**DIMACS Technical Report 2007-06**
**April 18, 2007**

# On complexity of algorithms for modeling disease transmission and optimal vaccination strategy [1]

by

Endre Boros
RUTCOR, Rutgers University
640 Bartolomew Road
Piscataway, NJ 08854-8003
`boros@rutcor.rutgers.edu`

Vladimir Gurvich
RUTCOR, Rutgers University
640 Bartolomew Road
Piscataway, NJ 08854-8003
`gurvich@rutcor.rutgers.edu`

[1] This research was supported by DIMACS, the NSF Center for Discrete Mathematics and Theoretical Computer Science

---

# ABSTRACT

We consider simple deterministic models of disease transmission. Given a set of individuals $I$, we assign a hypergraph $H_i = (I \setminus \{i\}, E_i)$ to each $i \in I$ and assume that $i$ will be infected whenever there is a fully infected edge $e \in E_i$. Along with this general model $M_H$ we also study two special cases $M_G$ and $M_D$ when for all $i \in I$ the hypergraphs $H_i$ are specified implicitly by a (directed) graph $G = (I, E)$ and integral positive thresholds $k(i)$ for all $i \in I$. Then we assume that $i$ will be infected whenever at least $k(i)$ of his neighbors (predecessors) are infected.

Given a set $S$ of the originally infected individuals (a source) we generate the closure $T(S) = cl(S)$, that is, the set of all individuals that will be infected if the above transmission rules are applied iteratively sufficiently many times. We study all minimal sources such that
(i) $T(S) = I$, or (ii) $T(S)$ contains a given individual $q \in I$, or
(iii) $T(S)$ contains an edge of a given "target" hypergraph $H$.

We denote these three types of "targets" by $T_I, T_q$, and $T_H$ respectively. We show that, given a threshold $t$, it is NP-complete to decide whether there is a source $S$ of size at most $t$. The problem remains NP-complete for each of the three models $M_R, M_G$ or $M_D$ and targets $T_I, T_q$ or $T_H$. We also consider enumeration problems and show that if the transmission rule is given explicitly, $M_R$, then all inclusion minimal sources can be generated in incremental polynomial time for all targets $T_I, T_q$, or $T_H$. On the other hand, generating minimal sources is hard for all targets if the transmission model is given by a (directed) graph, $M_G$ or $M_D$, since for these two cases the input size may be logarithmic in the input size of $M_R$. Indeed, given $G = (I, E)$ and $k(i)$ for all $i \in I$, a corresponding hypergraph $H_i$ for some $i \in I$ may be exponential in $|I|$ unless $k(i)$ is bounded by a constant.

**Key words**: closure, disease transmission, incremental polynomial, generation algorithm, graph, hypergraph, DNF, Horn DNF, pure Horn DNF.

# 1 Introduction

## 1.1 Basic assumptions

We consider simple models of a disease transmission within a set of individuals $I$. Our basic assumptions are as follows:

1. We restrict ourselves with deterministic models and do not consider any random mechanisms.

2. We also restrict ourselves with irreversible models, that is, if an individual is infected (s)he remains infected. In other words, we consider time intervals that are not long enough for an individual to recover.

3. We ignore dynamics and consider only the initial and final situation. Mathematically, such an approach is described by the concept of *closure* $cl : 2^I \to 2^I$ satisfying the following standard axioms:

a) expansion: $S \subseteq cl(S)$;

b) monotonicity: $S \subseteq S' \Rightarrow cl(S) \subseteq cl(S')$

c) idempotent rule: $cl(cl(S)) = cl(S)$

Here $S \subseteq I$ and $T = cl(S) \subseteq I$ denote the sets of initially and finally infected individuals. We will refer to them as the *source* and *target*, respectively.

## 1.2 Source and target groups, fatal sources

We will consider the following three types of targets.

$T_I$: the whole set $I$.

$T_q$: one special individual $q \in I$ that we will call a *queen*.
In this case we assume that $q$ cannot belong to a source, $q \notin S$.

$T_H$: Given a family of sets (a hypergraph) $H = (I, E)$ whose vertices are the individuals $i \in I$ and edges $e \subseteq I$ are target groups. Target is achieved if at least one target group is fully infected. Clearly, this case generalizes the previous two.

A set $S$ of the initially infected individuals is a source. In the next section we will introduce several models of a disease transmission. Given such a model $M$ and a target $T$, we say a source $S$ is fatal (or $(M, T)$-fatal) if $S$ will infect $T$, given transmission rules $M$. There are two problems to consider:

**O, “optimization”**: find a fatal source of minimum size, and

**E, enumeration**: generate all fatal sources.

Given $M$ and $T$, these two problems can be hard or tractable. In this paper we survey the cases.

A motivation for such a terminology and for the above assumptions, as well, is the need to study the aftermath of a possible terrorist attack.

## 1.3 Disease transmission modeling

We consider three transmission models $M_G, M_D$, and $M_R$ defined respectively by a graph, digraph, and a set of transmission rules or implications.

**Model** $M_G$. Let $G = (I, E)$ be a graph on the ground set $I$. Edges of $G$ are interpreted as contacts or connections between individuals. We assume that there are no loops but multiple edges are allowed. The multiplicities show how intensive are the contacts. To each $i \in I$ we assign a threshold $k(i)$ that indicates that $i$ will be infected whenever (at least) $k(i)$ of his neighbors (i.e., adjacent vertices) are infected.

Given a source $S \subseteq I$, the set of infected individuals is expanded iteratively until the limit (i.e., the closure $cl(S)$) is achieved. Without loss of generality, we can assume that $k(i) \leq deg(i)$ for every $i \in I$, since otherwise $i$ cannot be infected at all and hence, (s)he can be excused. Since the thresholds $k(i)$ may vary with $i \in I$, we can model a situation when some individuals are more vulnerable than others.

Let us remark that within model $M_G$ target $T_I$ can be easily reduced to $T_q$ as follows. Given a graph $G = (I, E)$, let us add to $I$ one new vertex $q$ and an edge $(i, q)$ between $i$ and $q$ for each $i \in I$. Then let us set $k(q) = deg(q) = |I|$. Clearly, given a source $S \subseteq I$, the queen $q$ is infected in the obtained graph $G' = (I', E')$ if and only if the whole set $I$ is infected in $G$.

**Model** $M_D$. Given a directed graph (digraph) $G = (I, E)$ on the ground set $I$ and a positive integral $k(i)$ for each $i \in I$, we assume that an individual $i$ will be infected whenever at least $k(i)$ of his predecessors are infected; in other words, if there is a set of infected individuals $I' \subseteq I$ such that $|I'| \geq k(i)$ and there is an arc $(i', i)$ from $i'$ to $i$ in $E$ for each $i' \in I'$. Without loss of generality we can assume that $k(i) \leq indeg(i)$ must hold for every $i \in I$, since otherwise $i$ can be excused.

This model $M_D$ may reflect an assymmetry between the individuals. Formally, $M_D$ generalizes $M_G$. Indeed, given a graph $G = (I, E)$, let us substitute each edge $(u, v) = e \in E$ by two oppositely directed arcs $(u, v) = e'$ and $(v, u) = e''$ and we obtain a digraph $G' = (I, E')$ that represents an equivalent model.

Let us also remark that for digraphs targets $T_I$ and $T_q$ are obviously equivalent. The reduction of $T_I$ to $T_q$ was described for graphs and it holds for digraphs as well. Now we reduce target $T_q$ to $T_I$ as follows. Given a digraph $G = (V, E)$ and $q \in I$, let us choose an integer $m > |I|$ and for each $i \in I \setminus \{q\}$ add $m$ parallel arcs $(q, i)$ from $q$ to $i$. Clearly, in the obtained graph $G' = (I, E')$ the whole set $I$ is infected if and only if $q$ is infected in the original digraph $G$.

**Model** $M_R$**; transmission rules (implications) and Horn DNFs.** More generally, the rules of transmission can be given explicitly by a set of implications. For example,

$i_5 \in I$ is infected whenever $i_1, i_2, i_3, i_4 \in I$ are all infected;

$i_4$ is infected whenever $i_1, i_2, i_3$ are all infected;

$i_3$ is infected whenever $i_1, i_2$ are both infected;

$i_2$ is infected whenever $i_1, i_3$ are both infected;

$i_1$ is infected whenever $i_2, i_3$ are both infected.

These transmission rules are summarized by the following pure Horn DNF:

$$D_{TR} = x_{i_1}x_{i_2}x_{i_3}x_{i_4}\overline{x}_{i_5} \vee x_{i_1}x_{i_2}x_{i_3}\overline{x}_{i_4} \vee x_{i_1}x_{i_2}\overline{x}_{i_3} \vee x_{i_2}x_{i_3}\overline{x}_{i_1} \vee x_{i_3}x_{i_1}\overline{x}_{i_2}.$$

Let us recall that $D$ is a (pure) Horn DNF if each its monomial contains at most one (respectively, exactly one) negation.

In Section 3 we will explain this approach with more details. Now let us notice that in the above example the following three properties of a source $S \subseteq I = \{i_1, i_2, i_3, i_4\}$ are equivalent:

(i) the whole set $I$ is infected by $S$;

(ii) the queen $q = i_5$ is infected by $S$;

(iii) $S$ contains at least one of the following three sets (minimal sources) $S_1 = \{i_2, i_3\}$, $S_2 = \{i_1, i_3\}$, or $S_3 = \{i_1, i_2\}$.

In general, let us assign a hypergraph $H_i = (I \setminus \{i\}, E_i$ to each individual $i \in I$ and assume that $i$ is infected whenever there is a fully infected edge $e \in E_i$. Thus, the family of hypergraphs $R = \{H_i \mid i \in I\}$ defines a transmission model $M_R$.

In Section 3, we assign a Horn DNF to $R$ as follows:

$$D_{TR} = \bigvee_{e \in E_i} (\overline{p} \bigwedge_{i \in e} x_i)$$

Given a source $S \subseteq I$ of originally infected individuals, we expand $S$ iteratively by applying the rules of $R$ until we get the limit set $T(S) = cl_R(S)$.

We will consider the same three types of targets: $T_I, T_q$, and $T_H$. Target $T_I$ can be reduced to $T_q$ similarly. Given a set of individuals $I$ and a family of rules $\mathcal{R} = \{H_i, i \in I\}$, let us add one new individual $q$ to $I$ and and expand $\mathcal{R}$ to $\mathcal{R}'$ by adding one more hypergraph $H_q$ that consists of one edge $I$; in other words,

$q$ is infected whenever the whole set $I$ is infected.

Obviously, $q$ is infected by a source $S \subseteq I$ in $R'$ if and only if the whole set $I$ is infected by $S$ in $R$.

In Section 3 to a target hypergraph $H = (I, E)$ we will assign another Horn DNF

$$D_{TA} = \bigvee_{e \in E} (\overline{q} \bigwedge_{i \in e} x_i)$$

and show that all minimal sources that achieve the target are in one-to-one correspondence with the minimal implicants of the pure Horn DNF $D_{TR} \vee D_{TA}$. Furthermore, these implicants can be efficiently generated by iterative applying consensus rules.

It is easy to see that model $M_R$ generalizes $M_D$ (that in its turn is more general than $M_G$). Indeed, given a digraph $G = (I, E)$ to each individual $i \in I$ let us assign the family of $m_i = \binom{indeg(i)}{k(i)}$ implications: $i$ is infected whenever (at least) $k_i$ of his predecessors are infected. Obviously, the obtained family $R$ and the digraph $G$ define the same model for any target.

Let us remark however that input sizes of these two models are polynomially equivalent only if all thresholds $k(i)$ are bounded. Yet, if $k(i)$ is a part of the input of the model (Di)Graph then the number of implications $m_i$ can be exponential in $|I|$.

Let us also remark that $M_G$ and $M_D$ correspond only to some very special cases of $M_R$. If a family $\mathcal{R} = \{H_i \mid i \in I\}$ is obtained from a (di)graph $G$ then each hypergraph $H_i$ consists of $m_i = \binom{\ell(i)}{k(i)}$ edges of size $k(i)$; these edges are all subsets of a set $P(i) \subseteq I$, where $P(i)$ is the set of all neighbors (predecessors) of $i$ in $G$ and $\ell(i) = |P(i)|$.

For example, $R$ can not contain two rules

$i_5$ is infected whenever $i_1, i_2$ are both infected; and

$i_5$ is infected whenever $i_2, i_3, i_4$ are both infected

since the sets $\{i_1, i_2\}$ and $\{i_2, i_3, i_4\}$ are of different cardinalities. Furthermore, if $R$ contains the rules

$i_4$ is infected whenever $i_1, i_2$ are infected; and

$i_4$ is infected whenever $i_1, i_3$ are infected

then $R$ must also contain the rule

$i_4$ is infected whenever $i_2, i_3$ are infected.

## 1.4 Main results

Given a transmission model $M$ and target $T$ we are looking for minimal sources $S$ that infect $T$. We shall consider two problems:

**”Optimization”**: find a source of minimum size that infects $T$.

**Generation**: output all inclusion minimal sources that infect $T$.

Optimization is hard already for graphs and for digraphs it remains hard even if $k_i = deg(i)$ for one $i \in I$ and $k_i = 1$ for all others. More precisely, the following claim holds.

**Theorem 1.** *Given a (di)graph $G = (I, E)$ and positive integral $t$ and $k(i)$ for each $i \in I$, the following two decision problems are NP-complete: if there is a fatal source $S$ of size (at most) $t$ that infects (i) the whole set $I$ or (ii) a given individual $q \in I$. Both problems, (i) and (ii), remain NP-complete $k(i)$ take only values $deg(i)$ or 1 (that is, $m_i = \binom{deg(i)}{k(i)} = deg(i)$). In case of digraphs (i) and (ii) remain NP-complete even when $k(i) = deg(i)$ only for one $i \in I$ and $k(i) = 1$ for all others. In case of graphs (i) and (ii) remain NP-complete even when $deg(i) \leq 3$ for all $i \in I$.*

Let us recall that $q \notin S$ in case (ii).

Let us also note that problems (i) and (ii) remain NP-complete for a more general transmission model Implications. Indeed, to a given a (di)graph we assign a list of implications whose size is bounded by $\Sigma = \sum_{i \in I} m_i$, where $m_i = \binom{deg(i)}{k(i)}$. Furthermore, we can assume that $\Sigma < 2|E|$, since for (di)graphs problems (i) and (ii) are NP-complete already for $m_i = deg(i)$ or 1 for each $i \in I$. Thus, the input size for Implications is polynomially bounded in the input size for (Di)graphs.

However, for generating problems the exponential increase of the input size is possible and it does make a difference between the models (Di)Graphs and Implications. In the latter case all minimal fatal sources can be generated in incremental polynomial time even if the target is an arbitrary hypergraph.

**Theorem 2.** *Given a family of rules $R$ over a set of individuals $I$ and a family of target sets $H \subseteq 2^I$, all inclusion minimal sources $S \subseteq I$ can be generated in incremental polynomial time.*

On the other hand, if the transmission model $M_G$ or $M_D$ is a (Di)Graph $G = (I, E)$ and neither $k_i$ nor $deg(i) - k(i)$ are polynomially bounded in $|I|$ then generating all minimal fatal sources becomes hard already for simple targets $T_I$ and $T_q$. Moreover, for digraphs these problems remain hard even if all thresholds $k(i)$, except one, take values 1 and 2. More precisely, the following claim holds.

**Theorem 3.** *Given a (di)graph $G = (I, E)$, queen $q \in I$, positive integral thresholds $t \leq |I|$ and $k(i)$ for each $i \in I$, and also a collection $\mathcal{S}$ of minimal fatal sources $S \subseteq I$ that infect (i) the whole set $I$ or (ii) the queen $q$, it is an NP-complete to decide whether there exists one more minimal fatal source $S \notin \mathcal{S}$ or the collection $\mathcal{S}$ is complete. Moreover, for digraphs the problem remains NP-hard even if all thresholds $k(i)$, except one, take only values 1 and 2.*

To summarize, let us count all problems. We consider three transmission models $M_G, M_D, M_R$, three targets $T_I, T_q, T_H$, and two goals: "optimization" and generation; totally $3 \times 3 \times 2 = 18$ problems. Three of them, generation problems with model $M_R$ are tractable, incremental polynomial, for all three possible targets; the remaining sixteen problems are NP-hard.

## 1.5 Generating optimal vaccination strategies

Let us say also a few words about possible vaccination strategies. Given a family (hypergraph) $\mathcal{S}$ of all fatal sources, let us consider so-called *dual (or transversal)* hepergraph $\mathcal{S}^d$ defined on the same ground set $I$ and whose edges are all minimal transversals to the edges of $\mathcal{S}$.

Let us note that in fact the hypergraph $\mathcal{S}$ is given implicitly, by a transmission model and target. The problem of simultaneous generating of two implicitly given dual hypergraphs was considered in [5]. The main result of this paper is positive: an incremental quasi-polynomial ($N^{o(\log N)}$) algorithm for joint generation of any dual pair $\mathcal{S}$ and $\mathcal{S}^d$ given by a (quasi-) polynomial oracle.

(Let us recall that a function $f(N) = 2^{polylog(N)} = 2^{log^c N}$ is called quasi-polynomial in $N$, where $c$ is a constant.)

Thus, generating all "inclusion minimal" vaccination strategies is a tractable (incremental quasi-polynomial) generation problem. Needless to say that "optimization", that is, getting a transversal to $\mathcal{S}$ of minimum cardinality, is NP-hard. Indeed, obviously, the complement to a (minimum) transversal to $\mathcal{S}$ is a (maximum) stable set of $\mathcal{S}$. Thus, the problem is reduced to the classical "Stability Number", which is NP-complete already for graphs.

# 2 Proofs of Theorems 1 and 3

**Proof of Theorem 1**. We reduce both cases (i) and (ii) for graphs and digraphs from a well-known NP-complete problem Minimal Vertex Cover (MVC). For an arbitrary graph $G = (I, E)$ a subset $S \subseteq I$ is called a *vertex cover* if each edge $e \in E$ is adjacent to an $i \in I$. A vertex cover $S$ is called *minimal* if it is not a proper subset of some other vertex cover. Given a graph $G = (V, E)$ and a positive integral threshold $t$, it is an NP-complete problem to decide if there is a vertex cover in $G$ of cardinality (at most) $t$. Moreover, this problem is NP-complete already for cubic ($deg(i) \leq 3$ for all $i \in I$) and planar graphs.

**(i) Target $T_I$ in model $M_G$**. Given an arbitrary graph $G = (I, E)$, let us assign to it a bipartite graph $G' = (I', E')$ where $I' = I \cup E$ and $i \in I$ is incident to $e \in E$ in $G'$ if and only if $i$ is adjacent to $e$ in $G$. Furthermore, let us set $k(e) = 1$ for each $e \in E$ and $k(i) = deg(i)$ for each $i \in I$. It is easy to see that each vertex cover $S \subseteq I$ in $G$ infects the whole set $I'$ in $G'$. Indeed, every $e \in E$ will be infected, by the definition of a vertex cover. Then, each $i \in I$ will be infected too. Furthermore, let us notice that it is irrelevant to include an individual $e \in E$ in a source $S$. Indeed, if instead of $e$ we include an $i \in I$ adjacent to $e$ then $e$ will be infected anyway. This concludes the proof of (i).

**(ii) target $T_q$ in model $M_G$**. In general, target Queen can be standardly reduced to target All; see Introduction. Let us add to $G' = (I', E')$ one new vertex $q$, connect it to each $i \in I$ and to each $e \in E$, and set $k(q) = deg(q) = |I| + |E|$. Obviously in the obtained graph $G'' = (I'', E'')$ a fatal source $S \subseteq I'' \setminus \{q\} = I' = I \cup E$ infects $q$ if and only if $S$ infects the whole set $I' = I \cup E$ in $G'$.

**Targets $T_I$ and $T_q$ for model $M_D$**. Given an arbitrary graph $G = (I, E)$, let us assign to it a digraph $G' = (I', E')$ as follows: $I' = I \cup E \cup \{q\}$ and $(i, e) \in E'$ if and only if $i$ is adjacent to $e$ in $G$; furthermore, $(e, q) \in E'$ for all $e \in E$ and $(q, i) \in E'$ for all $i \in I$. Thresholds $k(i)$ for $i \in I'$ are as follows: $k(i) = k(e) = 1$ for all $i \in I,\ e \in E$ and $k(q) = deg(q)$.

The same arguments as before show that the following three properties of a source $S \subseteq I$ are equivalent: (a) $S$ infects the whole set $I'$, (b) $S$ infects $q$, and (c) $S$ is a vertex cover in $G$.

It is also easy to see that $q$ herself infects the whole set $I'$.

Finally, it is pointless to include an individual $e \in E$ in a minimum source $S$. Indeed, if instead of $e$ we include an $i \in I$ adjacent to $e$ in $G$ then $e$ will be infected in $G'$ anyway. This concludes the proof. $\square$

**Proof of Theorem 3**. For both graphs and digraphs we reduce our decision problem from a well-known NP-complete problem: Stability Number of a Graph. For an arbitrary graph $G = (I, E)$ a subset $S \subseteq I$ is called a *stable set* if $S$ does not contain an edge, that is, $u, v \in S$ fot no edge $(u, v) = e \in E$. A stable set $S$ is called *maximal* if it is not a proper subset of some other stable set. Obviously, the complement of a (maximal) stable set in $I$ is a (minimal) vertex cover. Given a graph $G = (V, E)$ and a positive integral threshold $t$, it is an NP-complete problem to decide whether there is a stable set $S$ in $G$ of size (at least) $t$.

**Proof for digraphs, model $M_D$.** Given an arbitrary graph $G = (I, E)$, let us assign to it a digraph $G' = (I', E')$ where $I' = I \cup E \cup \{p\} \cup \{q\}$ and $E'$ are defined as follows: $(i, e)$ is an arc in $E'$ if and only if $i$ and $e$ are adjacent in $G$; next, $(i, p)$ and $(q, i) \in E'$ for each $i \in I$; furthermore, $(p, q)$ and $(e, q) \in E'$ for each $e \in e$. The thresholds $k(i)$ for $i \in I'$ are defined as follows: $k(i) = 1$ for $i \in I \cup \{q\}$, $k(i) = 2$ for $i \in E$, and $k(p) = t$. Let us note that $k(i) = 1$ or 2 for all $i \in I'$, except $i = p$.

It is easy to see that a single-individual-source $S = \{i\}$ will infect the whole set $I'$ whenever $i \in E \cup \{p\} \cup \{q\}$. Indeed, each $i \in E$ or $i = p$ will infect $q$, in her turn $q$ will infect the whole set $I$, and then $I$ will infect $E \cup \{p\}$. It is also easy to see that a pair $S = \{u, v\}$ will infect the whole set $I'$ whenever $(u, v) \in E$. Indeed, in this case $S = \{u, v\}$ will infect $e \in E$ such that $e = (u, v) \in E$, in his turn $e$ will infect $q$, then $q$ will infect $I$, and then $I$ will infect $E \cup \{p\}$. Thus we got a collection $\mathcal{S}$ of minimal sources that infect the whole set $I'$, or which is equivalent, the queen $q$. Yet, there may be more minimal sources with this property. It is easy to see that such a source exists if and only if graph $G = (V, E)$ contains a stable set $S \subseteq I$ of cardinality $|S| = t$. Indeed, in this case $S$ will infect $p$, in his turn $p$ will infect $q$, in her turn $q$ will infect the whole set $I$, and finally, $I$ will infect $E$. However, it is NP-complete to decide whether $G$ contains such a set $S$.

**Proof for graphs, model $M_G$.** Given an arbitrary graph $G = (I, E)$, let us assign to it a digraph $G' = (I', E')$ where $I' = I \cup E \cup \{p\} \cup \{q\}$ and $E'$ are defined as follows: $(i, e)$ is an arc in $E'$ if and only if $i$ and $e$ are adjacent in $G$; next, $(i, p)$ and $(q, i) \in E'$ for each $i \in I$; furthermore, in $E'$ there are $m$ multiple edges between $p$ and $q$ and between $e$ and $q$ for each $e \in E$, where $m$ is a positive integral such that $m > |I|$. The thresholds $k(i)$ for $i \in I'$ are defined as follows: $k(i) = 1$ for $i \in I$ ; $k(i) = 2$ for $i \in E$ , $k(p) = t$, and and $k(q) = m$.

It is easy to see that a source $S = \{i\}$ of cardinality 1 will infect the whole set $I'$ whenever $i \in E \cup \{p\} \cup \{q\}$. Indeed, each $i \in E$ or $i = p$ will infect $q$, in her turn $q$ will infect the whole set $I$, and then $I$ will infect $E \cup \{p\}$. It is also easy to see that a pair $S = \{u, v\}$ will infect the whole set $I'$ whenever $(u, v) \in E$. Indeed, in this case $S = \{u, v\}$ will infect $e \in E$ such that $e = (u, v) \in E$, in his turn $e$ will infect $q$, then $q$ will infect $I$, and then $I$ will infect $E \cup \{p\}$. Thus we got a collection $\mathcal{S}$ of minimal sources that infect the whole set $I'$, or which is equivalent, the queen $q$. Yet, there may be more minimal sources with this property. It is easy to see that such a source exists if and only if graph $G = (V, E)$ contains a stable set $S \subseteq I$ of cardinality $|S| = t$. Indeed, in this case $S$ will infect $p$, in his turn $p$ will infect $q$, in her turn $q$ will infect the whole set $I$, and finally, $I$ will infect $E \cup \{p\}$. However, it is NP-complete to decide whether $G$ contains such a set $S$.

Let us note finally that no subset $S \subseteq I$ can infect $q$ directly, since $m > |I|$. $\square$

# 3 Target and transmission pure Horn DNFs. Proof of Theorem 2

Let us assign a DNF

$$D_{TR} = \bigvee_{j \in J} (\overline{x}_{i_j} \bigwedge_{i \in K_j} x_i)$$

to a given set of transmission rules $R_j = \{r_j = (K_j, i_j) \mid j \in J\}$.

Also let us assign a DNF

$$D_{TA} = \bigvee_{e \in E} (\overline{x}_q \bigwedge_{i \in e} x_i)$$

to a given family of target sets $H = (I, E)$. In this DNF, $x_q$ is a special new variable, an “artificial queen”.

Finally, let us consider disjunction of the above two DNFs, $D_T = D_{TR} \vee D_{TA}$.

By construction, $D_{TR}$, $D_{TA}$, and $D_T$ are are pure Horn DNFs, that is, each monomial contains exactly one negated literal.

Obviously, a pure Horn monomial $x_1 \dots x_k \overline{x}_{k+1}$ is equal to 1 if and only if $x_1 = \dots = x_k = 1$, while $x_{k+1} = 0$, that is, if the implication $x_1 \wedge \dots \wedge x_k \Rightarrow x_{k+1}$ fails. In other words, this implication holds if and only if $x_1 \dots x_k \overline{x}_{k+1} = 0$. Thus, we obtain a one-to one correspondence between the fatal sources for the target hypergraph $H$ and the satisfying assignments for the Boolean equation $D_T = D_{TR} \vee D_{TA} = 0$ in which $x_q = 1$.

**Proposition 1.** *A source $S \subseteq I$ will infect a target set $e \in E = E(H)$ iff the corresponding elementary conjunction $\overline{x}_q \wedge_{i \in S} x_i$ is an implicant of $D_T$.* □

Respectively, the minimal sources that infect a target set $e \in E$ are in one-to-one correspondence with the *prime* implicants of $D_T$ that contains the literal $\bar{x}_q$. An efficient (incremental polynomial) algorithm generating all prime implicants that contain a given literal of an arbitrary Horn DNF was recently obtained by Eiter and Makino in [4]. (Let us remark that in [8, 9] it was conjectured that this generating problem is intractable.) Here we outline the algorithm for the case of pure Horn DNFs which appears to be simpler than the general case.

The following Boolean identity was introduced by Blake [1]:

$$xA \vee \overline{x}B = xA \vee \overline{x}B \vee AB.$$

Here $A$ and $B$ are elementary conjunctions (monomials) that may contain negated literals. The proof is immediate. Indeed, in both cases, $x = 0$ and $x = 1$, we obtain so-called absorption identities: $B = B \vee AB$ for $x = 0$ and $A = A \vee AB$ for $x = 1$. The above Boolean identity $xA \vee \overline{x}B = xA \vee \overline{x}B \vee AB$ is the so-called *consensus* operation and it plays an important role in Boolean algebra; see e.g. [2] Chapters 1-3. In particular, it is known that, successively applying this operation (together with the absorption $A \vee AB = A$), one can get all prime implicants of a Boolean function beginning with an arbitrary its DNF. However, in general, such a generation algorithm is exponential, since consensus operation can produce a new monomial that is larger than both operands; for example,

$$\overline{x}_1 x_2 x_3 \vee x_1 x_4 \overline{x}_5 = \overline{x}_1 x_2 x_3 \vee x_1 x_4 \overline{x}_5 \vee x_2 x_3 x_4 \overline{x}_5.$$

Hence, applying successively consensus operations, one can get exponentially large intermediate DNFs, while the input and output DNFs are both small. However, in case of (pure) Horn DNFs this problem can be resolved.

First, let us note that, obviously, consensus operation respects (pure) Horn DNFs; see the above example.

Given a DNF $D$, the Boolean equation $D = 0$ is called the *tautology* problem. (The corresponding equation $C = 1$ for a CNF $C$ is called the satisfiabilty (SAT) problem; obviously, these two problems are equivalent.) Tautology is NP-complete in general but it is polynomial for (pure) Horn DNFs; see e.g. [2] Chapters 1-3. Respectively, given a DNF $D$ and an arbitrary monomial $A$, it is NP-complete to verify whether $A$ is an implicant of $D$ (that is, if $A \Rightarrow D$ holds). However, this problem becomes polynomial if $D$ is a (pure) Horn DNF. This will be the main subroutine in the following algorithm of generating all prime implicants that contain $\overline{x}_q$ for the pure Horn DNF

$$D_T = D_{TR} \vee D_{TA} = (\bigvee_{j \in J} (\overline{x}_{i_j} \bigwedge_{i \in K_j} x_i)) \quad \bigvee \quad (\bigvee_{e \in E} (\overline{x}_q \bigwedge_{i \in e} x_i)).$$

Step 1. For each monomial $B\overline{x}_q$ of $D_{TA}$ let us verify whether is it prime in $D_T$ and if not reduce the set of literals of $B$ to get a prime one $B'\overline{x}_q$. This procedure is polynomial, since $D_T$ is a (pure) Horn DNF. Note also that we do not try to reduce $B$ in all possible ways but obtain only one $B'$. Let us denote the obtained (from $D_{TA}$) new target DNF by $D^1_{TA}$. Let us remark that this DNF is a ready part of the output.

Step 2. Now, let us apply consensus procedure to all proper pairs of monomials $A$ and $B$ from $D_{TR}$ and $D^1_{TA}$, respectively. Note that each $B$ contains $\overline{x}_q$, while $A$ does not; hence, all the obtained "new" monomials will contain it. Yet, some of them may be absorbed by the "old" monomials of $D_{TA}$. Let us delete them, add the remaining ones to $D^1_{TA}$, and denote the obtained "new" target DNF by $D^2_{TA}$.

Now we repeat Step 1 with $D^2_{TA}$, etc. If after $k$ rounds all new monomials produced in the Step $2k$ are absorbed then halt and output the target DNF.

It is clear that both steps 1 and 2 take time polynomial in size of the DNF $D_{TR}$ (which does not change at all) and the current target DNF. Since after Steps $1k$ and $2k$, for each round $k$, except the last one, we produce new outputs, we conclude that the above generating algorithm is incremental polynomial.

**Remark 1.** *Word "incremental" is used here because new prime implicants are generated in time polynomial in size of the input $D_T$ and the current output, that is, target DNF $D^{1k}_{TA}$. However, this DNF may increase exponentially in $k$; see [10, 7, 6] for the precise definitions related to complexity of the generation algorithms.*

**Acknowledgements.** We are thankful to Fred Roberts for the introduction to basic mathematical models of disease transmission; some of these models we learned from his works that are not published yet, [3].

# References

[1] A. Blake, Canonical expressions in Boolean algebras, Dissertation, Dep. of Math., Univ. of Chicago, 1937. Published by the Univ. of Chicago Libraries, 1938.

[2] Y. Crama and P.L. Hammer, Boolean functions; theory, algorithms, and applications, Springer Verlag, 2007, to appear.

[3] P. Dreyer and F.S. Roberts, Graph-theoretical Threshold Models of the Spread of Disease and of Opinion, in preparation, DIMACS Center, 2007.

[4] T. Eiter and K.Makino, On computing all abductive explanations, in Proc. of AAAI-02, pp. 62-67;
See the full version in INFSYS Research Report 183-02-04, April, 2002.

[5] V. Gurvich and L. Khachiyan, On generating the irredundant conjunctive and disjunctive normal forms of monotone Boolean functions, Discrete Applied Math., 96-97 (1999) 363-373.

[6] D.S. Johnson, M. Yanakakis, and C.H. Papadimitriou, On generating all maximal independent sets, Inform. Process. Letters, 27 (1988), 119-123.

[7] E. Lowler, J.K. Lenstra, and A.H.G. Rinnooy Kan, Generating all maximal independent sets: NP-hardness and polynomial algorithms. SIAM J. of Computing 9 (1980), 558-565.

[8] B.Selman and H.J. Levesque, Abductive and default reasoning, a computational core, in Proc. of AAAI-90, pp.343-348.

[9] B.Selman and H.J. Levesque, Support set selection for abductive and default reasoning, Artif. Intell. 82, (1996), 259-272.

[10] L.G. Valiant, The complexity of enumeration and reliability problems, SIAM J. of Computing 8 (1979), 410-421.