Excitation and Measurement Patterns for the Identifiability of Directed Acyclic Graphs

Eduardo Mapurunga, Michel Gevers, *Life Fellow, IEEE,* and Alexandre S. Bazanella, *Senior Member, IEEE*

Abstract—This paper deals with the design of Excitation and Measurement Patterns (EMP) for the identification of a class of dynamical networks whose topology has the structure of a Directed Acyclic Graph (DAG). In addition to the by now well known condition that the identifiabiltiy of any dynamical network requires that the sources be excited, the sinks be measured, and all other nodes be either excited or measured, we show that for DAGs two other types of nodes have special excitation and measurement requirements. Armed with this result, we propose a systematic procedure for the design of EMPs that guarantees identifiability of a network with DAG topology.

Index Terms—Dynamic Networks, Generic Identifiability, Network Identification, Directed Acyclic Graphs.

I. INTRODUCTION

This work deals with identifiability of dynamic networks, which has been an active research topic in the control community over the last decade. The network framework used here was introduced in [1], where signals were represented as nodes of the network which were related to other nodes through transfer functions. These networks can be interpreted as directed graphs where the transfer functions, also called modules, are the edges of the graph and the node signals are the vertices.

In [1], it was assumed that all nodes are excited and measured. As a result, an input-output matrix of the network, denoted $T(z)$, can be defined, which can always be identified from these data. The network identifiability question is then whether the network matrix, denoted $G(z)$ (whose elements are the transfer functions relating the nodes) can be recovered from this closed-loop transfer matrix $T(z)$. In subsequent works, a range of new objectives were defined, from the identification of the whole network to identification of some specific part of the network [2]–[9]. As for the assumptions on the signals, up to 2019, all contributions assumed that either all nodes are excited, or all nodes are measured. A typical question would be: given that all nodes are excited,

Eduardo Mapurunga and Alexandre S. Bazanella are with the Data Driven Control Group, Department of Automation and Energy, Universidade Federal do Rio Grande do Sul, Porto Alegre-RS, Brazil, {eduardo.mapurunga, bazanella}@ufrgs.br

Michel Gevers is with the Institute of Information and Communication Technologies, Electronics and Applied Mathematics (ICTEAM), UCLouvain, Louvain la Neuve, Belgium, michel.gevers@uclouvain.be.

which nodes must be measured in order to identify the whole network?

The first identifiability results for networks where not all nodes are excited AND not all nodes are measured were presented in [10]. That paper first provided a necessary condition for identifiability of any network: each node must be either excited or measured, at least one node must be excited and at least one node measured. The paper [10] also presented identifiability conditions for two special classes of networks, namely trees and loops.

The results of [10] inspired the definition of an excitation and measurement pattern (EMP), namely the combination of excited nodes and measured nodes. The concept of EMP was introduced in [11] where an EMP was called *valid* if it guarantees the identifiability of the whole network. An EMP was called *minimal* if it guarantees the identifiability of the network using the smallest possible combination of excited and measured nodes [11]. This number is the *cardinality of the EMP*. Achieving identifiability of a network with a minimal EMP is of both theoretical and practical interest. The excitation of a node typically requires an actuator, while its measurement requires a measurement device. On the other hand, having some flexibility in the choice of a valid EMP is also of practical interest. It may be that exciting node 42, say, is prohibitively expensive while its measurement is easy; conversely, measuring a node may be difficult while its excitation is cheap. In evaluating the choice of an EMP for the identification of a network, one must of course remember that each node must be either excited or measured or both. As a result, the cardinality of a valid EMP is always at least equal to n, the number of nodes.

The search for valid, and possibly minimal, EMPs began by looking at special structures. In [10] a necessary and sufficient condition was given for the identifiability of a tree, which shows that a tree can possibly be identified with an EMP of cardinality n. In [12] necessary and sufficient conditions were derived for the identifiability of some classes of parallel networks. In [13] necessary and sufficient conditions were given for the identifiability of loops. This result showed that any loop with more than 3 nodes can also be identified with a minimal EMP of cardinality n. In addition, it was shown that constructing EMPs for loops is very easy and that the number of minimal EMPs grows very quickly with the number of nodes.

In this paper, we generalize the results derived in [10] for the identification of trees to a much wider class of networks,

namely those that have the structure of a Directed Acyclic Graph (DAG), i.e. a directed graph that has no cycles. DAGs have been widely studied in the literature [14]. A specific feature of a DAG is that the corresponding network matrix $G(z)$ can be rewritten in a lower triangular form via a relabeling of the nodes. The corresponding input-output transfer matrix $T(z)$ is then also lower-triangular and this greatly simplifies the relations between the elements $G_{ij}(z)$ of the network matrix and the elements $T_{ij}(z)$ of this transfer matrix.

The main contributions of this paper are as follows. First, we provide an explicit solution for the elements $G_{ij}(z)$ as a function of elements of the matrix $T(z)$ and elements of its inverse: $S(z) = T^{-1}(z)$, with the property that the elements $S_{ij}(z)$ of $S(z)$ are expressed in terms of $T_{kl}(z)$ only (i.e. they do not involve inverses of elements $T_{kl}(z)$). Next we focus on the construction of valid EMPs for the identification of a DAG. We provide a necessary condition on the excitation and measurement of specific nodes within a DAG. Finally, we provide a simple procedure for the construction of a valid EMP for the identification of a DAG.

The paper is organized as follows. In Section II we introduce the notations, and recall the definitions of generic identifiability of a network, and of a valid, as well as a minimal, Excitation and Measurement Pattern. We also recall the main necessary condition for identifiability of any network. In Section III we present networks that have the topology of Directed Acyclic Graphs and establish the key results that will allow one to identify such networks. In Section IV we establish necessary conditions for the identifiability of a DAG. The results of Sections III and IV allow us to propose, in Section V, a recursive procedure for the construction of a valid EMP, i.e. one that secures identifiability of the network. Finally, we present conclusions in Section VI.

II. DEFINITIONS, NOTATIONS, AND CONCEPTS

In this section, we briefly state the identifiability problem for a dynamical network and recall the main necessary condition for the identifiability of any such network. We then define the concept of a valid Excitation and Measurement Pattern, namely a choice of excited nodes and measured nodes that guarantees identifiability. We also introduce the notations used throughout the paper.

We consider dynamic networks composed of n nodes (or vertices) which represent internal scalar signals $\{w_k(t)\}$ for $k \in \{1, 2, \ldots, n\}$. These nodes are interconnected by discrete time transfer functions, represented by edges, which are entries of a *network matrix* $G(z)$. The dynamics of the network is given by the following equations:

$$w(t) = G(z)w(t) + Br(t), \tag{1a}$$

$$y(t) = Cw(t), \tag{1b}$$

where $w(t) \in \mathbb{R}^n$ is the node vector, $r(t) \in \mathbb{R}^m$ is the input and $y(t) \in \mathbb{R}^p$ is the network's output. The matrix $B \in \mathbb{Z}_2^{n \times m}$, where $\mathbb{Z}_2 \triangleq \{0, 1\}$, is a binary selection matrix with a single 1 and $n-1$ zeros in each column; it selects the inputs affecting the nodes of the network. Similarly, $C \in \mathbb{Z}_2^{p \times n}$ is a matrix

with a single 1 and $n-1$ zeros in each row that selects which nodes are measured.

We now introduce some definitions and notations concerning these dynamical networks and their network matrix $G(z)$. To each $G(z)$ we can associate a directed graph $\mathcal{G}$ defined by the tuple $(\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of vertices and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges. The graph $\mathcal{G}$ defines the topology of the network. A particular transfer function $G_{ij}(z)$ of the network matrix is called an incoming edge of node i and outgoing edge of node j. Furthermore, for this transfer function, we say that node i is an out-neighbor of node j, and that node j is an in-neighbor of node i. A node j is *connected* to node i if there exists a directed edge from node j to node i. For the graph $\mathcal{G}$ associated to the network matrix $G(z)$ we introduce the following notations.

- $\mathcal{W}$ - the set of all n nodes;
- $\mathcal{B}$ - the set of excited nodes, defined by B in (1a);
- $\mathcal{C}$ - the set of measured nodes, defined by C in (1b);
- $\mathcal{F}$ - the set of sources: nodes with no incoming edges;
- $\mathcal{S}$ - the set of sinks: nodes with no outgoing edges;
- $\mathcal{I}$ - the set of internal nodes, i.e. nodes that are neither a source nor a sink: $\mathcal{I} \triangleq \mathcal{W} \backslash (\mathcal{F} \cup \mathcal{S})$;
- $\mathcal{N}_j^-$ - the set of in-neighbors of node j;
- $\mathcal{N}_j^+$ - the set of out-neighbors of node j.

Additionally, we introduce the following two types of nodes:

- A node j is called a **dource** if all its in-neighbors have a directed edge to at least one of its out-neighbors;
- A node j is called a **dink** if it has at least one in-neighbor that has a directed edge to all its out-neighbors.

Assumptions on the network matrix $G(z)$:

- the diagonal elements are zero;
- $(I - G(z))^{-1}$ is proper and all its elements are stable.

One can represent the dynamic network in (1a)-(1b) as an input-output model as follows

$$y(t) = M(z)r(t), \text{ with } M(z) \triangleq CT(z)B. \tag{2}$$

where

$$T(z) \triangleq (I - G(z))^{-1}. \tag{3}$$

Observe that the matrix $T(z)$ is nonsingular by construction.

It is assumed that the input-output model $M(z)$ is known; the identification of $M(z)$ from input-output (IO) data $\{y(t), r(t)\}$ is a standard identification problem, provided the input signal $r(t)$ is sufficiently rich. The question of identifiability of the network is whether the network matrix $G(z)$ can be fully recovered from the transfer matrix $M(z)$. We now give a formal definition of generic identifiability of the network matrix from the data $\{y(t), r(t)\}$ and from the graph structure.

Definition II.1. *([5]) The network matrix $G(z)$ is generically identifiable from excitation signals applied to $\mathcal{B}$ and measurements made at $\mathcal{C}$ if, for any rational transfer matrix parametrization $G(P, z)$ consistent with the directed graph associated with $G(z)$, there holds*

$C[I - G(P, z)]^{-1}B = C[I - \tilde{G}(z)]^{-1}B \implies G(P, z) = \tilde{G}(z),$

for all parameters P except possibly those lying on a zero measure set in $\mathbb{R}^N$, where $\tilde{G}(z)$ is any network matrix consistent with the graph.

In this paper, we discuss identifiability in terms of which nodes must be excited and/or measured in the subsets $\mathcal{B}$ and $\mathcal{C}$ in order to guarantee identifiability of the network. This approach is inspired by [3]. The following Proposition gives a necessary condition for generic identifiability of a network; it combines Theorem III.1 and Corollary III.1 of [10].

Proposition II.1. *The network matrix $G(z)$ is generically identifiable only if $\mathcal{B}, \mathcal{C} \neq \emptyset$, $\mathcal{F} \subset \mathcal{B}$, $\mathcal{S} \subset \mathcal{C}$ and $\mathcal{B} \cup \mathcal{C} = \mathcal{W}$.*

As a consequence, every node of the network must be either excited or measured. Generic identifiability of a given network can thus be equivalently characterized by the network's Excitation and Measurement Pattern, denoted EMP. The concept of EMP, which led to the concept of minimal EMP, was introduced in [11]. They are defined in the following.

Definition II.2. *A pair of selection matrices B and C, with its corresponding pair of node sets $\mathcal{B}$ and $\mathcal{C}$, is called an excitation and measurement pattern - EMP for short. An EMP is said to be* valid *if it is such that the network (1a)-(1b) is generically identifiable. Let $\nu = |\mathcal{B}| + |\mathcal{C}|$ [1] be the cardinality of an EMP. A given EMP is said to be* minimal *if it is valid and there is no other valid EMP with smaller cardinality.*

The following result establishes a lower and upper bound for the cardinality of a valid EMP for any network.

Theorem II.1. *The cardinality of a valid EMP for the identification of a dynamical network with n nodes is at least equal to n and at most equal to $2n - f - s$, where f is the number of sources and s the number of sinks.*

Proof. The lower bound results from Proposition II.1; it can actually be achieved for trees and loops [10], [13]. As for the upper bound, we know by Proposition II.1 that all sources must be excited and all sinks measured, while the remaining $n - f - s$ nodes must be excited or measured. Assuming that these are all excited and measured, then the cardinality of the EMP is $f + s + 2(n - f - s) = 2n - f - s$. ∎

From now on, we drop the arguments z and t used in (1a)-(1b) whenever there is no risk of confusion.

III. DIRECTED ACYCLIC GRAPHS AND THEIR PROPERTIES

In this section we investigate the generic identifiability of dynamic networks whose topologies are associated with directed acyclic graphs, denoted DAG[2]. These are very general classes of graphs, of which trees are a special case. We will derive necessary conditions for the generic identifiability of these classes of networks and characterize which nodes need to be excited or/and measured to obtain a valid EMP for these networks. Directed Acyclic Graphs are defined as follows.

---

[1] $|\cdot|$ - Denotes the cardinality of a set.

[2] For simplicity, we shall in the future just refer to a DAG rather than a network that has the topology of a DAG.

Definition III.1. *A directed acyclic graph is a directed graph that has no cycles.*

A property of DAGs is that the sequence of their nodes can be relabeled by a topological sorting algorithm [15] in such a way that $G_{ij} = 0$ for $i < j$. In the sequel, we assume without loss of generality that the nodes of the dynamic networks we study in this section have been relabeled this way. As a result, the network matrix G can be written as a lower triangular matrix.

$$
G = \begin{bmatrix}
0 & 0 & 0 & \ldots & 0 \\
G_{21} & 0 & 0 & \ldots & 0 \\
G_{31} & G_{32} & 0 & \ldots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
G_{n1} & G_{n2} & \ldots & G_{n,n-1} & 0
\end{bmatrix}, \tag{4}
$$

where some $G_{ij}, i > j$ are typically zero. In [12] it was shown that for a network matrix (4) with all $G_{ij} \neq 0$ for $i > j$ generic identifiability is achieved if and only if all sources are excited, all sinks are measured, and every other node is both excited and measured. Exciting and measuring all internal nodes is of course a very strong condition; we shall explain in the next section why it occurs when all $G_{ij} \neq 0$.

The following Lemma establishes relationships between such network matrix G with the structure of a DAG and the corresponding I/O matrix T.

Lemma III.1. *Let G be as in (4) and define $T = (I - G)^{-1}$. Then the following relationships hold.*

$$T_{ll} = 1, \tag{5}$$

$$T_{lj} = 0, \ \text{for } j > l, \tag{6}$$

$$T_{lj} = \sum_{i=j}^{l-1} G_{li} T_{ij}, \ \text{for } l > j \tag{7}$$

$$T_{lj} = \sum_{i=j+1}^{l} T_{li} G_{ij}, \ \text{for } l > j \tag{8}$$

$$G_{lj} = T_{lj} - \sum_{i=j+1}^{l-1} G_{li} T_{ij}, \ \text{for } l > j \tag{9}$$

$$G_{lj} = T_{lj} - \sum_{i=j+1}^{l-1} T_{li} G_{ij}, \ \text{for } l > j \tag{10}$$

Proof. The relations (5)-(7) follow directly from $[I - G]T = I_n$, while (8) follows from $T[I - G] = I_n$. Observing that $T_{jj} = 1$ in (7) yields (9), while (10) follows similarly from (8). ∎

In the sequel of this paper we shall illustrate all our results with the following 7-node DAG.

For this network, equation (10) allows us, for example, to write:

$$G_{43} = T_{43}$$
$$G_{53} = T_{53} - T_{54} G_{43}$$
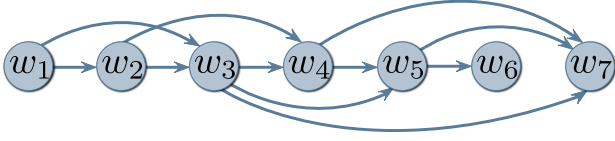$$G_{73} = T_{73} - T_{74} G_{43} - T_{75} G_{53}$$

Fig. 1. A 7-node DAG network

Equations (9)-(10) define the expressions of the G_{ij} recursively as a function of the T_{ij} and of the previously computed G_{kl}. But the G_{ij} can also be expressed explicitly as functions of the T_{ij} as is shown in the following Theorem, which is one of the main results of this paper. It shows that for DAGs one can compute the unknown transfer functions G_{ij} explicitly as a function of the elements of the I/O matrix.

Theorem III.1. *Define $S \triangleq T^{-1}$ with elements S_{ij}. Then the following results hold:*
(1) $G_{lj} = -S_{lj}$, and hence $S_{lj} = 0$ for each pair $\{l, j\}$, $l \neq j$, for which G_{lj} is known to be zero;
(2) G_{lj} can be written as a sum of products of T_{ki} with $k \leq l$ and $i \geq j$.

Proof. Item (1) follows from $G = I - T^{-1}$ and the fact that the inverse of a lower triangular matrix whose diagonal elements are all equal to one is a lower triangular matrix with the same property. As for item (2), it follows from the direct computation of T^{-1}, with the property that the inverse of a lower triangular matrix whose diagonal elements are all equal to one does not contain any T_{ij}^{-1}. But it also follows by substituting the G_{ij} on the right hand side of (10) by their expressions computed from the same equation. ∎

To illustrate the result (2) of Theorem III.1, we observe that if we substitute G_{53} and G_{43} in the expression of G_{73} above by their expressions, we obtain

$$G_{73} = T_{73} - T_{74}T_{43} - T_{75}T_{53} + T_{75}T_{54}T_{43}$$

as claimed.

Not only does Theorem III.1 provide an explicit expression for the G_{lj} as a function of the input-output elements T_{lk}, but it also proves very useful to establish relations among the T_{lk}. Indeed, for each G_{lj} that is known to be zero, result (2) of the Theorem allows one to compute one of the T_{lk} as a function of others. This may then allow one to eliminate the need for the excitation or measurement of some nodes as we shall show in Section V. Returning again to the 7-node DAG of Figure 1, since $G_{52} = 0$, it follows that $S_{52} = 0$, which implies

$$T_{52} = T_{53}T_{32} + T_{54}T_{42} + T_{54}T_{43}T_{42}.$$

IV. Necessary conditions for the identification of a DAG

It was shown in [10] that a necessary condition for the generic identifiability of any network is that all sources must be excited, all sinks must be measured, and that each other node must be either excited or measured. Here we show that, for DAGs, some additional necessary conditions are required

for two special classes of nodes, namely the dources and the dinks that were defined in Section II. Our main result is the following.

Theorem IV.1. *Consider a dynamic network with the topology of a directed acyclic graph whose network matrix is given in (4) with some known $G_{ij} = 0$ for $i > j$. Then this network is generically identifiable only if the following conditions hold.*
(1) each node is either excited or measured;
(2) all sources are excited and all sinks are measured;
(3) all dources are excited and all dinks are measured.

Proof. Conditions (1) and (2) have been shown to be necessary for the identification of any network in [10].
We turn to item (3). We first prove that the excitation of all dources is necessary. Consider a node l that is an out-neighbor of i such that all in-neighbors of node i are connected to that out-neighbor l. This means that node i is a dource. It then follows that for each in-neighbor j of node i, we have

- $G_{ij} \neq 0$ since j is an in-neighbor of i;
- $G_{lj} \neq 0$ since all in-neighbors of i are connected to l.

To show that the dource i needs to be excited, we focus on the transfer function T_{li} between the dource i and the out-neighbor l that is connected to all in-neighbors of i. It follows from (10) that in the equations defining G_{ij} as a function of the T_{lk} and the other G_{ij}, this specific transfer function T_{li} appears only in the expressions of G_{lj} in which l is the considered out-neighbor of dource i, and j is either the dource i itself or one of its $\mathcal{N}_i^-$ in-neighbors. We now observe that in these $1 + \mathcal{N}_i^-$ equations, T_{li} is either multiplied by '1' or by the G_{ij} that relate the in-neighbors of dource i to itself. These G_{ij} are nonzero (see above). Each of these $1 + \mathcal{N}_i^-$ equations therefore contains an unknown G_{lj} on the left hand side and the unknown T_{li} on the right hand side. Hence, node i must be excited in order to compute T_{li} and therefore the edges G_{lj} linking the dource i and all its in-neighbors to this particular out-neighbor.
The proof for the necessity of measuring all dinks is the exact dual of the proof for the necessity of exciting all dources and will therefore be omitted. ∎

We illustrate the result of Theorem IV.1 with the 7-node DAG network pictured in Figure 1. This network has one source (node 1), two dources (nodes 2 and 5), two sinks (nodes 6 and 7) and one dink (node 4).

We first show why node 5 must be excited. It is a dource because all its in-neighbors (nodes 3 and 4) are connected to out-neighbor 7. We write the equations (10) that relate the out-neighbor to the dource and to its two in-neighbors.

$$G_{75} = T_{75}$$
$$G_{74} = T_{74} - T_{75}G_{54}$$
$$G_{73} = T_{73} - T_{74}G_{43} - T_{75}G_{53}$$

We observe that, even if all quantities other than T_{75} on the right hand side were known (i.e. $T_{74}, T_{73}, G_{54}, G_{43}, G_{53}$), it would be impossible to identify the transfer functions G_{75}, G_{74} and G_{73} without knowing T_{75}, which requires

exciting node 5. We add that, if an edge G_{76} was added, the first equation would be replaced by $G_{75} = T_{75} - G_{76}T_{65}$, and the conclusion would thus be identical.

We now illustrate why node 4 must be measured. It is a dink because it has an in-neighbor (node 3) that is connected to all its out-neighbors (nodes 5 and 7). We write the equations (9) that relate its in-neighbor 3 to the dink and to all its out-neighbors.

$$G_{43} = T_{43}$$
$$G_{53} = T_{53} - G_{54}T_{43}$$
$$G_{73} = T_{73} - G_{74}T_{43} - G_{75}T_{53}$$

We observe that T_{43} is required to compute G_{43}, and that it cannot be computed from the other two equations for the same reasons as above. Thus, node 4 must be measured for the identification of G_{43}, G_{53} and G_{73}.

Notice that a given node can be both a dource and a dink. In a "full" DAG as in (4), with all $G_{ij} \neq 0$, all internal nodes are dources and dinks, which explains the need for exciting and measuring them all, as shown in [12].

V. Constructing a valid EMP for a DAG

In this section, we show how to construct a valid EMP while trying to keep the cardinality of this EMP low. Recall that this cardinality is always between n and $2n - f - s$.

An easy solution for the construction of a valid EMP results from Theorem III.1. Indeed, each G_{ij} can be expressed explicitly as a combination of T_{kl}. The collection of all these T_{kl} indicates which node excitations and which node measurements will lead to a valid EMP. In order to reduce the cardinality of this explicit solution, one can, in a second step, take advantage of the possible replacement of some of these T_{kl} by others using the equations $S_{ij} = 0$ for each G_{ij} that is known to be zero: see item (1) of Theorem III.1.

For the 7-node DAG of Figure 1 the explicit solution provided by Theorem III.1 yields the following solution for the G_{ij}:

$$G_{21} = T_{21}$$
$$G_{31} = T_{31} - T_{21}T_{32}, G_{32} = T_{32}$$
$$G_{42} = T_{42} - T_{32}T_{43}, G_{43} = T_{43}$$
$$G_{53} = T_{53} - T_{43}T_{54}, G_{54} = T_{54}$$
$$G_{65} = T_{65}$$
$$G_{73} = T_{73} - T_{75}T_{53} - T_{74}T_{43} + T_{75}T_{54}T_{43}$$
$$G_{74} = T_{74} - T_{54}T_{75}, G_{75} = T_{75}$$

Collecting all the indices that appear as inputs and outputs of the T_{kl} yields the valid EMP $\mathcal{B} = \{1, 2, 3, 4, 5\}$ and $\mathcal{C} = \{2, 3, 4, 5, 6, 7\}$. It has cardinality 11. Observe that nodes 2, 3, 4, 5 are both excited and measured. According to Theorem IV.1, nodes 2 and 5 are dources and node 4 is a dink. Hence, nodes 2 and 5 must be excited, as well as node 1, which is a source; while node 4 must be measured, as well as nodes 6 and 7 which are sinks. Starting from this initial EMP with cardinality 11, one can then use the

equations of Theorem III.1 and Lemma III.1 (in particular the equations $S_{ij} = 0$) to check whether one can eliminate any one of nodes 3 and 4 from $\mathcal{B}$, or any one of nodes 2, 3, 5 from $\mathcal{C}$. However, such procedure is tedious, and there appears to be no systematic way to proceed with the elimination of excitations or measurements. Therefore, we propose a recursive procedure for the construction of a valid EMP which is based on Theorem III.1 and Lemma III.1.

Recursive procedure for the construction of a valid EMP. First, build the matrix S, replacing each S_{lj} by $-G_{lj}$, where these G_{lj} are computed using equation (10) of Lemma III.1. For each G_{lj} that is known to be zero, equate the zero element in S_{ji} to the expression resulting from the same equation (10).

Now, construct the preliminary EMP with the excitations and measurements required by the structure of the DAG: sources and dources must be excited, dinks and sinks must be measured. Call it EMP$_0$, thus defining a $\mathcal{B}_0$ and a $\mathcal{C}_0$. List the corresponding known T_{ij}, i.e. all T_{ij} for which $j \in \mathcal{B}_0$ and $i \in \mathcal{C}_0$.

Now, proceed stepwise within the columns of S, say from column 1 to column n, as explained below.

1) Column 1: the unknown G_{j1} appear in the first column of S. Add to EMP$_0$ whatever excited node or measured node is required to be able to identify all G_{j1}. There may be several choices. Use the remaining $S_{j1} = 0$ equations to compute new elements T_{kl}. Update EMP$_0$ to EMP$_1$, update the known T_{kl} and update the known G_{j1}.

2) Column 2: the unknown G_{j2} appear in the second column of S. Add to EMP$_1$ whatever excited node or measured node is required to be able to identify all G_{j2}. There may be several choices. Use the remaining $S_{j2} = 0$ equations to compute new T_{kl}. Update EMP$_1$ to EMP$_2$, update the known T_{kl} and update the known G_{j2}.

3) Continue until all columns of S containing elements G_{lj} have been covered.

Before we illustrate this procedure with our Example, let us make the following comments.
Comments

- The computation of elements G_{lj} based on column j of S may require that several elements of that column be used jointly, leading to the solution of a linear system of equations. One must check that these equations are linearly independent.
- The procedure proposed above uses a column by column approach, going from left to right. Other approaches can be used, such as covering the columns from right to left, using a row by row approach, etc. These different approaches will typically lead to different valid EMPs.

The EMP procedure applied to the DAG of Figure 1.
The construction of a valid EMP from the successive columns of S is shown in Table I. For brevity of notation, for this example with 7 nodes, we represent an EMP that has $\mathcal{B} = \{1, 2, 4, 5\}$ and $\mathcal{C} = \{2, 3, 4, 6, 7\}$, say, by E1245, M23467.

TABLE I

CONSTRUCTION OF AN EMP COLUMN-WISE.

column of S	EMP update	Known T_{ji}	Known G_{ji}
	EMP$_0$: E125, M467	$T_{41}, T_{42}, T_{61}, T_{62}, T_{65}, T_{71}, T_{72}, T_{75}$	–
1	+M23: E125, M23467	$T_{21}, T_{31}, T_{32}, T_{43}$ from $S_{41}=0$, T_{63} from $S_{61}=0$, T_{73} from $S_{71}=0$	G_{21}, G_{31}
2	E125, M23467	T_{64} from $S_{62}=0$, T_{74} from $S_{72}=0$	G_{32}, G_{42}
3	E125, M23467	–	G_{43}, G_{53} from S_{63}, G_{73}
4	E125, M23467	–	G_{54} from S_{64}, G_{74}
5	E125, M23467	–	G_{65}, G_{75}

We know a priori that nodes 1, 2 and 5 must be excited, being a source and two dources, while nodes 4, 6 and 7 must be measured, being a dink and two sinks. As a result, the starting EMP, denoted EMP$_0$ above, is E125, M467. With this initial EMP$_0$, the elements $T_{41}, T_{42}, T_{61}, T_{62}, T_{65}, T_{71}, T_{72}, T_{75}$ are known. In Table I, we have listed the a priori information that results from EMP0. By adding M23, i.e. the measurement of nodes 2 and 3, we know T_{21}, T_{31}, T_{32}. In addition, the expressions $S_{l1} = -G_{l1}$ of the first column of S, established in (10), allow one to compute G_{21}, G_{31}, as well as T_{43}, T_{63}, T_{73}; these are derived from $S_{41} = 0, S_{61} = 0$, and $S_{71} = 0$, respectively. Proceeding similarly with the elements of the second column of S yields directly G_{32}, G_{42}. It also allows to compute T_{64} from $S_{62} = 0$ and T_{74} from $S_{72} = 0$. Continuing with columns 3, 4 and 5 of the matrix S subsequently yields all other elements G_{lj}. The procedure ends with a valid EMP defined as E125, M23467, which has cardinality 8.

An alternative is to add E3 for the identification of the elements of column 1, and E4 for the identification of the elements of column 2. This yields the alternative valid EMP defined as E12345, M467, with the same cardinality 8. A row by row procedure, from top to bottom, applied to the same example leads to the valid EMP: E125, M234567, which has cardinality 9. In comparison, the explicit solution based on the decomposition of each G_{lj} as a function of the T_{kl} only, as explained above, has cardinality 11.

VI. CONCLUSIONS

We have pursued our study of specific substructures of dynamic networks, with the aim of designing EMPs that guarantee identifiability for these specific substructures. In [5] a necessary and sufficient condition was established for trees, with a corresponding valid (and minimal) EMP. In [12] necessary and sufficient conditions for network structures with parallel paths were derived. In [13] it was shown that any loop with more than 3 nodes can also be identified with a minimal EMP and the design of such loops was presented. In the present paper we have focused on networks that have the structure of a DAG. We have shown that DAGs have specific properties which facilitate the design of valid EMPs. A rather surprising result of our analysis has been to show that two specific and well-defined nodes, which we have called dources and dinks, have specific excitation and measurement requirements. Dources, just like sources, must be excited and dinks, just like sinks, must be measured in order for a DAG network to be identifiable. With these constraints under our belt, we have produced a recursive EMP design procedure that takes advantage of the properties of DAGs. The procedure we have proposed guarantees that the constructed EMP is valid. However, there is no guarantee that it is minimal. The search for an algorithm that constructs a minimal EMP is a topic for future research.

REFERENCES

[1] P. M. J. Van den Hof, A. Dankers, P. S. C. Heuberger, and X. Bombois, "Identification of dynamic models in complex networks with prediction error methods—Basic methods for consistent module estimates," *Automatica*, vol. 49, no. 10, pp. 2994–3006, Oct. 2013.

[2] M. Gevers, A. S. Bazanella, and A. Parraga, "On the identifiability of dynamical networks," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 10 580–10 585, Jul. 2017.

[3] A. S. Bazanella, M. Gevers, J. M. Hendrickx, and A. Parraga, "Identifiability of dynamical networks: Which nodes need be measured?" in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, Dec. 2017, pp. 5870–5875.

[4] S. Shi, X. Cheng, and P. M. J. Van den Hof, "Generic identifiability of subnetworks in a linear dynamic network: The full measurement case," *Automatica*, vol. 137, p. 110093, Mar. 2022.

[5] J. M. Hendrickx, M. Gevers, and A. S. Bazanella, "Identifiability of Dynamical Networks With Partial Node Measurements," *IEEE Transactions on Automatic Control*, vol. 64, no. 6, pp. 2240–2253, Jun. 2019.

[6] N. Everitt, G. Bottegal, and H. Hjalmarsson, "An empirical Bayes approach to identification of modules in dynamic networks," *Automatica*, vol. 91, pp. 144–151, May 2018.

[7] M. Gevers, A. S. Bazanella, and G. V. da Silva, "A practical method for the consistent identification of a module in a dynamical network," *IFAC-PapersOnLine*, vol. 51, no. 15, pp. 862–867, Jan. 2018.

[8] H. J. Van Waarde, P. Tesi, and M. K. Camlibel, "Necessary and Sufficient Topological Conditions for Identifiability of Dynamical Networks," *IEEE Transactions on Automatic Control*, vol. 65, no. 11, pp. 4525 – 4537, 2019.

[9] S. Jahandari and D. Materassi, "Sufficient and necessary graphical conditions for miso identification in networks with observational data," *IEEE Transactions on Automatic Control*, pp. 1–1, 2021.

[10] A. S. Bazanella, M. Gevers, and J. M. Hendrickx, "Network identification with partial excitation and measurement," in *2019 IEEE 58th Conference on Decision and Control (CDC)*, Dec. 2019, pp. 5500–5506.

[11] E. Mapurunga and A. S. Bazanella, "Optimal Allocation of Excitation and Measurement for Identification of Dynamic Networks," *IFAC-PapersOnLine*, vol. 54, no. 7, pp. 43–48, Jan. 2021.

[12] ——, "Identifiability of Dynamic Networks from Structure," *IFAC-PapersOnLine*, vol. 54, no. 7, pp. 55–60, Jan. 2021.

[13] E. Mapurunga, M. Gevers, and A. S. Bazanella, "Necessary and sufficient conditions for the identifiability of isolated loops," *IEEE Control Systems Letters*, pp. 2276–2280, 2022.

[14] J. Bang-Jensen and G. Gutin, *Digraphs: Theory, Algorithms, and Applications*, 2nd ed., ser. Springer Monographs in Mathematics. London: Springer, 2009.

[15] A. B. Kahn, "Topological sorting of large networks," *Commun. ACM*, vol. 5, no. 11, p. 558–562, nov 1962.