The Complexity of Boolean Failure Identification

Nicola Galesi[1,*], Fariba Ranjbar[2]

[1]*Department of Computer, Control and Management Engineering (DIAG), Sapienza Università Roma, Rome, Italy*
[2]*Department of Business and Management, Luiss Guido Carli, Rome, Italy*

Abstract

We consider the problem of identifying failure nodes in networks under the *Boolean Network Tomography* (BNT) approach, which is based on end-to-end measurements routed in a network along paths and producing a boolean (failure/not-failure) outcome[1]. Such end-to-end measurements paths are usually described by an incidence boolean matrix $\mathbb{M}$ with m rows (the measurements paths) and n columns (the nodes of the network). A key notion used in practice in this approach is that of *k-identifiability*. Loosely speaking, a set of m boolean measurements paths over n nodes is k-identifiable, where k is a non-negative integer, if, whenever there are fewer than $k+1$ failures, it is always possible to identify unambiguously and uniquely which nodes are failing.

Following the focus of some recent results analyzing maximal identifiability from a theoretical point of view [1, 2, 3, 4], this work establishes the complexity of the optimization problem that determines the *maximal* k for which a set of measurement paths is k-identifiable (MID). We prove that such *a problem* is NP-hard by a reduction from the *Minimum Hitting Set* problem. To our knowledge the NP-hardness of MID and the relation with the Minimum Hitting Set problem are new and not known before.

We further consider the following extremal combinatoric question: given the number n of nodes of the network and a non-negative integer value k for the identifiability, what is the minimal number m of measurement paths over the n nodes to consider in such a way that the maximal identifiability *value* is at least k? A folklore result shows that to have maximal identifiability at least 1, then $m \geq \log(n+1)$ (or, equivalently, that if $n > 2^m - 1$, then the maximal identifiability is less than or equal 0). In this work we answer this question for each $n \in \mathbb{N}$ and for each $k \geq 2$, proving that, there exists a constant C such that if $n > Cm^{1+\frac{m}{k-1}}$, then the maximal identifiability value is strictly smaller than k (and when $k=2$, $n > Cm^m$ suffices). To show these results we consider two notions that we prove to be equivalent to k-identifiability: one is from the field of *non-adaptive group testing* (NAGT) and the other is the notion of *union-free* set families [5]. The connection between identifiability and group testing was mentioned in [6]; we make this connection precise towards a solution of our problem.

Keywords

Boolean network tomography, k-identifiability, Union free sets, Group testing, Hitting set problem

1. Introduction

Network Tomography is a general inference technique based on end-to-end measurements aimed to extract not only internal network characteristics such as link delays and link loss rates but also defective items. In *Boolean Network Tomography* (BNT) the outcome of the measurements

[1]In this paper we consider only identifying failure nodes but all the results work for links as well.

ICTCS'24: Italian Conference on Theoretical Computer Science, September 11–13, 2024, Torino, Italy
*Supported by the project PRIN 2022 "Logical Methods in Combinatorics" N. 2022BXH4R5 of the Italian Ministry of University and Research (MIUR).
nicola.galesi@uniroma1.it (N. Galesi); fariba.ranjbar@luiss.it (F. Ranjbar)
0000-0002-8522-362X (N. Galesi); 0000-0001-6432-3683 (F. Ranjbar)

is a Boolean value. Duffield in [7], introduced the Boolean network tomography approach to identify sets of failure links in networks and later this was also applied for node failure identification, in the works [8, 9, 10, 1, 3, 2, 4]. The BNT approach is rather simple: each measurement path is routed with a suitable data packet and the received data at *the end* of the path is one bit capturing the presence or the absence of failures along the path: for a *failure* the output bit is intended to be 1 and for a *properly working* path the output bit is 0. Such a method can be applied to detect both node and link failures in a network. In this work we always deal with nodes, but all the results can be similarly applied to capture defective links.

Since in each path the outcome of a measurement indicates only whether a failure occurred somewhere in the path, in the BNT approach *the position of the nodes in the paths is not taken into account and paths are regarded as sets of nodes.*

The problem of identifying failing nodes is approached by studying the solutions $\vec{x}$ of a Boolean system $\mathbb{M}\vec{x} = \vec{b}$, where $\mathbb{M}$ is the incidence $\{0, 1\}$-matrix of the m measurement paths over the n nodes, $\vec{b}$ is the vector of length m of the Boolean outcomes of the measurement paths and $\vec{x}$ is the Boolean vector of length n indicating whether each node is failing or not. To be consistent with *the* BNT terminology, in this work we keep the names *nodes* and *paths* respectively for the n columns and the m rows of $\mathbb{M}$. The challenge of localizing failure nodes is that different sets of failure nodes can produce the same measurement along the paths and so are indistinguishable from each other with *only* using the measurements. This led to the following question: given the set of paths, what is the *maximal* set of defective nodes one can hope to identify unambiguously?

k-identifiability (for a set $\mathbb{M}$ of m paths over n nodes) states that any two distinct node sets U and W of size at most k can be separated by at least *one* path in $\mathbb{M}$, that is, there is at least *one* path touching nodes of only one of them. It was observed in [9] that having maximal k-identifiability for a set of paths $\mathbb{M}$ ensures that if there are at most k failing nodes, then these nodes can be identified unambiguously using the BNT approach: this was exactly what was needed towards node failure identification. Concretely, we are interested in understanding the maximal k such that $\mathbb{M}$ is k-identifiable. This measure - which we call $\mu(\mathbb{M}, n, m)$ - was introduced and investigated especially from an applied perspective in the works [8, 9, 10]. The book [6] presents a comprehensive treatment of k-identifiability and boolean network tomography.

Identifiability however is a precise combinatorial definition and therefore it was interesting to research it also from a theoretical and combinatorial perspective complementing the applied results. In a series of recent works [1, 2, 3, 4] focusing on theoretical aspects of identifiability we studied the relations of maximal k-identifiability with the topology of the network and some of its structural properties, like vertex connectivity.

This work contributes to the research line of boolean network tomography and maximal k-identifiability. We consider two problems: the first question we approach is that of understanding the computational complexity of maximal k-identifiability. We consider the following optimization problem: given a set $\mathbb{M}$ of m measurement paths over n nodes, determine the *maximal* k for which $\mathbb{M}$ is k-identifiable (MID).

We consider the question of proving that MID is NP-hard. In Theorem 3.4 we prove this result using a reduction from the well-known Minimum Hitting Set problem (MHS). The complexity of MID was not known *before* and to our knowledge this is the first time that the optimization problem of *maximizing* k-identifiability is shown to be strictly related to the Minimum Hitting Set problem. We complement the MID hardness result: first we prove that the problem of deciding whether a set of paths is *not* k-identifiable, for a given $k \in \mathbb{N}$, is in NP (Theorem 3.7). Second we prove that if MHS is computable in polynomial time, then also MID is computable in polynomial time (Theorem 3.6). Together with the hardness result this establish a strict relationships between Minimum Hitting Set problem and the MID problem. This relation might be interesting from an applied point view for algorithms approximating k-identifiability: indeed it is known that there are c-approximation algorithms for the Minimum Hitting Set problem where edges have cardinality at most c [11, 12], which is the case of our reduction in theorem 3.6 when the paths in $\mathbb{M}$ has length at most c. The reduction used in Theorem 3.6 to prove the hardness result works by first reducing MID to a maximal k-identifiability problem *scaled* to each single node: a node u is k-identifiable in $\mathbb{M}$ if any two node sets U and W of size at most k and differing on u (that is such that u is only in one of them) are separated by at least a path in $\mathbb{M}$ (that is there exists a path intersecting one of them but not the other).

The second problem we face is an extremal combinatoric question: imagine a network topology with n nodes is given, and let us say we establish a non-negative integer value k for the number of defective nodes we aim to identify: what is the minimal number m of measurement paths over the n nodes we have to consider in such a way that the maximal identifiability is at least k ? This is an extremal question on the incidence matrix $\mathbb{M}$. The answer to this question for $k = 1$ follows immediately by the pigeonhole principle argument: if we have more than 2^m binary strings of length m there are at least two identical strings. The result for $k = 1$ can be formalized as follows (see Lemma 4.1 for the details): if $\mu(\mathbb{M}, m, n) \geq 1$, then $m \geq \log(n + 1)$, or, equivalently, saying that if $n > 2^m - 1$, then $\mu(\mathbb{M}, m, n) < 1$. Nothing is known for a generic $k > 1$ and here we answer this question for any $n \in \mathbb{N}$ and for any $k \in \mathbb{N}$, $1 < k \leq n$ in Theorem 4.7.

It is not difficult to see, and *it* is also explicitly mentioned in [6], that k-identifiability is related to similar concepts in *Group Testing* [13]. Approaching our second question we first make this connection precise to solve our question. We identify precisely in [13] a central notion defined in *non-adaptive group testing* (NAGT) (the so-called $\bar{k}$-separability) which we prove to be equivalent to k-identifiability (see Definition 2.3 and Lemma 2.4 for the precise definition of $\bar{k}$-separability and the proof of the equivalence). Using a known theorem in NAGT (Theorem 2.6) we can answer our question but only for $k + 1$ (Theorem 4.2). However to prove the result for k, using known results in NAGT is not sufficient, and new techniques are needed. Towards this goal and looking again at $\mathbb{M}$ as a hypergraph, we observe that k-identifiability is strictly related to the combinatorial notion called *union-free families of sets*. This was a combinatorial notion on sets introduced in [5] and studied in [5, 14] which we observe to be strictly related to k-identifiability (Theorem 4.4). Using a recent result on *uniform union-free* families, proved in [14], we can eventually fully answer to our question: Theorem 4.7 states that there is a constant C such that

$$n > Cm^{1+\frac{m}{k-1}} \Rightarrow \mu(\mathbb{M}, m, n) < k \quad \text{if } k \geq 3 \qquad \text{and} \qquad n > Cm^m \Rightarrow \mu(\mathbb{M}, m, n) < 2.$$

Finally, in the last subsection, we use these results to provide bounds on the maximal number of k-identifiable nodes, for each $k \in \mathbb{N}$.

2. Preliminary definitions and results

Let $n, k \in \mathbb{N}$, $k \leq n$. $[n] = \{1, \ldots, n\}$ and $\binom{[n]}{k}$ is the set of subsets of $[n]$ of size k. 2^A is the set of subsets of the set A.

Let n and m be positive integers. Following previous works on Boolean network tomography [6] we see *a set* $\mathbb{M}$ *of* m *paths over nodes in* $[n]$ as a $\{0,1\}$-matrix of m rows and n columns[1]. On $\mathbb{M}$ we use the following notations:

1. We see $\mathbb{M}$ as a collection of n m-bit vectors, that is the columns of the matrix $\mathbb{M}$, that are vectors of length m, all different from the m-bit zero vector[2].
2. For $u \in [n]$, $\mathbb{M}(u)$ is the set of rows $r \in [m]$, such that u belongs to r, that is such that $\mathbb{M}[r, u] = 1$. In the simpler BNT terminology, we say that $\mathbb{M}(u)$ is the set of paths in $[m]$ passing through (or intersecting) the node u. If $U \subseteq [n]$, then $\mathbb{M}(U) = \bigcup_{u \in U} \mathbb{M}(u)$ so, if $U \subseteq W$, then $\mathbb{M}(U) \subseteq \mathbb{M}(W)$. In the following we will use the BNT terminology talking of nodes and paths in $\mathbb{M}$ to mean columns and rows of $\mathbb{M}$ and see them simply as sets of Boolean values.

2.1. Identifiability and non-adaptive group testing

We consider the following definition given in [8].

Definition 2.1. *(Identifiability) A Boolean matrix* $\mathbb{M}$ *over* m *rows and* n *columns is* k-identifiable *if for all* $U, W \subseteq [n]$ *such that* $|U|, |W| \leq k$ *and* $U \neq W$, *it holds that* $\mathbb{M}(U) \neq \mathbb{M}(W)$. *We denote by* $\mu(\mathbb{M}) = \mu(\mathbb{M}, m, n)$ *the maximal* $k \leq n$ *such that* $\mathbb{M}$ *is* k*-identifiable.*

We observe next that k-identifiability is strictly related to some notions in non-adaptive group testing. Consider the following definitions given in [13]. First, notice that the union (or Boolean sums) of any k columns in a Boolean matrix $\mathbb{M}$ is, the bitwise OR that is a binary operation that takes two bit patterns of equal length and performs the logical inclusive OR operation on each pair of corresponding bits. The result in each position is 0 if both bits are 0, while otherwise the result is 1. Moreover, in terms of the union (or Boolean sums) of columns in a Boolean matrix $\mathbb{M}$ over m rows and n columns, in the definition 2.1 we have for $U \subseteq [n]$ with $|U| \leq k$, $\mathbb{M}(U) = \bigcup_{u \in U} \mathbb{M}(u)$ which is the union (or Boolean sums) of up to k columns.

Definition 2.2. *(Disjunctness) A Boolean matrix* $\mathbb{M}$ *with* m *rows and* n *columns is called* k-disjunct *if the union (or Boolean sums) of any* k *columns in* $\mathbb{M}$ *does not contain any other column in* $\mathbb{M}$. *Notice that this also implies that the union of any up to* k *columns does not contain any other column.*

Definition 2.3. *(k-separability and $\overline{k}$-separability) A Boolean matrix* $\mathbb{M}$ *of* m *rows and* n *columns is called* k-separable *(respectively* $\overline{k}$-separable) if the unions (or Boolean sums) of k columns (respectively of up to k columns) are all distinct.

[1]Notice that this encoding does not keep track of the order of the nodes in the path, but this is usual in Boolean network tomography approaches for identifying failing nodes since the position of the nodes in the paths is not taken into account and hence paths are regarded as sets of nodes.

[2]Notice that when $\mathbb{M}$ is a real set of paths, the condition means that each node is used in at least one path.

Lemma 2.4. *A Boolean matrix $\mathbb{M}$ with m rows and n column is $\bar{k}$-separable if and only if it is k-identifiable*

Proof. First assume that $\mathbb{M}$ is k-identifiable. Then for all $U, W \subseteq [n]$ such that $|U|, |W| \leq k$ and $U \neq W$, we have $\mathbb{M}(U) \neq \mathbb{M}(W)$. Moreover $\mathbb{M}(U) = \bigcup_{u \in U} \mathbb{M}(u)$ which is the union of up to k columns. The unions of up to k columns are thus all distinct and $\mathbb{M}$ is $\bar{k}$-separable. Similarly if $\mathbb{M}$ is $\bar{k}$-separable, the unions of up to k columns are all distinct i.e., for all $U, W \subseteq [n]$ such that $|U|, |W| \leq k$ and $U \neq W$, $\bigcup_{u \in U} \mathbb{M}(u) = \mathbb{M}(U) \neq \mathbb{M}(W) = \bigcup_{w \in W} \mathbb{M}(w)$. Therefore $\mathbb{M}$ is k-identifiable. □

A close relation between disjunctness and separability of $\mathbb{M}$ was proved in [13] (Lemma 7.2.2 and Lemma 7.2.4)

Lemma 2.5 ([13]). *For a $\{0, 1\}$-matrix $\mathbb{M}$ of m rows and n columns:*

1. *if $\mathbb{M}$ is k-disjunct, then $\mathbb{M}$ is $\overline{k}$-separable.*
2. *if $\mathbb{M}$ is $\overline{(k+1)}$-separable, then $\mathbb{M}$ is a k-disjunct.*

Let $t(k, n)$ denotes the minimum number of rows for a k-disjunct matrix with n columns. We have the following theorem (Theorem 7.2.13 in [13]):

Theorem 2.6 ([13]). *For k fixed and $n \to \infty$, there is a constant C_k such that*

$$t(k, n) \geq C_k(1 + o(1)) \log n.$$

3. NP-Hardness of maximal k-identifiability

We have seen that k-identifiability is a reasonable combinatorial notion. In this section we clarify what is its computational complexity. We consider the following optimization problem:

MID:
Input: A Boolean $m \times n$ matrix $\mathbb{M}$.
Output: The maximal $k \leq n$ such that $\mathbb{M}$ is k-identifiable.

Let $\mathbb{M}$ be a $m \times n$ Boolean matrix. We say that two sets of nodes $U, W \subseteq [n]$ *differ on* u if u belongs to exactly one of them, that is $U \cap \{u\} \neq W \cap \{u\}$. Furthermore we say that a path $p \in [m]$ *separates* U *and* W *in* $\mathbb{M}$ if p belongs to exactly one between $\mathbb{M}(U)$ and $\mathbb{M}(W)$. The definition of k-identifiability can be scaled to nodes $u \in [n]$ as follows:

Definition 3.1. *(k-identifiable nodes) A node $u \in [n]$ is* k-identifiable in $\mathbb{M}$, *if for all $U, W \subseteq [n]$ of size at most k differing on u, it holds that $\mathbb{M}(U) \neq \mathbb{M}(W)$.*

Scaling identifiability to nodes does not affect the k-identifiability of the whole $\mathbb{M}$. The next theorem has been proven in [9], but we decide to show the proof again for the sake of completeness.

Theorem 3.2. *([9]) Let $\mathbb{M}$ be a set of m paths over n nodes. $\mathbb{M}$ is k-identifiable if and only if every node in $[n]$ is k-identifiable in $\mathbb{M}$.*

Proof. First assume that $\mathbb{M}$ is k-identifiable and let $u \in [n]$. Since $\mathbb{M}$ is k-identifiable, for all $U, W \subseteq [n]$ of size at most k such that $U \neq W$, $\mathbb{M}(U) \neq \mathbb{M}(W)$ holds. Therefore for all $U, W \subseteq [n]$ such that $|U|, |W| \leq k$ and differing on u, we also have $\mathbb{M}(U) \neq \mathbb{M}(W)$. This proves that u is k-identifiable.

Now let every node in $[n]$ be k-identifiable in $\mathbb{M}$. Assume by contradiction that $\mathbb{M}$ is not k-identifiable. This means that there exist two subsets $U, W \subseteq [n]$ of size at most k such that $U \neq W$ and $\mathbb{M}(U) = \mathbb{M}(W)$. Since $U \neq W$, without loss of generality we can say that there is a node $u \in U \setminus W$ (or in $W \setminus U$ if $U \subset W$). Hence, for node u, we have the two subsets $U, W \subseteq [n]$ such that $|U|, |W| \leq k$ and differing on u and we also have $\mathbb{M}(U) = \mathbb{M}(W)$ which means the node u is not k-identifiable and this is a contradiction. □

Let $\mathsf{ID}_k(\mathbb{M})$ be the set of k-identifiable nodes in $\mathbb{M}$.

Lemma 3.3. $\mathsf{ID}_k(\mathbb{M}) \subseteq \mathsf{ID}_{k'}(\mathbb{M})$ *for* $k' \leq k \leq n$.

Proof. From Definitions 2.1 and 3.1 it is immediate to see that k-identifiability implies k'-identifiability for *any* $k' < k$. Hence the claim. □

We can now prove the main theorem of this section.

Theorem 3.4. MID *is* NP*-hard.*

Proof. We consider the following optimization problem

NID:
Input: A Boolean $m \times n$ matrix $\mathbb{M}$, an element $u \in [n]$.
Output: The maximal $k \leq n$ such that $u \in \mathsf{ID}_k(\mathbb{M})$.

By Theorem 3.2 MID and NID are polynomially equivalent (MID $\equiv_p$ NID, that is polynomially reducible to each other). So to prove the NP-hardness of MID it is sufficient to prove the NP-hardness of NID.

As noticed before $\mathsf{ID}_k(\mathbb{M}) \subseteq \mathsf{ID}_{k'}(\mathbb{M})$ for *any* $k' \leq k$, hence to solve NID it is sufficient to know the *minimal* ℓ such $u \notin \mathsf{ID}_\ell(\mathbb{M})$: indeed, given such an ℓ, it is sufficient to set $k = \ell - 1$ to get a solution k for NID. We call this problem $\mathsf{NID}^\top$.

$\mathsf{NID}^\top$:
Input: A Boolean $m \times n$ matrix $\mathbb{M}$, an element $u \in [n]$.
Output: The minimal $\ell \leq n$ such that $u \notin \mathsf{ID}_\ell(\mathbb{M})$.

NID and $\mathsf{NID}^\top$ are clearly polynomially equivalent and to prove the NP-hardness of NID, we work with the problem $\mathsf{NID}^\top$.

Consider a hypergraph (a set-system) $\mathcal{H} = ([n], E)$, where $E \subseteq 2^{[n]}$ with $|E| = m$. A set $T \subseteq [n]$ is a *hitting set* for $\mathcal{H}$ if $T \cap e \neq \emptyset$ for all $e \in E$. T is *minimal* if no other subset T' of $[n]$ smaller than T has the same property.

The optimization problem *Minimum Hitting Set*, MHS, is

MHS:
Input: A hypergraph $\mathcal{H} = ([n], E)$.
Output: A minimal hitting set T of $\mathcal{H}$.

MHS is a well-known NP-hard problem [15, 16]. The next claim concludes the proof of the theorem.

Claim 3.5. $\mathsf{MHS} \leq_p \mathsf{NID}^\top$.

Proof. Let $\mathcal{H} = ([n], E)$ be a hypergraph. We define an instance $(\mathbb{M}_\mathcal{H}, u_\mathcal{H})$ for $\mathsf{NID}^\top$ as follows:

- $\mathbb{M}_\mathcal{H}$ has $n+1$ columns,
- the set of rows of $\mathbb{M}_\mathcal{H}$ is E,
- $u_\mathcal{H} = n+1$ and $\mathbb{M}(u_\mathcal{H}) = E$, namely every path touches the node $u_\mathcal{H}$.

We prove that $\mathcal{H} \in \mathsf{MHS}$ if and only if $(\mathbb{M}_\mathcal{H}, u_\mathcal{H}) \in \mathsf{NID}^\top$. We first prove the soundness of the reduction, that is $(\mathbb{M}_\mathcal{H}, u_\mathcal{H}) \in \mathsf{NID}^\top$. Let T be a minimal hitting set of size k for the instance $\mathcal{H}$ of MHS. We need to prove that

1. $u_\mathcal{H} \notin \mathsf{ID}_k(\mathbb{M}_\mathcal{H})$, and
2. $u_\mathcal{H} \in \mathsf{ID}_r(\mathbb{M}_\mathcal{H})$ for all $r \leq k-1$.

To show $u_\mathcal{H} \notin \mathsf{ID}_k(\mathbb{M}_\mathcal{H})$, by the definition of k-identifiability we have to find two subsets U and W of $[n+1]$ and of size at most k differing on $u_\mathcal{H}$ such that $\mathbb{M}(U) = \mathbb{M}(W)$. So we fix $U = T$ and $W = \{u_\mathcal{H}\}$. Since $U = T$ and T is an hitting set in $\mathcal{H} = ([n], E)$, then $U \subseteq [n]$ and since $u_\mathcal{H} = n+1$, U and W differ on $u_\mathcal{H}$, but by construction of $(\mathbb{M}_\mathcal{H}, u_\mathcal{H})$, $\mathbb{M}(U) = \mathbb{M}(W) = E$. Hence $u_\mathcal{H} \notin \mathsf{ID}_k(\mathbb{M})$. To prove the optimality condition, that is, $u_\mathcal{H} \in \mathsf{ID}_r(\mathbb{M}_\mathcal{H})$ for all $r \leq k-1$, assume by contradiction that there exist two subsets of $[n+1]$, U' and W' of size $r \leq k-1$ differing on $u_\mathcal{H}$ such that $\mathbb{M}(U') = \mathbb{M}(W')$. Since U' and W' differ on $u_\mathcal{H}$, $u_\mathcal{H}$ is in exactly one of them, say U'. Given that $\mathbb{M}(u_\mathcal{H}) = E$, it follows that $\mathbb{M}(W') = E$. Hence $W' \subseteq [n]$ is a hitting set for $\mathcal{H}$ of size strictly smaller than $k = |T|$, and this is a contradiction.

To prove the completeness of the reduction, given $(\mathbb{M}_\mathcal{H}, u_\mathcal{H}) \in \mathsf{NID}^\top$ we prove that $\mathcal{H} \in \mathsf{MHS}$. Since $(\mathbb{M}_\mathcal{H}, u_\mathcal{H}) \in \mathsf{NID}^\top$, we know that:

1. in $(\mathbb{M}_\mathcal{H}, u_\mathcal{H})$, there exist two subsets of $[n+1]$, U and W of size at most k differing on $u_\mathcal{H}$ such that $\mathbb{M}(U) = \mathbb{M}(W)$ (i.e., $u_\mathcal{H} \notin \mathsf{ID}_k(\mathbb{M}_\mathcal{H})$), and
2. for any pair of distinct subsets U' and W' of size $r \leq k-1$ differing on $u_\mathcal{H}$ we have $\mathbb{M}(U') \neq \mathbb{M}(W')$ (namely $u_\mathcal{H} \in \mathsf{ID}_r(\mathbb{M}_\mathcal{H})$ for all $r \leq k-1$).

Observe that since U and W differ on $u_\mathcal{H}$, $u_\mathcal{H}$ belongs to only one of them, say U. Hence $W \subseteq [n]$. Furthermore, since $E = \mathbb{M}(u_\mathcal{H})$ and $u_\mathcal{H} \in U$, thus $E = \mathbb{M}(U) = \mathbb{M}(W)$. Therefore, if we fix $T = W$, we have that $e \cap T \neq \emptyset$ for all $e \in E$ and T is a hitting set in $\mathcal{H}$. To prove the optimality of T, assume by contradiction that there exists a set $T' \subseteq [n]$, with $|T'| < |T| \leq k$ such that T' is also a hitting set in $\mathcal{H}$. Let $W' = T'$ and $U' = \{u_\mathcal{H}\}$. These are two subsets of $[n+1]$ of size at most $k-1$ such that $E = \mathbb{M}(U') = \mathbb{M}(W')$ and this contradicts Condition (2). Hence T is a minimal hitting set. ☐

3.1. Further observations on the complexity of MID

We conclude this section investigating the inverse reduction between MHS and MID. First we show that an algorithm solving MHS raises an algorithm to solve $\mathsf{NID}^\top$ and therefore MID.

Theorem 3.6. *Let $\mathcal{A}$ be an algorithm solving* MHS. *Then, there is an algorithm $\mathcal{B}$ solving* $\mathsf{NID}^\top$. *Furthermore if $\mathcal{A}$ works in polynomial time then $\mathcal{B}$ works in polynomial time too.*

Proof. The algorithm $\mathcal{B}$ solving $\mathsf{NID}^\top$ works in this way:

Algorithm 1 Algorithm $\mathcal{B}$

Input: a Boolean $m \times n$ matrix $\mathbb{M}$ and a node $u \in [n]$

1. Define $\mathcal{H}_\mathbb{M} = (V \setminus \{u\}, \mathbb{M}(u))$, where $\mathbb{M}(u) = \{p \in [m] \mid u \in p\}$
2. Run $\mathcal{A}$ on $\mathcal{H}_\mathbb{M}$ and let T be its output
3. **Output**: $k = |T|$

The algorithm $\mathcal{B}$ works clearly in polynomial time if $\mathcal{A}$ works in polynomial time. To prove its correctness we need to prove that k is the minimal ℓ such that $u \notin \mathsf{ID}_\ell(\mathbb{M})$. First we argue that $u \notin \mathsf{ID}_k(\mathbb{M})$. By Definition 3.1 we have to find two sets U, W differing on u and of size at most k such that $\mathbb{M}(U) = \mathbb{M}(W)$. Define $W = T$ and $U = \{u\}$. Since T is a hitting set in $\mathcal{H}_\mathbb{M}$, it contains all the edges in $\mathcal{H}_\mathbb{M}$, which are exactly $\mathbb{M}(u)$. Hence $\mathbb{M}(W) = \mathbb{M}(u) = \mathbb{M}(U)$.

To prove that k is the minimal value with that property, assume by contradiction that $u \notin \mathsf{ID}_\ell(\mathbb{M}_\mathcal{H})$ for some $\ell < |T| = k$. Since $u \notin \mathsf{ID}_\ell(\mathbb{M})$, there exist two subsets U and W of size at most ℓ differing on u such that $\mathbb{M}(U) = \mathbb{M}(W)$. Say without loss of generality that $u \in U$. Hence $\mathbb{M}(u) \subseteq \mathbb{M}(U) = \mathbb{M}(W)$. Therefore W is covering $\mathbb{M}(u)$ which is the set of edges in $\mathcal{H}_\mathbb{M}$. Hence W is a hitting set of $\mathcal{H}_\mathbb{M}$. But $|W| < |T|$ where T was the minimal hitting set. A contradiction.

□

Finally we consider the decision version DMID of the problem MID, that is - given a Boolean $m \times n$ matrix $\mathbb{M}$ and an integer k, $0 \leq k \leq n$, decide whether $\mathbb{M}$ is k-identifiable. We prove that DMID in coNP.

Theorem 3.7. DMID *is in* coNP. *Therefore the problem of deciding, given a set of m paths over n nodes $\mathbb{M}$ and an integer $k \leq n$, whether $\mathbb{M}$ is not k-identifiable is in* NP.

Proof. We have to prove that DMID $\in$ coNP. A certificate for this problem is any pair of sets $U, W \subseteq [n]$ with $|U|, |W| \leq k$ and such that $U \neq W$. This certificate is linear in the size of the input of DMID. According to Definition 2.1 to decide DMID, an algorithm has to verify whether $\mathbb{M}(U) \neq \mathbb{M}(W)$. Given U and W, this task can can be clearly accomplished in polynomial time in the size of $\mathbb{M}$ and k. Notice that DMID is $\forall$-problem: it follows that DMID $\in$ coNP. Of course the dual of this problem, that is decide if $\mathbb{M}$ is not k-identifiable is, by the same proof, in NP. □

4. Minimal number of paths for k-identifiability

In this section we consider the following question: given a network on n nodes and a non-negative integer value k for the identifiability, what is the minimal number m of measurement paths over the n nodes we have to consider in such a way that we will be able to identify uniquely and unambiguously at least k failing nodes (or in other words, in such a way that the maximal identifiability *value* is at least k)?

Let us consider first a toy example for $k = 1$.

Lemma 4.1. *Let $\mathbb{M}$ be a Boolean $m \times n$ matrix. If $m < \log(n+1)$ then $\mu(\mathbb{M}, m, n) < 1$.*

Proof. We prove the equivalent statement that if $n > 2^m - 1$, then $\mu(\mathbb{M}, m, n) < 1$. Since $n > 2^m - 1$ and the 0-column (i.e., the column with all entries equal to zero) cannot be in $\mathbb{M}$, in $\mathbb{M}$ there are at least two identical columns, that is two distinct nodes $u, w \in [n]$ which belong to the same set of paths. Therefore there are $U = \{u\}$ and $W = \{w\}$ of size 1 such that $\mathbb{M}(U) = \mathbb{M}(W)$. It follows that $\mu(\mathbb{M}, m, n) < 1$. □

In this subsection we generalize this result for a generic integer k, with $1 < k \leq n$. First notice that some partial result can be obtained from Theorem 2.6 but only for $k + 1$.

Theorem 4.2. *Let $\mathbb{M}$ be a Boolean matrix with m rows and n columns. Let C_k be the constant in Theorem 2.6. If $m < C_k(1 + o(1)) \log n$, then $\mu(\mathbb{M}, m, n) < (k + 1)$.*

Proof. Assume by contradiction that for $\mathbb{M}$ we have $m < C_k(1+o(1)) \log n$ and that $\mathbb{M}$ is $(k+1)$-identifiable. By Lemma 2.4 this means $\mathbb{M}$ is $(k + 1)$-separable, which implies that $\mathbb{M}$ is k-disjunct by Lemma 2.5. Now by Theorem 2.6 for a k-disjunct matrix we have $t(k, n) \geq C_k(1 + o(1)) \log n$. Since $t(k, n)$ is the minimum number of rows we need for $\mathbb{M}$ to be a k-disjunct, therefore $m \geq C_k(1 + o(1)) \log n$, which is a contradiction. □

Proving the result for k instead of $k + 1$, cannot be obtained from Theorem 2.6. This requires a different approach. We start by giving some preliminary definitions following [14]. A hypergraph $\mathcal{F}$ on the set $[m]$ is a family of distinct subsets of $[m]$, called *(hyper-)edges* of $\mathcal{F}$. If each edge is of fixed size $r \leq m$, then $\mathcal{F}$ is said to be r*-uniform*, i.e., $\mathcal{F} \subset \binom{[m]}{r}$.

Definition 4.3. *([5, 14]) For a positive integer k, $\mathcal{F}$ is called* k-union-free *if for any two distinct subsets of edges $\mathcal{A}, \mathcal{B} \subseteq \mathcal{F}$, with $1 \leq |\mathcal{A}|, |\mathcal{B}| \leq k$, it holds that $\cup_{A \in \mathcal{A}} A \neq \cup_{B \in \mathcal{B}} B$.*

Union-free uniform hypergraphs are investigated in extremal combinatorics [5].

A set $\mathbb{M}$ of m paths over n nodes defines a hypergraph $\mathcal{F}_{\mathbb{M}}$ with vertices in $[n]$ and hyper-edges included in $[m]$ in the following way: for $i \in [n]$ let A_i be the set of paths $j \in [m]$ the node i belongs to, that is $A_i = \{j \in [m] | \mathbb{M}[i, j] = 1\}$. Define $\mathcal{F}_{\mathbb{M}} = \{A_i, \mid i \in [n]\}$.

Given a set of nodes $U \subseteq [n]$ in $\mathbb{M}$, let $\mathcal{U}$ the subset of $\mathcal{F}_{\mathbb{M}}$, made by the A_i such that $i \in U$, that is $\mathcal{U} = \{A_i \in \mathcal{F}_{\mathbb{M}} | i \in U\}$. Notice that, by definition of $\mathbb{M}(U)$, $\mathbb{M}(U) = \bigcup_{i \in U} A_i$ which can be written as $\bigcup_{A \in \mathcal{U}} A$.

Theorem 4.4. *If $\mathbb{M}$ is a set of m paths over n nodes and $\mu(\mathbb{M}, m, n) \geq k$, then $\mathcal{F}_{\mathbb{M}}$ is k-union free.*

Proof. Assume that $\mu(\mathbb{M}, m, n) \geq k$. Let $\mathcal{U}$ and $\mathcal{W}$ be two distinct subsets of $\mathcal{F}_{\mathbb{M}}$ of size at most k. Let U and W such that $\mathcal{U} = \{A_i \in \mathcal{F}_{\mathbb{M}} | i \in U\}$ and $\mathcal{W} = \{A_i \in \mathcal{F}_{\mathbb{M}} | i \in W\}$. Since $\mathcal{U}$ and $\mathcal{W}$ are distinct, then also U and W are distinct and furthermore U and W are of cardinality at most k by the cardinality constraint on $\mathcal{U}$ and $\mathcal{W}$. That means that $\mathbb{M}(U) \neq \mathbb{M}(W)$ since $\mu(\mathbb{M}, m, n) \geq k$. But $\bigcup_{A \in \mathcal{U}} A = \mathbb{M}(U) \neq \mathbb{M}(W) = \bigcup_{A \in \mathcal{W}} A$. The claim is proved.

□

Notice that $\mathcal{F}_{\mathbb{M}}$ is not necessarily a uniform hypergraph. Let $r \in [m]$, then the subfamily of $\mathcal{F}_{\mathbb{M}}$ defined by $\mathcal{F}_{\mathbb{M}}(r) = \{A \in \mathcal{F}_{\mathbb{M}} | |A| = r\}$ is trivially a r-uniform hypergraph on $[m]$ for any $r \in [m]$. Notice that if $r_1, r_2 \in [m]$ with $r_1 \neq r_2$, then $\mathcal{F}_{\mathbb{M}}(r_1) \cap \mathcal{F}_{\mathbb{M}}(r_2) = \emptyset$. Therefore the subfamilies $\mathcal{F}_{\mathbb{M}}(1), \ldots \mathcal{F}_{\mathbb{M}}(m)$ form a partition of $\mathcal{F}_{\mathbb{M}}$ and therefore $|\mathcal{F}_{\mathbb{M}}| = \sum_{r \in [m]} |\mathcal{F}_{\mathbb{M}}(r)|$.
Since $|\mathcal{F}_{\mathbb{M}}| = n$, it follows that:

Lemma 4.5. $\sum_{r \in [m]} |\mathcal{F}_{\mathbb{M}}(r)| = n.$

Furthermore notice that if $\mathcal{F}_{\mathbb{M}}$ is k-union free, then $\mathcal{F}_{\mathbb{M}}(r)$ for each $r \in [m]$ will also be k-union free.

Let $m > r$ and $k \in [m]$ with $k \geq 2$, and let $f(k, r, m)$ denote the maximum cardinality of a k-union-free r-uniform hypergraph over $[m]$. The next theorem for $k \geq 3$ is Theorem 1.3 in [14] and for the case $k = 2$ in [5, 14].

Theorem 4.6 ([5, 14])**.** *For fixed integers $k, r \geq 3$ it holds that $f(k, r, m) \leq O(m^{\lceil \frac{r}{k-1} \rceil})$. Furthemore $f(2, r, m) = \Theta(m^{\frac{\lceil 4r/3 \rceil}{2}})$.*

Theorem 4.7. *There exist two constants $m_0 \in \mathbb{N}$ and C, such that for all $m \geq m_0$ if $\mathbb{M}$ is a set of m paths over n nodes, then*

1. *for $k \geq 3$, if $n > Cm^{1+\frac{m}{k-1}}$ then $\mu(\mathbb{M}) < k$, and*
2. *if $n > Cm^m$, then $\mu(\mathbb{M}) < 2$.*

Proof. Let $m_0 \in \mathbb{N}$ be the integer and let C be the constant obtained from the $O(\cdot)$-notations of the previous theorem such that for all $m \geq m_0$, we have both $f(k, r, m) \leq Cm^{\lceil \frac{r}{k-1} \rceil}$ and $f(2, r, m) \leq Cm^{\frac{\lceil 4r/3 \rceil}{2}}$.

Let us prove the case $k \geq 3$. Assume by contradiction that $n > Cm^{1+\frac{m}{k-1}}$ and $\mu(\mathbb{M}) \geq k$. Since $\sum_{r \in [m]} Cm^{\lceil \frac{r}{k-1} \rceil} \leq Cmm^{\frac{m}{k-1}}$, therefore $n > \sum_{r \in [m]} Cm^{\lceil \frac{r}{k-1} \rceil}$. By Theorem 4.4 $\mathcal{F}_{\mathbb{M}}$ is k-union free. Hence, by definition of k-identifibility and k-union freeness (see observation after Lemma 4.5) it follows that for each $r \in [m]$, $\mathcal{F}_{\mathbb{M}}(r)$ is a r-uniform k-union free hypergraph and hence by the previous theorem $|\mathcal{F}_{\mathbb{M}}(r)| \leq Cm^{\lceil \frac{r}{k-1} \rceil}$. The $\mathcal{F}_{\mathbb{M}}(r)$ partition $\mathcal{F}_{\mathbb{M}}$ and by Lemma 4.5 we have $n = \sum_{r \in [m]} |\mathcal{F}_{\mathbb{M}}(r)| \leq \sum_{r \in [m]} Cm^{\lceil \frac{r}{k-1} \rceil} < n$ and this is a contradiction.

The case $k = 2$ follows exactly the same reasoning observing that $Cm^{\frac{\lceil 4r/3 \rceil}{2}} \leq Cm^m$.

□

4.1. Upper bounds on the number of k-identifiable nodes

In many practical applications it might be useful to know an upper *bound* on the maximal number of failure nodes we can hope to identify unambiguously by Boolean methods based on k-identifiability. Previous results in this section can be used to obtain upper bounds on $\mathsf{ID}_k(\mathbb{M})$.

Corollary 4.8. *Let C and m_0 be as in Theorem 4.7. For all $m \geq m_0$, let $\mathbb{M}$ be a set of m paths over n nodes. Then*

1. $|\mathsf{ID}_k(\mathbb{M})| \leq \min\{n, Cm^{1+\frac{m}{k-1}}\}$, *for all $k \geq 3$;*
2. $|\mathsf{ID}_2(\mathbb{M})| \leq \min\{n, Cm^m\}$.

Proof. Let us prove the result for $k = 1$ using Lemma 4.1. The claims for $k > 1$ follow the same argument using Theorem 4.7.

$|\mathsf{ID}_1(\mathbb{M})| \leq n$ since it is a set of nodes. Assume that $n > 2^m - 1$, hence by Lemma 4.1 $\mu(\mathbb{M}) = 0$, hence there are at least two nodes $u_1 \neq u_2$ not 1-identifiable. Hence $|\mathsf{ID}_1(\mathbb{M})| \leq 2^m - 1$. □

Notice that our results can be expressed also in terms of the number of paths as follows: for example for $k = 2$ we observe that if $n > Cm^m$, then $m < \sqrt[1+\epsilon]{\log(n/C)}$ for any $\epsilon > 0$. Here we use the bound that $m \log m < m^{1+\epsilon}$ for any $\epsilon > 0$. A similar result can be easily obtained for the case $k \geq 3$.

Acknowledgements

We would like to thank Navid Talebanfard to point us the paper [14] and the anonymous reviewers of ICTCS 24 for their useful comments that helped us to improve the presentation and the readability of our paper.

References

[1] N. Galesi, F. Ranjbar, Tight bounds for maximal identifiability of failure nodes in boolean network tomography, in: 38th IEEE International Conference on Distributed Computing Systems, ICDCS 2018, Vienna, Austria, July 2-6, 2018, IEEE Computer Society, 2018, pp. 212–222. URL: https://doi.org/10.1109/ICDCS.2018.00030. doi:10.1109/ICDCS.2018.00030.

[2] N. Galesi, F. Ranjbar, Tight bounds to localize failure nodes on trees, grids and through embeddings under boolean network tomography, Theor. Comput. Sci. 919 (2022) 103–117. URL: https://doi.org/10.1016/j.tcs.2022.03.035. doi:10.1016/j.tcs.2022.03.035.

[3] N. Galesi, F. Ranjbar, M. Zito, Vertex-connectivity for node failure identification in boolean network tomography, in: F. Dressler, C. Scheideler (Eds.), Algorithms for Sensor Systems - 15th International Symposium on Algorithms and Experiments for Wireless Sensor Networks, ALGOSENSORS 2019, Munich, Germany, September 12-13, 2019, Revised Selected Papers, volume 11931 of *Lecture Notes in Computer Science*, Springer, 2019, pp. 79–95. URL: https://doi.org/10.1007/978-3-030-34405-4_5. doi:10.1007/978-3-030-34405-4_5.

[4] N. Galesi, F. Ranjbar, M. Zito, Vertex-connectivity for node failure identification in boolean network tomography, Information Processing Letters 184 (2024) 106450. URL: https://www.sciencedirect.com/science/article/pii/S0020019023000935. doi:https://doi.org/10.1016/j.ipl.2023.106450.
[5] P. Frankl, Z. Füredi, Union-free families of sets and equations over field, Journal of Number Theory 23 (1986) 210–218.
[6] T. He, L. Ma, A. Swami, D. Towsley, Network Tomography: Identifiability, Measurement Design, and Network State Inference,, Cambridge University Press, 2021.
[7] N. G. Duffield, Network tomography of binary network performance characteristics, IEEE Trans. Information Theory 52 (2006) 5373–5388. URL: https://doi.org/10.1109/TIT.2006.885460. doi:10.1109/TIT.2006.885460.
[8] L. Ma, T. He, A. Swami, D. Towsley, K. K. Leung, J. Lowe, Node failure localization via network tomography, in: C. Williamson, A. Akella, N. Taft (Eds.), Proceedings of the 2014 Internet Measurement Conference, IMC 2014, Vancouver, BC, Canada, November 5-7, 2014, ACM, 2014, pp. 195–208. URL: http://doi.acm.org/10.1145/2663716.2663723. doi:10.1145/2663716.2663723.
[9] L. Ma, T. He, A. Swami, D. Towsley, K. K. Leung, On optimal monitor placement for localizing node failures via network tomography, Perform. Eval. 91 (2015) 16–37. URL: https://doi.org/10.1016/j.peva.2015.06.003. doi:10.1016/j.peva.2015.06.003.
[10] L. Ma, T. He, A. Swami, D. Towsley, K. K. Leung, Network capability in localizing node failures via end-to-end path measurements, IEEE/ACM Trans. Netw. 25 (2017) 434–450. URL: https://doi.org/10.1109/TNET.2016.2584544. doi:10.1109/TNET.2016.2584544.
[11] D. S. Hochbaum (Ed.), Approximation algorithms for NP-hard problems, PWS Publishing Co., USA, 1996.
[12] R. Bar-Yehuda, S. Even, A linear-time approximation algorithm for the weighted vertex cover problem, Journal of Algorithms 2 (1981) 198–203. URL: https://www.sciencedirect.com/science/article/pii/0196677481900201. doi:https://doi.org/10.1016/0196-6774(81)90020-1.
[13] D.-Z. Du, F. Hwang, Combinatorial Group Testing and Its Applications, World Scientific, 1993.
[14] C. Shangguan, I. Tamo, New turán exponents for two extremal hypergraph problems, SIAM Journal on Discrete Mathematics (2020).
[15] G. Ausiello, A. D'Atri, M. Protasi, Structure preserving reductions among convex optimization problems, J. Comput. Syst. Sci. 21 (1980) 136–153. URL: https://doi.org/10.1016/0022-0000(80)90046-X. doi:10.1016/0022-0000(80)90046-X.
[16] M. R. Garey, D. S. Johnson, Computers and Intractability, A Guide to the Theory of NP-Completeness, 22 ed., W. H. Freeman and Company, New York, 2000.