Network Identifiability: Advances in Separating Systems and Networking Applications

Viviana Arrigoni and Novella Bartolini, *Senior Member, IEEE*

***Abstract*—This letter discusses the parallelism between the concept of network identifiability introduced by Boolean Network Tomography (BNT), and the theory of separating systems, highlighting applications of interest to research in networking. This letter evidences how recent results of BNT have direct implications to the formulation of new bounds to the size of separating systems over finite sets. Grounding on these theoretical results, we provide an efficient algorithm for the design of separating systems that meet the bound tightly. We extend the proposed results, bounds and algorithm, to networking applications, including network failure assessment, robust network design and compressed sensing.**

***Index Terms*—Computer network reliability, information theory, compressed sensing, combinatorial testing.**

I. Introduction

BOOLEAN Network Tomography (BNT) is a powerful technique for inferring the binary state (failed/working) of internal elements of a network (nodes/links) by means of end-to-end measurement paths. Network Tomography (NT) has gained wide attention in the networking research community in the past years. Its applications span from failure localization, [1], [2], [3], delay and bandwidth inference [4], [5], topology discovery [6], as it provides techniques that overcome monitoring issues due to network heterogeneity and multi-proprietary nature. These techniques are very versatile and prove to be useful in many application contexts, including urban traffic monitoring [7]. Although born before the advent of the latest centralized network frameworks (e.g., software-defined networks and network function virtualization), NT techniques can provide efficient monitoring also in these contexts with negligible traffic overhead [8], [9]. However, BNT also brings about many challenges, mostly related to the computational intractability of data interpretation. *Identifiability* is a core concept in BNT [3]. Roughly, we say that a network node *v* is identifiable with respect to a set of paths traversing the network, *P*, if, when it fails, it is possible to identify its state without ambiguity by observing the binary outcome (failed or working) of the paths in *P*. Our works in [10], [11] provide fundamental bounds for network node identifiability. We notice a strong connection between the concepts of BNT identifiability, separating systems, combinatorial group testing and compressed sensing. This relation has already been highlighted in previous works [4], [5], [12]. In this letter, for the first time, we define the structural analogy between separating systems and network identifiability. In Section II, we summarize our results proven in [11]. In particular, after providing the formal definition of node identifiability, we report our bound of the minimum number of paths of constrained path length necessary to identify an arbitrary number of nodes, and an algorithm (Incremental Crossing Arrangement - ICA) for network design that meets such bound tightly. In Section III, we provide background on separating systems and existing bounds on their size. We show the equivalence between sets of paths that guarantee identifiability and separating systems (Proposition 1), and we show that our bounds improve previous results. We provide an example of how ICA can be used to design group testing experiments with minimum number of probes (Section III-A), and discuss the applicability of NT in Compressed Sensing (Section III-A). We finally show experimental results that highlight the tightness of our bounds (Section IV).

Manuscript received 30 April 2022; revised 29 June 2022; accepted 12 August 2022. Date of publication 18 August 2022; date of current version 21 November 2022. This work was supported by the NGIAtlantic.eu under Grant 871582. The associate editor coordinating the review of this article and approving it for publication was C. Wang. *(Corresponding author: Viviana Arrigoni.)*

The authors are with the Department of Computer Science, Sapienza University of Rome, 00185 Rome, Italy (e-mail: arrigoni@di.uniroma1.it; novella@di.uniroma1.it).

Digital Object Identifier 10.1109/LNET.2022.3200002

II. Background in Boolean Network Tomography and Existing Results

In BNT, some peripheral nodes, hereafter called *monitors*, send probing packets along given paths. Each path probe has a binary outcome: it can either be *working* or *failed*. A path works if it only traverses working nodes and fails if it traverses at least one failed node. In the example of Fig. 1a, working nodes are highlighted in green (v_1, v_2, v_4), and failed nodes in red (v_3). Path p_1 fails whereas path p_2 works. The idea of BNT is to infer the binary state of the network nodes by testing end-to-end paths from a set *P*. This problem can be formalized as follows: we can define $R \in \{0,1\}^{m \times n}$ to be the *routing matrix*, where $m = |P|$ is the number of paths in *P* and $n = |V|$ is the number of nodes in the graph $G = (V, E)$ that describes the topology of the network. $R_{ij} = 1$ if path p_i traverses node v_j, $R_{ij} = 0$ otherwise. The columns of *R* are the so called *binary encodings* of the nodes. The binary encoding of node v_j is denoted as $b(v_j)$ and represents the characteristic vectors of the subset of paths that traverse v_j. Notice that $\chi(v_j) = ||b(v_j)||_1$, i.e., the 1-norm of $b(v_j)$, is the number of paths traversing node v_j, hereafter called the *crossing number* of v_j. Paths are modeled as the set of nodes they traverse, and the length of a path *p*, $|p|$, is the size of such set. For each path p_i we define a variable y_i such that $y_i = 0$ if p_i works, $y_i = 1$ otherwise. The value of y_i is observable by probing path p_i. For each node v_j we define the variable x_j

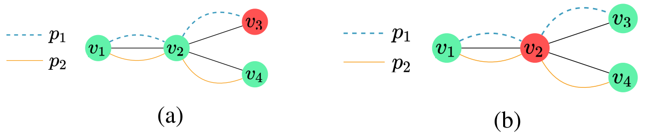


Fig. 1. Two paths p_1 and p_2 traversing a small network. Working nodes are highlighted in green (v_1, v_2, v_4 left - v_1, v_3, v_4 right) and failed nodes in red (v_3 left - v_2 right).

likewise. To classify the state of the network nodes, we need to solve the following system of Boolean equations:

$$y_i = \bigvee_{j=1}^{n} R_{ij} x_j \qquad \forall i = 1, \ldots, m. \tag{1}$$

For the example of Fig. 1a, the system is the following:

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \tag{2}$$

Because path p_1 fails and p_2 works, $y_1 = 1$, $y_2 = 0$. Since nodes v_1, v_2, v_4 work, it holds that $x_j = 0$ for $j = 1, 2, 4$, whereas $x_3 = 1$.

The authors of [3] tackle the case in which the number of failures is known to be lower than or equal to *k*, formalizing the notion of *k*-identifiability. A node *v* is *k*-identifiable with respect to a set of paths *P* if, when the number of failed nodes in the network is $\leq k$, it is possible to assess the state of *v* without ambiguity, by observing the binary outcomes of the paths in *P*.

In this letter we focus on the case of 1-identifiability, that is formally defined as follows:

Definition 1: A node v_j is 1-identifiable with respect to $P = \{p_1, \ldots, p_m\}$ if $b(v_j) \neq 0^m$ and if for all $v_k \neq v_j$, $b(v_k) \neq b(v_j)$, i.e., its binary encoding is not null and not identical to that of any other node.

For the rest of this letter, we shall refer to 1-identifiability as *identifiability*. Notice that in the simple example of Fig. 1a, nodes v_1 and v_2 are not identifiable, contrarily to nodes v_3 and v_4. As a matter of fact, $x = [0, 0, 1, 0]^T$ is the only possible solution to the system in Equation (2).

Fig. 1b represents the same topology as Fig. 1a, with v_2 as the only failed node. The associated Boolean system has two possible solutions ($x = [1, 0, 0, 0]^T$ and $x = [0, 1, 0, 0]^T$) because v_2 is non-identifiable.

In our works [10], [11] we provide topology agnostic theoretical bounds on network identifiability. We summarize them in the remainder of this section. We analyze different routing scenarios, one of which is called *unconstrained routing*, i.e., paths do not contain loops nor circles. In particular, in [11] we show that under unconstrained routing, topology and path length, the minimal number of paths $m^{id}_{\mathtt{min}}$ necessary to identify *n* nodes is

$$m^{id}_{\mathtt{min}}(n) = \lceil \log_2(n+1) \rceil. \tag{3}$$

We then focus on bounding the minimal number of paths $m^{id}_{\mathtt{min}}(n, d_{\mathtt{max}})$ necessary to identify *n* nodes under the constraint $|p_i| \leq d_{\mathtt{max}}$ $\forall p_i \in P$, and prove the following bound

---

Algorithm 1: Incremental Crossing Arrangement (ICA)

Input: n and $d_{\mathtt{max}}$.

Output: A set of encodings B_V which can be mapped onto a topology graph $G = (V, E)$ where all $|V| = n$ nodes are identifiable by using $m^{id}_{\mathtt{min}}$ paths with average length $d_{\mathtt{max}}$, i.e., a routing matrix R whose columns are the binary encodings B_V.

Calculate $m = m^{id}_{\mathtt{min}}(n, d_{\mathtt{max}})$ according to Equation (4)
For $i = 1, \ldots, m$ **do** set $d_i = \min\{2^{m^{id}_{\mathtt{min}}-1}, d_{\mathtt{max}}\}$
$B_V = \emptyset$
For $i = 1, \ldots, i_{\mathtt{max}}$ **do** $B_V = B_V \cup \mathcal{B}(i)$
Calculate the family $\mathcal{F}$ defined as $\mathcal{F} \triangleq \{B\colon B \subseteq \mathcal{B}(i_{\mathtt{max}}+1) \wedge \ell_k(B \cup B_V) \in [d_k - 1, d_k],\ \forall k = 1, \ldots, m\}$
Choose $B^* = \arg\max_{B \in \mathcal{F}} |B|$
Return $B_V = B_V \cup B^*$

(Theorem 2 in [11]):

$$m^{id}_{\mathtt{min}}(n, d_{\mathtt{max}}) = \min m\ s.t. \left\lfloor \frac{l_{\mathtt{max}} \cdot m}{i_{\mathtt{max}} + 1} \right\rfloor + \sum_{i=1}^{i_{\mathtt{max}}} \binom{m}{i} \geq n, \tag{4a}$$

$$\text{where } i_{\mathtt{max}} = \max\left\{ j : \sum_{i=1}^{j} i \cdot \binom{m}{i} \leq m \cdot D \right\}, \tag{4b}$$

$$D = \min\left\{ d_{\mathtt{max}}, 2^{m-1} \right\}, \tag{4c}$$

$$\text{and } l_{\mathtt{max}} = D - \sum_{i=0}^{i_{\mathtt{max}}-1} \binom{m-1}{i}. \tag{4d}$$

We highlight that an increasing value of $d_{\mathtt{max}}$ guarantees higher identifiability as long as $d_{\mathtt{max}} \leq 2^{m-1}$; this is captured by Equation (4c). Finally, we report a polynomial time algorithm called Incremental Crossing Arrangement (ICA) [11] for designing a topology and a set of paths traversing it, such that the bound in Equation (4) is met tightly. We report the pseudocode of ICA in Algorithm 1, where $\mathcal{B}(i) \subset \{0,1\}^m$ is the set of binary strings of length *m* with exactly *i* "1"s, and $\ell_i(B)$ is the number of encodings of $B \subseteq \{0,1\}^m$ having a 1 in the *i*-th position.

Intuitively, as already brought up in [10] the strategy to follow in order to minimize the number of paths *m* of maximum path length $d_{\mathtt{max}} \leq 2^{m-1}$ such that *n* nodes are 1-identifiable (i.e., their binary encoding is unique) is equivalent to the following: given a number of paths *m* with maximum length constraint, find the maximum number of identifiable nodes. We reason in terms of binary strings of length *m* and argue that, under path length constraints, the number of distinct encodings is maximized when we minimize the number of "1"s in the encodings. This is achieved when we have $\binom{m}{1}$ binary encodings with one 1 (each path crosses a different node and has remaining length equal to $d_{\mathtt{max}} - 1$), $\binom{m}{2}$ encodings with two "1"s (each path crosses $\binom{m-1}{0} = 1$ node of crossing number 1, and $\binom{m-1}{1} = m - 1$ nodes of crossing number 2, and has remaining path length equal to $d_{\mathtt{max}} - m - 2$), and so forth, until the path length $d_{\mathtt{max}}$ allows to continue, that is until all binary strings with up to $i_{\mathtt{max}}$ "1"s are formed. If $i_{\mathtt{max}}$ is such that $l_{\mathtt{max}} > 0$, it means that the maximum path length has not been reached yet, and therefore some of the binary strings with $\binom{m}{i_{\mathtt{max}}+1}$ "1"s can be selected in our set of encodings B_V.

Thanks to ICA, we are able to design a topology and to define a set of paths traversing it such that all the nodes are identifiable, i.e., their state can be assessed without ambiguity. This capability provides guarantees for robust network monitorability, as discussed in [2], [11], [13]. In the next section, we show how these results can be mapped onto the theory of separating systems, combinatorial group testing and compressed sensing.

III. Separating Systems and Applications

Introduced in [14], separating systems found a wide set of applications in Causal Inference Theory, whose employment is extremely useful to respond to the growing data-driven decision-making approach that populate machine learning applications. They are defined as follows:

Definition 2: Let *H* be a finite set. The system $\mathcal{A} \subseteq 2^H$ is a *separating system* if for any $x, y \in H$, $x \neq y$:$\exists S \in \mathcal{A}$, such that $x \in S, y \notin S$ or $x \notin S, y \in S$.
The work in [15] shows how the theory of separating systems can be used to understand causal relationships from purely empirical data. In particular, it allows to select sequences of experiments that can identify the underlying causal structure relating observable data, or to craft an experiment that maximizes the insight one can expect to gain from a set of observations. These questions relapse also in the task of inferring the state of the nodes of a network via NT (i.e., through end-to-end measurement paths) as also observed in [12], making it a perfectly compatible field of application of the theory of separating systems. It is trivial to prove the following:

Proposition 1: Let *G* = (*V*, *E*) be the graph describing the topology of a network, being $V = \{v_1, \dots, v_n\}$ the set of its nodes. Let $P = \{p_1, \dots, p_m\}$ be a set of paths traversing *G* such that $\forall v \in V\ \exists p_i \in P{:}b(v)|_i = 1$. Each path p_i can be modeled as the set of nodes it traverses, i.e., $P \subseteq 2^V$. Then the set of paths *P* is a separating system if and only if all nodes *V* are 1-identifiable.
Observe that the one difference between the two concepts lays on the fact that the empty set may belong to a separating system, whereas an empty path is meaningless in the context of BNT.

The result of Proposition 1 defines the parallelism between sets of paths guaranteeing network identifiability and separating systems. In light of this observation, in the remainder of this section, we show how the results shown in Section II can be exploited to solve problems related to separating systems. Separating systems play a central role in combinatorial group testing (CGT) [12]. The essential goal of CGT is to localize some *defective* items in a pool of items by testing them in groups instead of singularly. A good CGT strategy will achieve this goal by driving less tests than the total number of testing groups. Intuitively, the parallelism between separating systems, CGT and BNT can be discussed as follows: in CGT, separating systems are sets of experimental elements ($\mathcal{A}$, in Definition 2) that guarantee to identify one defective item in a collection of items, *H*. Equivalently, in BNT, sets of elements to tests are paths *P* to probe with end-to-end packets sends. Each path can be modeled as the set of nodes it traverses and can be interpreted as the group of elements (nodes) to test. The binary encoding of a node says in which groups the node is selected for testing. As we said, one trivial and undesired solution is $\mathcal{A} = \{\{u\}, \forall u \in H\}$, that corresponds to testing each element of *H* individually. Contrarily, one would want to drive less tests to localize the defective item. Also in BNT the choice of $P = \{\{v_i\}, \forall v_i \in V\}$ is extremely inconvenient: degenerate paths (i.e., paths traversing only one node) are particularly unlikely and expensive to test in real networking applications.

From Proposition 1, it follows that bounding the size of a set of paths *P* such that a set of nodes *V* is 1-identifiable with respect to *P* is equivalent to bounding the size of a separating system. It is no surprise then that the minimal size for a separating system $\mathcal{A}$ is $\lceil \log_2(n) \rceil$, where $n = |H|$ [16]; this result is essentially what we derive in Equation (3), with the only difference being the "+1", that is due to the fact that separating systems can include the empty set. The work in [16] also provides a lower bound to the size of a separating system $\mathcal{A} \subseteq 2^H$ with the constraint that $\forall S \in \mathcal{A}$, $|S| \leq d_{\mathtt{max}}$. We call this lower bound $m^{ss}_{\mathtt{min}}(n, d_{\mathtt{max}})$. The same work also shows that the maximum size $d_{\mathtt{max}}$ has to be $\leq n/2$. This is in accordance with what we observed in Section II; in fact, $n/2 = 2^{m^{id}_{\mathtt{min}}-1}$ (by Equation (3)), and in Equation (4d) we actually define the maximum path length as $D = \min\{d_{\mathtt{max}}, 2^{m-1}\}$. The bound has been recently improved in [12] as follows:

$$m^{ss}_{\mathtt{min}}(n, d_{\mathtt{max}}) = \min m{:} \sum_{i=0}^{j} i \binom{m}{i} + a(j+1) \leq m d_{\mathtt{max}}, \quad \text{(5a)}$$

$$\text{where} \sum_{i=0}^{j} \binom{m}{i} + a = n, \quad \text{(5b)}$$

$$\text{and } j \leq m-1,\ a < \binom{m}{j+1},\ d_{\mathtt{max}} < \frac{n}{2}. \quad \text{(5c)}$$

This result is analogous to our bound in Equation (4), shifted by one unit, as a consequence of the fact that we exclude nodes whose binary encoding is 0^m (corresponding to a node not traversed by any path, or to an empty set in the terminology of separating systems). We can observe that the term *j* in Equation (5) embodies $i_{\mathtt{max}}$ in Equation (4b), with the only difference being that $i_{\mathtt{max}} \leq m$ (instead of $\leq m-1$) as a consequence of the one unit shift. In fact, as we discuss in [11], $i_{\mathtt{max}}$ represents the maximum integer such that all binary encodings with up to $i_{\mathtt{max}}$ "1"s can be selected, and therefore $i_{\mathtt{max}} \leq m$. The term $\lfloor (l_{\mathtt{max}} m)/(i_{\mathtt{max}}+1) \rfloor$ in Equation (4a) represents the number of binary encodings with $(i_{\mathtt{max}}+1)$ "1"s that can be selected. By construction, this number is always $< \binom{m}{i_{\mathtt{max}}+1}$, and it plays the same role as the term *a* in Equation (5). Having seen that a set of paths *P* defining sets of identifiable nodes is a separating system, we can state the following:

Proposition 2:
1) The bound in Equation (4) is a bound to the minimal size of $\mathcal{A} \subseteq 2^H$ such that $\mathcal{A}$ is a separating system and $\forall S \neq \emptyset \in \mathcal{A}$, $|S| \leq d_{\mathtt{max}}$.
2) The bound is tight (see Proposition 2 in [11]).
3) ICA in Algorithm 1 can be used to design a separating system of minimal size.

A. Incremental Crossing Arrangement for Combinatorial Group Testing: An example

We now show a simple example, inspired by the one in [17], of how ICA (Algorithm 1) can be used to design a CGT experiment. Suppose we have 7 gold coins, one of which we know is counterfeit and so of a different mass than the other coins. Assume that we know that a regular coin weights w grams, and that we have an accurate electronic scale. The scale only fits up to three coins at a time. By solving Equation (4) with $n = 7$ and $d_{\mathtt{max}} = 3$, we get that $m^{id}_{\mathtt{min}}(7, 3) = 4$ and $i_{\mathtt{max}} = 1$. We then run ICA. We set $d_i = 3$ for each experiment (i.e., at most three coins can be measured at a time), line 2, Algorithm 1. We derive $B_V = \{\{1, 0, 0, 0\}, \{0, 1, 0, 0\}, \{0, 0, 1, 0\}, \{0, 0, 0, 1\}\}$, line 4. Among all possible $\mathcal{B}(i_{\mathtt{max}} + 1) \in \mathcal{F}$, we choose $\mathbf{B} = \{\{1, 1, 0, 0\}, \{1, 0, 1, 0\}, \{0, 0, 1, 1\}\}$, lines 5-6, and set $B_V = B_V \cup \mathbf{B}$. Let x_j be the random variable such that $x_j = 0$ if the j-th coin is normal, and $x_j = 1$ if it is counterfeit. Similarly, for each measurement i on d_i coins, we define $y_i = 0$ if the weight is $d_i \cdot w$ grams (i.e., the measurement contains only normal coins), $y_i = 1$ otherwise (i.e., the measurement contains the counterfeit coin). We can then set each string in B_V as a column of the routing matrix R, as follows:

$$R = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}.$$

Each row of R says which coins to include in one measurement. Thanks to the choice of R, each counterfeit coin c_j will trigger different sets of measurements. In fact, if $\mathcal{Y} = \{i = 1, \ldots, m^{id}_{\mathtt{min}}(7, 3) = 4 : y_i = 1\}$, it follows that:

$$x_1 = 1 \Leftrightarrow \mathcal{Y} = \{1\}, \quad x_2 = 1 \Leftrightarrow \mathcal{Y} = \{2\}, \quad x_3 = 1 \Leftrightarrow \mathcal{Y} = \{3\},$$
$$x_4 = 1 \Leftrightarrow \mathcal{Y} = \{4\}, \quad x_5 = 1 \Leftrightarrow \mathcal{Y} = \{1, 2\}, x_6 = 1 \Leftrightarrow \mathcal{Y} = \{1, 3\},$$
$$x_7 = 1 \Leftrightarrow \mathcal{Y} = \{3, 4\}.$$

Hence, any counterfeit coin will produce a different and identifiable outcome in the measurements. By Proposition 2, the bound $m^{id}_{\mathtt{min}}(7, 3) = 4$ is tight, whereas it results that $m^{ss}_{\mathtt{min}}(7, 3) = 2$ (Equation (5)), which is a looser bound.

B. Network Tomography and Compressed Sensing

Compressed Sensing (CS) is an effective technique for acquiring and reconstructing sparse signals from a series of sampling measurements. Mathematically, this problem is equivalent to recover the exact solution to an under-determined linear system $Rx = y$ that, if consistent, has infinite possible solutions. CS aims at recovering the signal x with few observations, y. In order to be able to recover x out of the infinitely many solutions, two requirements must be verified. The first one is that x is sparse in some domain (i.e., x must be k-sparse, that is $||x||_0 \leq k$ for a small k). The second requirement is incoherence, which is guaranteed by the Restricted Isometry Property (RIP). In recent years, many researchers investigated binary compressed sensing, which considers $R \in \{0, 1\}^{m \times n}$, see, e.g., [18], [19]. These studies have several interesting applications, including wireless communication and error correcting codes, where signals are strings of bites. In this section we focus on this special case. The following result, proven in [17], can be exploited to establish the connection between network identifiability and uniqueness of the solution to the 1-sparse signal problem in CS.

Proposition 3: Every subset of $2k$ columns of R is linearly independent $\iff$ all non-zero vectors in the null space of R have at least $2k + 1$ non-zero entries $\implies$ any k-sparse solution to $Rx = y$ is unique.

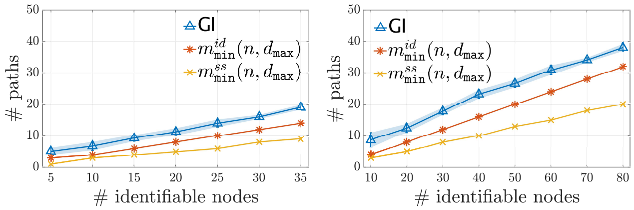
By choosing $k = 1$ and assuming that R is a binary matrix, the linear independence between every possible couple of columns of R is obtained by letting all columns be different from one another, that corresponds to the definition of 1-identifiability (Definition 1). The direct implication of this result is the following proposition:

Proposition 4: The bound of Equation (4) establishes the minimum number of observations (i.e., tests) m necessary to retrieve the signal that solves the 1-sparse CS problem, constrained to at most $d_{\mathtt{max}}$ samples for each test. Furthermore, ICA can be used to design the sampling matrix R.

The work in [5] was the first to use CS for NT inference. More recently, Chen *et al.* [4] show how NT and CS can be used together for localizing congested links in a communication network. In this scenario, the goal is to retrieve the solution x to the Boolean system in Equation (1), where $x_j = 1$ if link l_j is congested, $x_j = 0$ otherwise. The work in [4] proves that a routing matrix with no null columns satisfies the Restricted Isometry Property (RIP). By ensuring matrix incoherence, this property allows to find an unique k-sparse solution to a linear system with high probability. The authors also prove that network identifiability in Definition 1 guarantees statistical identifiability of the probability vector obtained by testing sufficiently many times the set of paths P for dynamic state inference. These facts imply that ICA can be employed to design a topology and a routing strategy that guarantee the conditions for the inference of additive and Boolean metrics over nodes and links via CS.

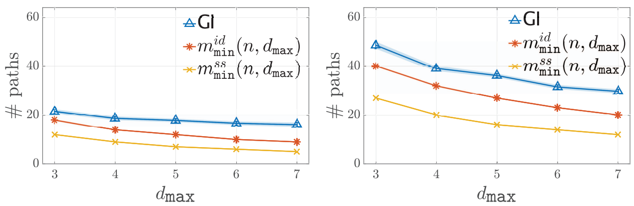
IV. Experimental Results

In this section, we provide experimental results that compare our bounds in Equation (4) to the bounds in Equation (5), which are computed as in [12, Algorithm 1]. Similarly to the performance evaluation that we provide in [11], we evaluate the tightness of the bounds in comparison with the Greedy for Identifiability algorithm, GI [3]. GI works as follows: given a network, represented as a graph, a number of nodes to identify, n, and the maximum path length, $d_{\mathtt{max}}$, it iteratively chooses the next path to place as the one that maximizes the number of identifiable nodes in the network and whose length is $\leq d_{\mathtt{max}}$. The work in [3] proves that the number of paths placed by this heuristic is a constant approximation of the optimal solution. We highlight that, as a consequence of Proposition 2, GI behaves as a greedy algorithm for minimizing the number of elements in a separating system with the constraint that each element (i.e., set of the partition) has a fixed maximal size. We drive our tests on the following topologies: **Valley**



(a) Valley Net (b) Jellyfish

Fig. 2. Number of paths to identify variable numbers of nodes on different topologies with $d_{\mathtt{max}} = 4$.



(a) Valley Net, $n = 35$ (b) Jellyfish, $n = 80$

Fig. 3. Number of paths to identify n nodes with variable values of $d_{\mathtt{max}}$ on different topologies.

Net is a fibre optical network located in the USA, provided by the Topology Zoo dataset [20]. It has 39 nodes and 53 edges.

Jellyfish is a topology structure that was recently proposed for data centres [21] as it offers high throughput and capacity, high scalability and failure resiliency. The internal nodes of the Jellyfish (nodes with degree strictly greater than one) are network switches, whereas leaf nodes are servers. We generate a Jellyfish topology with 100 nodes and 150 edges.

The plots show average results obtained through 20 runs. Shades represent standard deviations. In Fig. 2 we plot the number of required paths to identify a variable number of nodes for GI, n, and the corresponding bounds $m^{id}_{\mathtt{min}}(n, d_{\mathtt{max}})$ and $m^{ss}_{\mathtt{min}}(n, d_{\mathtt{max}})$. The maximum path length $d_{\mathtt{max}}$ is fixed to 4. Fig. 2a shows these results for the Valley Net network and Fig. 2b for the Jellyfish topology. Fig. 3 shows the curves of the number of required paths to identify a fixed number of nodes (35 nodes for Valley Net, Fig. 3a and 80 nodes for the Jellyfish, Fig. 3b), and varying the maximum path length $d_{\mathtt{max}}$. Both experiments show that our bounds $m^{id}_{\mathtt{min}}(n, d_{\mathtt{max}})$ are tighter than $m^{ss}_{\mathtt{min}}(n, d_{\mathtt{max}})$ in all the configurations, and close to the number of paths used by the GI algorithm.

V. Conclusion

This letter provides constructive comments on how to extend results of Boolean Network Tomography (BNT) to other domains of interest to researchers in communication networks. It shows practical applications of BNT results to the theory of separating systems and combinatorial group testing, providing evidence of their power in robust network design, network failure assessment and compressed sensing. In particular, we consider bounds on node identifiability provided by the research on BNT. We show their extension to the problem of separating system, showing how they improve currently known bounds for such a field. We evaluate our bounds in comparison with the most recently published bounds for separating systems. We show that our bounds are tighter and close to the results obtained by the sub-optimal Greedy for Identifiability algorithm.

References

[1] V. Arrigoni, N. Bartolini, A. Massini, and F. Trombetti, “Failure localization through progressive network tomography,” in *Proc. IEEE INFOCOM*, 2021, pp. 1–10.
[2] S. Tati, S. Silvestri, T. He, and T. La Porta, “Robust network tomography in the presence of failures,” in *Proc. IEEE ICDCS*, 2014, pp. 481–492.
[3] T. He, N. Bartolini, H. Khamfroush, I. Kim, L. Ma, and T. La Porta, “Service placement for detecting and localizing failures using end-to-end observations,” in *Proc. IEEE ICDCS*, 2016, pp. 560–569.
[4] J. Chen, X. Qi, and Y. Wang, “An efficient solution to locate sparsely congested links by network tomography,” in *Proc. IEEE ICC*, 2014, pp. 1278–1283.
[5] M. H. Firooz and S. Roy, “Network tomography via compressed sensing,” in *Proc. GLOBECOM*, 2010, pp. 1–5.
[6] B. Eriksson, G. Dasarathy, P. Barford, and R. Nowak, “Efficient network tomography for Internet topology discovery,” *IEEE/ACM Trans. Netw.*, vol. 20, no. 3, pp. 931–943, Jun. 2012.
[7] R. Zhang, S. Newman, M. Ortolani, and S. Silvestri, “A network tomography approach for traffic monitoring in smart cities,” *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 7, pp. 2268–2278, Jul. 2018.
[8] M. M. Tajiki, S. H. G. Petroudi, S. Salsano, S. Uhlig, and I. Castro, “Optimal estimation of link delays based on end-to-end active measurements,” *IEEE Trans. Netw. Service Manag.*, vol. 18, no. 4, pp. 4730–4743, Dec. 2021.
[9] M. Rahali, J.-M. Sanner, and G. Rubino, “TOM: A self-trained tomography solution for overlay networks monitoring,” in *Proc. IEEE CCNC*, 2020, pp. 1–6.
[10] N. Bartolini, T. He, V. Arrigoni, A. Massini, F. Trombetti, and H. Khamfroush, “On fundamental bounds on failure identifiability by boolean network tomography,” *IEEE/ACM Trans. Netw.*, vol. 28, no. 2, pp. 588–601, Apr. 2020.
[11] V. Arrigoni, N. Bartolini, and A. Massini, “Topology agnostic bounds on minimum requirements for network failure identification,” *IEEE Access*, vol. 9, pp. 6076–6086, 2021.
[12] G. Wiener, É. Hosszu, and J. Tapolcai, “On separating systems with bounded set size,” *Discrete Appl. Math.*, vol. 276, pp. 172–176, Apr. 2020.
[13] L. Ma, T. He, K. K. Leung, A. Swami, and D. Towsley, “Monitor placement for maximal identifiability in network tomography,” in *Proc. IEEE INFOCOM*, 2014, pp. 1447–1455.
[14] A. Rényi, “On random generating elements of a finite Boolean algebra,” *Acta Scientiarum Mathematicarum*, vol. 22, nos. 1–2, pp. 75–81, 1961.
[15] A. Hyttinen, F. Eberhardt, and P. O. Hoyer, “Experiment selection for causal discovery,” *J. Mach. Learn. Res.*, vol. 14, pp. 3041–3071, Jan. 2013.
[16] G. Katona, “On separating systems of a finite set,” *J. Comb. Theory*, vol. 1, no. 2, pp. 174–194, 1966.
[17] K. Bryan and T. Leise, “Making do with less: An introduction to compressed sensing,” *SIAM Rev.*, vol. 55, no. 3, pp. 547–566, 2013.
[18] W. Lu, T. Dai, and S.-T. Xia, “Binary matrices for compressed sensing,” *IEEE Trans. Signal Process.*, vol. 66, no. 1, pp. 77–85, Jan. 2018.
[19] M. Lotfi and M. Vidyasagar, “Compressed sensing using binary matrices of nearly optimal dimensions,” *IEEE Trans. Signal Process.*, vol. 68, pp. 3008–3021, May 2020.
[20] S. Knight, H. X. Nguyen, N. Falkner, R. Bowden, and M. Roughan, “The Internet topology zoo,” *IEEE J. Sel. Areas Commun.*, vol. 29, no. 9, pp. 1765–1775, Oct. 2011.
[21] A. Singla, C.-Y. Hong, L. Popa, and P. B. Godfrey, “Jellyfish: Networking data centers randomly,” in *Proc. USENIX NSDI*, 2012, pp. 225–238.

Open Access funding provided by ‘Università degli Studi Roma “La Sapienza”’ within the CRUI CARE Agreement