermore it has been shown that the completeness results for weak and strong variants of Global Monotonicity are preserved when succinct input is used, rather than explicit conditional probability tables. Unfortunately, a similar complexity result for Global E-Monotonicity with explicit representation could not be established. This problem is to be situated either in co-NP$^{\mathrm{PP}}$, in co-NP$^{\mathrm{NP}^{\mathrm{PP}}}$, or somewhere in between.

Acknowledgements

A preliminary version of this paper was presented at the 16th International Symposium on Fundamentals of Computation Theory, held August 27-30, 2007, in Budapest, Hungary [20]. The author was partially supported by the Netherlands Organisation for Scientific Research NWO. He wishes to thank Hans Bodlaender, Gerard Tel, and Linda van der Gaag, for fruitful discussions on this subject and useful comments on earlier drafts of this paper.

References

[1] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann, Palo Alto, 1988.

[2] B. Kappen, W. Wiegerinck, E. Akay, M. Nijman, J. Neijt, and A. van Beek. Promedas: A diagnostic decision support system. In *Proceedings of the 15th Belgian-Dutch Conference on Artificial Intelligence (BNAIC'03)*, pages 455–456, 2003.

[3] L. C. van der Gaag, S. Renooij, C. L. M. Witteman, B. M. P. Aleman, and B. G. Taal. Probabilities for a probabilistic network: a case study in oesophageal cancer. *Artificial Intelligence in Medicine*, 25:123–148, 2002.

[4] M. L. Littman, S. M. Majercik, and T. Pitassi. Stochastic boolean satisfiability. *Journal of Automated Reasoning*, 27(3):251–296, 2001.

[5] D. Roth. On the hardness of approximate reasoning. *Artificial Intelligence*, 82(1-2):273–302, 1996.

[6] S. E. Shimony. Finding MAPs for belief networks is NP-hard. *Artificial Intelligence*, 68(2):399–410, 1994.

[7] J. D. Park and A. Darwiche. Complexity results and approximation settings for MAP explanations. *Journal of Artificial Intelligence Research*, 21:101–133, 2004.

[8] L. C. van der Gaag, H.L. Bodlaender, and A. Feelders. Monotonicity in Bayesian networks. In *Twentieth Conference on Uncertainty in Artificial Intelligence*, pages 569–576. AUAI Press, 2004.

[9] G. F. Cooper. NESTOR: A computer-based medical diagnostic aid that integrates causal and probabilistic knowledge. Technical Report HPP-84-48, Stanford University, 1984.

[10] M. P. Wellman. Fundamental concepts of qualitative probabilistic networks. *Artificial Intelligence*, 44(3):257–303, 1990.

[11] J. Kwisthout, H. Bodlaender, and G. Tel. Local monotonicity in probabilistic networks. In K. Mellouli, editor, *Ninth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty. October 31-November 2, 2007, Hammamet, Tunisia*, volume 4724 of *LNCS*, pages 548–559. Springer-Verlag, 2007.

[12] L. C. van der Gaag, P. L. Geenen, and H. J. M. Tabachneck-Schijf. Verifying monotonicity in Bayesian networks with domain experts. In L. C. van der Gaag and R. Almond, editors, *Proceedings of the 4th Bayesian Modelling Applications Workshop: Bayesian Models Meet Cognition*, 2006.

[13] M. R. Garey and D. S. Johnson. *Computers and Intractability. A Guide to the Theory of NP-Completeness.* W. H. Freeman and Co., San Francisco, 1979.

[14] C. H. Papadimitriou. *Computational Complexity.* Addison-Wesley, 1994.

[15] K. W. Wagner. The complexity of combinatorial problems with succinct input representation. *Acta Informatica*, 23:325–356, 1986.

[16] J. Torán. Complexity classes defined by counting quantifiers. *Journal of the ACM*, 38(3):752–773, 1991.

[17] H. Galperin and A. Wigderson. Succinct representations of graphs. *Information and Control*, 56:183–193, 1983.

[18] J. L. Balcázar, J. Diaz, and J. Gabarró. *Structural Complexity*, volume I. SPringer-Verlag, 1988.

[19] D. E. Knuth. *The Art of Computer Programming*, volume 2: Seminumerical Algorithms. Addison-Wesly, third edition, 1997.

[20] J. Kwisthout. The computational complexity of monotonicity in probabilistic networks. In E. Csuhaj-Varj and Z. Esik, editors, *Sixteenth International Symposium on Fundamentals of Computation Theory, August 27-30, 2007, Budapest, Hungary*, volume 4639 of *LNCS*, pages 388–399. Springer-Verlag, 2007.