



# Identifiable Paths and Cycles in Linear Compartmental Models

Cashous Bortner<sup>1</sup> · Nicolette Meshkat<sup>2</sup> 

Received: 19 January 2021 / Accepted: 27 January 2022 / Published online: 19 March 2022  
© The Author(s), under exclusive licence to Society for Mathematical Biology 2022

## Abstract

We introduce a class of linear compartmental models called *identifiable path/cycle models* which have the property that all of the monomial functions of parameters associated to the directed cycles and paths from input compartments to output compartments are identifiable and give sufficient conditions to obtain an identifiable path/cycle model. Removing leaks, we then show how one can obtain a locally identifiable model from an identifiable path/cycle model. These identifiable path/cycle models yield the *only* identifiable models with certain conditions on their graph structure and thus we provide necessary and sufficient conditions for identifiable models with certain graph properties. A sufficient condition based on the graph structure of the model is also provided so that one can test if a model is an identifiable path/cycle model by examining the graph itself. We also provide some necessary conditions for identifiability based on graph structure. Our proofs use algebraic and combinatorial techniques.

**Keywords** Structural identifiability · Linear compartmental model · Identifiable functions of parameters · Identifiable combinations

## 1 Introduction

The parameter identifiability problem is the question of whether or not the unknown parameters of a mathematical model can be determined from known data. This paper is concerned with *structural identifiability* analysis, that is, whether the model parameters can be identified from perfect input–output data (noise-free and of any duration required). Structural identifiability is a necessary condition for *practical identifiability* which is identifiability analysis in the presence of noisy and imperfect data. Thus, structural identifiability is an important step in the parameter estimation problem,

---

✉ Nicolette Meshkat  
nmeshkat@scu.edu

<sup>1</sup> North Carolina State University, NC, USA

<sup>2</sup> Santa Clara University, Santa Clara, USA

since failure to recover parameters in the ideal case implies failure in the imperfect case as well. If all of the parameters of a model can be determined, we say the model is (at least) locally identifiable, but if some subset of the parameters can take on an infinite number of values yet yield the same input–output data, the model is said to be unidentifiable.

In this work, we examine a special class of models called *linear compartmental models*. Linear compartmental models are an important class of biological models used in the areas of cell biology, pharmacology, toxicology, ecology, physiology, and many other areas (DiStefano 2015). In a typical biological application, the mass or concentration of a substance (e.g., drug concentration in an organ) is represented by a compartment, and the transfer of material from one compartment to another is given by a constant rate parameter, called an *exchange rate*. The transfer of material from a compartment leaving the system is given by a constant rate parameter called the *leak rate*, and any compartment containing such a leak is called a *leak compartment*. An input represents the input of material to a particular compartment of the system (e.g., IV drug input) and an output represents a measurement from a compartment (e.g., drug concentration in an organ), where such compartments are called *input compartments* and *output compartments*, respectively. The resulting ODE system of equations (see Eq. (1)) is linear. This linearity feature has a nice mathematical consequence in that the model can be represented by a directed graph. Thus, we can analyze identifiability problems in terms of the combinatorial structure of that graph.

Recent work on identifiable reparametrizations (Meshkat and Sullivant 2014; Baaijens and Draisma 2016), sufficient conditions for identifiability (Meshkat et al. 2015; Gross et al. 2019; Gerberding et al. 2020), and identifiable functions of parameters (Meshkat and Sullivant 2014; Meshkat et al. 2018) has examined the combinatorial structure of the graph in a linear compartmental model to answer questions about what to do with an unidentifiable model. One approach to dealing with an unidentifiable model is to reparametrize the model over identifiable functions of parameters in the model. In other words, although not all the parameters are identifiable, one can attempt to reparametrize over a set of functions of parameters that can be determined from input–output data. In Meshkat and Sullivant (2014), necessary and sufficient conditions were given to obtain an identifiable scaling reparametrization in the case of a linear compartmental model with a single input and output in the same compartment, leaks from every compartment, and having a strongly connected graph. In Meshkat et al. (2015), these models were called *identifiable cycle models* because the monomial functions associated to the directed cycles are identifiable and it was shown that removing all but one leak from such models results in identifiability. Additionally, it was shown that removing a subset of leaks, but adding inputs or outputs to the remaining leak compartments, results in identifiability.

In this paper, we expand upon the results in Meshkat and Sullivant (2014); Meshkat et al. (2015) in the following ways. First, we consider the case of inputs and outputs not necessarily in the same compartment and define the analogous *identifiable path/cycle model* (Definition 3.4), which is a model where all the monomial functions associated to the directed cycles and paths from input to output are identifiable. Just as in Meshkat and Sullivant (2014), this occurs when the model has a coefficient map whose image has maximal dimension (Theorem 3.13). We then take these identifiable path/cycle models

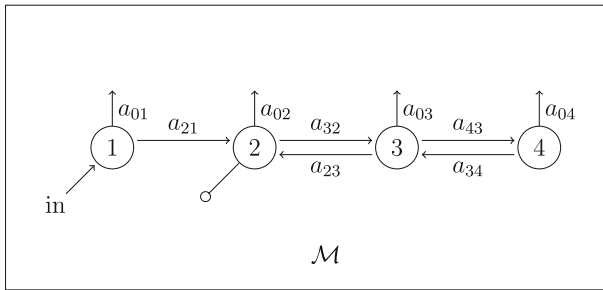


Fig. 1 Graph for Example 2.1

and remove leaks from all compartments except input/output compartments to achieve identifiable models (Theorem 3.19). A similar result was demonstrated in Meshkat et al. (2015), but in that version, the intersection of input and output compartments was nonempty, whereas in the present work the input and output compartments need not coincide. We then show that these identifiable path/cycle models yield the *only* identifiable models with certain conditions on their graph structure (Theorem 4.1). We thus provide necessary and sufficient conditions for identifiable models with certain graph properties (Corollary 4.3). We also give a sufficient condition for a model to be an identifiable path/cycle model which can be tested simply by examining the graph itself (Theorem 3.22). In addition, we weaken the conditions on the graph structure to obtain some necessary and sufficient conditions for identifiability (Corollary 5.11). We also give some necessary conditions for identifiability in terms of the structure of the graph (Theorems 6.1, 6.3, 6.4). Finally, we give a construction of identifiable models using results from Baaijens and Draisma (2016) (Algorithm 9.4).

Our results apply to a large class of linear compartmental models which arise in many real-world applications. Path models of the form in Proposition 3.29 arise in physiological models involving metabolism, biliary, or excretory pathways (DiStefano 2015) and models of neuronal dendritic trees (Bressloff and Taylor 1993). Path models also arise when modeling the delayed response to input and are called *time-delay models* (DiStefano 2015). One such example is Example 4.13 from DiStefano (2015) on oral dosing losses and delays in the gastrointestinal tract. Some other path models are considered in Sect. 7. More generally, we consider models that are *strongly input–output connected*. Mammillary and catenary models (DiStefano 2015) fall into this category, as well as a variation of mammillary and catenary models where input and output are in distinct neighboring compartments, but the edge from output to input is missing (see Fig. 1). More generally, our results apply to models that can be thought of as path models combined with catenary models and are considered in Sect. 7. Such a model could, for example, represent a time-delay model coupled with a catenary model.

The organization of the paper is as follows. Section 2 gives the necessary background. Section 3 gives the definition of an identifiable path/cycle model and how to obtain one. Section 4 gives a classification of all identifiable models with certain graph properties. Section 5 examines weaker conditions on the graph structure for necessary and sufficient conditions for identifiability. Section 6 gives necessary conditions for

**Table 1** Summary of main results

Result	Explanation
Corollary 4.3	Gives necessary and sufficient conditions for a strongly input–output connected model to be an identifiable path/cycle model
Theorem 3.22	Gives a sufficient condition to be an identifiable path/cycle model based on graph structure
Corollary 5.11	Gives necessary and sufficient conditions for an output connectable model to be an identifiable path/cycle model

**Table 2** Summary of which new results in this paper generalize the prior results from Meshkat and Sullivant (2014) and Meshkat et al. (2015)

Prior result	New result	Explanation
Theorem 1.2 of Meshkat and Sullivant (2014)	Theorem 3.13	Generalizes conditions for identifiable cycle model to identifiable path/cycle model
Theorem 5 from Meshkat et al. (2015)	Theorem 3.19	Generalizes removing leaks to obtain identifiability
Theorem 5.13 from Meshkat and Sullivant (2014)	Theorem 3.22	Generalizes inductively strongly connected to almost inductively strongly connected
Proposition 5.4 from Meshkat and Sullivant (2014)	Proposition 3.29	Generalizes identifiable cycle to identifiable path
Proposition 5.5 from Meshkat and Sullivant (2014)	Proposition 3.30	Generalizes adding a new vertex
Proposition 5.3 from Meshkat and Sullivant (2014)	Theorems 6.3 and 6.4	Generalizes necessary condition of having an exchange to having a path

identifiability in terms of the graph structure of the model. Section 7 demonstrates our results on some real-world examples. Section 8 gives computations on the number of models with maximal dimension with a certain number of inputs and outputs. Sect. 9 gives a construction of identifiable models. Section 10 gives a conclusion and a conjecture on identifiable scaling reparametrizations of identifiable path/cycle models. We give a summary of our main results in Table 1 and a summary of which of our results generalize previous results in Table 2.

## 2 Background

Let  $G$  be a directed graph with vertex set  $V$  and set of directed edges  $E$ . Each vertex  $i \in V$  corresponds to a compartment in our model, and an edge  $j \rightarrow i$  denotes a direct flow of material from compartment  $j$  to compartment  $i$ . Also introduce three subsets of the

vertices  $In, Out, Leak \subseteq V$  corresponding to the set of input compartments, output compartments, and leak compartments respectively. To each edge  $j \rightarrow i$  we associate an independent parameter  $a_{ij}$ , the rate of flow from compartment  $j$  to compartment  $i$ . To each leak node  $i \in Leak$ , we associate an independent parameter  $a_{0i}$ , the rate of flow from compartment  $i$  leaving the system.

We associate a matrix  $A(G)$ , called the *compartmental matrix* to the graph and the set  $Leak$  in the following way:

$$A(G)_{ij} = \begin{cases} -a_{0i} - \sum_{k:i \rightarrow k \in E} a_{ki} & \text{if } i = j \text{ and } i \in Leak \\ -\sum_{k:i \rightarrow k \in E} a_{ki} & \text{if } i = j \text{ and } i \notin Leak \\ a_{ij} & \text{if } j \rightarrow i \text{ is an edge of } G \\ 0 & \text{otherwise} \end{cases}$$

For brevity, we will often use  $A$  to denote  $A(G)$ . Also, define the vector  $\mathcal{A} \in \mathbb{R}^{|E|+|Leak|}$  consisting of nonzero parameters of  $A$ .

Then we construct a system of linear ODEs with inputs and outputs associated to the quadruple  $(G, In, Out, Leak)$  as follows:

$$x'(t) = Ax(t) + u(t) \quad y_i(t) = x_i(t) \text{ for } i \in Out \tag{1}$$

where  $u_i(t) \equiv 0$  for  $i \notin In$ . The coordinate functions  $x_i(t)$  are the state variables, the functions  $y_i(t)$  are the output variables, and the nonzero functions  $u_i(t)$  are the inputs. The resulting model is called a *linear compartmental model*.

We will indicate output compartments by this symbol:  $\swarrow$ . Input compartments are labeled by “in,” and leaks are indicated by edges which go to no vertex.

**Example 2.1** The model  $\mathcal{M} = (G, \{1\}, \{2\}, V)$  with  $G$  given in Figure 1 is a linear compartmental model with equations given by:

$$\begin{pmatrix} x_1' \\ x_2' \\ x_3' \\ x_4' \end{pmatrix} = \begin{pmatrix} -a_{01} - a_{21} & 0 & 0 & 0 \\ a_{21} & -a_{02} - a_{32} & a_{23} & 0 \\ 0 & a_{32} & -a_{03} - a_{23} - a_{43} & a_{34} \\ 0 & 0 & a_{43} & -a_{04} - a_{34} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} + \begin{pmatrix} u_1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \tag{2}$$

with output equation  $y_2 = x_2$ .

For a model  $(G, In, Out, Leak)$  where there is a leak in every compartment (i.e.,  $Leak = V$ ), it can greatly simplify the representation to use the fact that the diagonal entries of  $A(G)$  are the only places where the parameters  $a_{0i}$  appear. Since these are algebraically independent parameters, we can introduce a new algebraically independent parameter  $a_{ii}$  for the diagonal entries (i.e., we make the substitution  $a_{ii} = -a_{0i} - \sum_{k:i \rightarrow k \in E} a_{ki}$ ) to get generic parameter values along the diagonal. Identifiability questions in such a model are equivalent to identifiability questions in the model with this reparametrized matrix.

We will be considering graphs that have some special connectedness properties. We define these properties now, as well as the basic algebraic structures (monomial paths and cycles) we will be working over.

**Definition 2.2** A directed graph  $G$  is *connected* if each pair of vertices in the graph is joined by an undirected path. A directed graph  $G$  is *strongly connected* if there exists a directed path from each vertex to every other vertex. A directed graph  $G$  is *inductively strongly connected* with respect to vertex 1 if each of the induced subgraphs  $G_{\{1, \dots, i\}}$  is strongly connected for  $i = 1, \dots, n$  for some ordering of the vertices  $1, \dots, i$  which must start at vertex 1.

**Definition 2.3** A *closed path* in a directed graph  $G$  is a sequence of vertices  $i_0, i_1, i_2, \dots, i_k$  with  $i_k = i_0$  and such that  $i_{j+1} \rightarrow i_j$  is an edge for all  $j = 0, \dots, k-1$ . A *cycle* in  $G$  is a closed path with no repeated vertices. To a cycle  $C = i_0, i_1, i_2, \dots, i_k$ , we associate the monomial  $a^C = a_{i_0 i_1} a_{i_1 i_2} \cdots a_{i_k i_0}$ , which we refer to as a *monomial cycle*. If a monomial cycle  $a^C$  has length  $k$ , we refer to it as a  $k$ -cycle.

Note that we also include the monomial cycles  $a_{ii}$  which are 1-cycles, or *self-cycles*.

**Definition 2.4** A *path* from vertex  $i_k$  to vertex  $i_0$  in a directed graph  $G$  is a sequence of vertices  $i_0, i_1, i_2, \dots, i_k$  such that  $i_{j+1} \rightarrow i_j$  is an edge for all  $j = 0, \dots, k-1$ . To a path  $P = i_0, i_1, i_2, \dots, i_k$ , we associate the monomial  $a^P = a_{i_0 i_1} a_{i_1 i_2} \cdots a_{i_{k-1} i_k}$ , which we refer to as a *monomial path*. If a monomial path  $a^P$  has length  $k$ , we refer to it as a  $k$ -path.

**Remark** We will sometimes drop the word “monomial” from “monomial path” or “monomial cycle” and simply refer to these as paths and cycles, as is the case in the next definition.

We now define the path/cycle map for a linear compartmental model  $\mathcal{M} = (G, In, Out, V)$ :

**Definition 2.5** Let  $\mathcal{P} = \mathcal{P}(G)$  be the set of all directed cycles and paths from input to output vertices in the graph  $G$ . Define the *path/cycle* map by:

$$\pi : \mathbb{R}^{|E|+|V|} \rightarrow \mathbb{R}^{|\mathcal{P}|}, \mathcal{A} \mapsto (a^C)_{C \in \mathcal{P}} \quad (3)$$

**Example 2.6** (Continuation of Example 2.1) Consider the model  $\mathcal{M} = (G, \{1\}, \{2\}, V)$  as described in Example 2.1. The path/cycle map for this model is

$$\begin{aligned} \pi : \mathbb{R}^9 &\rightarrow \mathbb{R}^7 \\ (a_{11}, a_{22}, a_{33}, a_{44}, a_{21}, a_{23}, a_{32}, a_{34}, a_{43}) &\mapsto (a_{21}, a_{11}, a_{22}, a_{33}, a_{44}, a_{23}a_{32}, a_{34}a_{43}) \end{aligned}$$

Now we give some definitions from Gross et al. (2019) regarding an important subgraph to this work:

**Definition 2.7** For a linear compartmental model  $\mathcal{M} = (G, In, Out, Leak)$ , let  $i \in Out$ . The *output-reachable subgraph to  $i$*  (or *to  $y_i$* ) is the induced subgraph of  $G$  containing all vertices  $j$  for which there is a directed path in  $G$  from  $j$  to  $i$ . A linear compartmental model is *output connectable* if every compartment has a directed path leading from it to an output compartment.

We add the following definition:

**Definition 2.8** A linear compartmental model  $\mathcal{M} = (G, In, Out, Leak)$  is *output connectable to every output* if every compartment has a directed path leading from it to every output compartment.

We will be using the so-called *differential algebra approach* to structural identifiability (Ljung and Glad 1994; Ollivier 1990). Other methods to test identifiability include the Taylor Series approach (Pohjanpalo 1978), the generating series approach (Walter and Lecourtier 1982), the similarity transformation approach (Vajda et al. 1989), and the observability-identifiability condition (Villaverde 2019). In the differential algebra approach, we view the model equations as differential polynomials in a differential polynomial ring  $R(p)[u, y, x]$ , i.e., the ring of polynomials in state variable vector  $x$ , output vector  $y$ , input vector  $u$ , and their derivatives, with coefficients in  $R(p)$  for parameter vector  $p$ . Since the unmeasured state variables  $x_i$  cannot be determined, the goal in this approach is to use differential elimination to eliminate all unknown state variables and their derivatives. The resulting equations are only in terms of input variables, output variables, their derivatives, and parameters, so these equations have the following form:

$$\sum_i c_i(p) \Psi_i(u, y) = 0. \quad (4)$$

An equation of the form (4) is called an *input–output equation* for  $\mathcal{M}$ .

For nonlinear models, one standard “reduced” generating set for these input–output equations is formed by those equations in a *characteristic set* (defined precisely in Glad 1990) that do not involve the  $x_i$ ’s or their derivatives. In a characteristic set, which can be computed using the software DAISY (Bellu et al. 2007), each  $\Psi_i(u, y)$  in each input–output equation (4) is a differential monomial, i.e., a monomial purely in terms of input variables, output variables, and their derivatives. The terms  $c_i(p)$  are called the coefficients of the input–output equations. These coefficients can be fixed uniquely by normalizing the input–output equations to make them monic (Bellu et al. 2007).

However, for linear models, it has been shown that these input–output equations can be found much more easily using the Transfer Function approach (Bellman and Åström 1970) or even a trick with Cramer’s Rule (Meshkat and Sullivant 2014). We will be taking the latter approach to get an explicit formula for the input–output equations.

We now state Theorem 3.8 from Gross et al. (2019) with input  $i$  and output  $j$ , which gives the input–output equation in  $y_j$  in terms of the output-reachable subgraph to  $y_j$ .

**Theorem 2.9** Let  $\mathcal{M} = (G, In, Out, Leak)$  be a linear compartmental model with at least one input. Let  $j \in Out$ , and assume that there exists a directed path from some input compartment to compartment- $j$ . Let  $H$  denote the output-reachable subgraph to  $y_j$ , let  $A_H$  denote the compartmental matrix for the restriction  $\mathcal{M}_H$ , and let  $\partial I$  be the the product of the differential operator  $d/dt$  and the  $|V_G| \times |V_G|$  identity matrix.

Then the following is an input–output equation for  $\mathcal{M}$  involving  $y_j$ :

$$\det(\partial I - A_H)y_j = \sum_{i \in In \cap V_H} (-1)^{i+j} \det(\partial I - A_H)_{ij} u_i, \quad (5)$$

where  $(\partial I - A_H)_{ij}$  denotes the matrix obtained from  $(\partial I - A_H)$  by removing the row corresponding to compartment- $i$  and the column corresponding to compartment- $j$ . Thus, this input–output equation (5) involves only the output-reachable subgraph to  $y_j$ .

**Example 2.10** (Continuation of Example 2.1) The model  $\mathcal{M} = (G, \{1\}, \{2\}, V)$  with  $G$  given by the graph  $\{1 \rightarrow 2, 2 \rightarrow 3, 3 \rightarrow 2, 3 \rightarrow 4, 4 \rightarrow 3\}$  has leaks from every compartment, thus writing the diagonal elements as  $a_{ii}$ , we have the following input–output equation:

$$\begin{aligned} & y_2^{(4)} + (-a_{11} - a_{22} - a_{33} - a_{44})y_2^{(3)} \\ & + (a_{11}a_{22} - a_{23}a_{32} + a_{11}a_{33} + a_{22}a_{33} - a_{34}a_{43} + a_{11}a_{44} + a_{22}a_{44} + a_{33}a_{44})y_2'' \\ & + (a_{11}a_{23}a_{32} - a_{11}a_{22}a_{33} + a_{11}a_{34}a_{43} + a_{22}a_{34}a_{43} - a_{11}a_{22}a_{44} \\ & + a_{23}a_{32}a_{44} - a_{11}a_{33}a_{44} - a_{22}a_{33}a_{44})y_2' \\ & + (-a_{11}a_{22}a_{34}a_{43} - a_{11}a_{23}a_{32}a_{44} + a_{11}a_{22}a_{33}a_{44})y_2 \\ & = (a_{21})u_1'' \\ & + (-a_{21}a_{33} - a_{21}a_{44})u_1' \\ & + (a_{21}a_{33}a_{44} - a_{21}a_{34}a_{43})u_1. \end{aligned}$$

## 2.1 Identifiability

In the differential algebra approach to structural identifiability, one tests identifiability by using the input–output equations determined from the *characteristic set* (Bellu et al. 2007). However, for the case of a linear model, we have shown that the input–output equations can be formed using Eq. (5) and thus we would like to define identifiability using these equations. An important question arises here: can we use the input–output equations in Eq. (5) to test identifiability?

In the case of a single output, there is just a single input–output equation of the form in Eq. (5). However, in the case of multiple outputs, there is an input–output equation for each output in *Out*. The equations formed in Theorem 2.9 are not necessarily minimal, i.e., of lowest degree, but this condition of minimality is required in order for identifiability to be well-defined (Saccomani et al. 2003). In Theorem 3 of (Ovchinnikov et al. 2022), it is shown that the input–output equations in Eq. (5) can be used to analyze identifiability in the case of a linear compartmental model with at least one input and whose graph is strongly connected. Thus, in the rest of this paper, we will be assuming  $G$  is strongly connected and the model has at least one input when we are analyzing models with more than 1 output so that our definition of identifiability is well-defined.



The next step of the differential algebra approach assumes that the coefficients  $c_i(p)$  of the input–output equations can be recovered uniquely from input–output data, and thus are presumed to be known quantities (Soderstrom and Stoica 1989). While there are examples of linear models where this assumption can lead to incorrect conclusions about identifiability (see Example 2.14 of Hong et al. 2020), this assumption that the coefficients can be recovered uniquely from input–output data holds for all linear compartmental models with corresponding graphs such that one can reach a leak or an input from every vertex (see Theorem 2 of Ovchinnikov et al. 2022). We will only be considering either strongly connected or strongly input–output connected models where the output compartment contains a leak, so the coefficients of the input–output equations can be uniquely recovered.

We give some definitions of identifiability from Meshkat et al. (2015).

**Definition 2.11** Let  $(G, In, Out, Leak)$  be a linear compartment model and let  $c$  denote the vector of all nonzero and nonmonic coefficient functions of all the linear input–output equations derived in Theorem 2.9 for each  $i \in Out$ . The function  $c$  defines a map  $c : \mathbb{R}^{|E|+|Leak|} \rightarrow \mathbb{R}^k$ , where  $k$  is the total number of coefficients which we call the *coefficient map*. The linear compartment model  $(G, In, Out, Leak)$  is:

- *globally identifiable* if  $c$  is a one-to-one function, and is *generically globally identifiable* if global identifiability holds everywhere in  $\mathbb{R}^{|E|+|Leak|}$ , except possibly on a set of measure zero.
- *locally identifiable* if around any neighborhood of a point in  $\mathbb{R}^{|E|+|Leak|}$ ,  $c$  is a one-to-one function, and is *generically locally identifiable* if local identifiability holds everywhere in  $\mathbb{R}^{|E|+|Leak|}$ , except possibly on a set of measure zero.
- *unidentifiable* if  $c$  is infinite-to-one.

**Remark** Throughout the rest of this paper, we will be concerned with generic local identifiability. Thus, we will drop the word “generic” and just state a model is locally identifiable. Additionally, if a model is “identifiable,” this means it is (at least) locally identifiable.

**Example 2.12** (Continuation of Example 2.1) Consider the model  $\mathcal{M} = (G, \{1\}, \{2\}, V)$  as described in Example 2.1. The coefficient map for this model is

$$c: \mathbb{R}^9 \rightarrow \mathbb{R}^7$$

$$(a_{11}, a_{22}, a_{33}, a_{44}, a_{21}, a_{23}, a_{32}, a_{34}, a_{43}) \mapsto$$

$$\begin{pmatrix} -a_{11} - a_{22} - a_{33} - a_{44} \\ a_{11}a_{22} - a_{23}a_{32} + a_{11}a_{33} + a_{22}a_{33} - a_{34}a_{43} + a_{11}a_{44} + a_{22}a_{44} + a_{33}a_{44} \\ a_{11}a_{23}a_{32} - a_{11}a_{22}a_{33} + a_{11}a_{34}a_{43} + a_{22}a_{34}a_{43} - a_{11}a_{22}a_{44} + a_{23}a_{32}a_{44} - a_{11}a_{33}a_{44} - a_{22}a_{33}a_{44} \\ -a_{11}a_{22}a_{34}a_{43} - a_{11}a_{23}a_{32}a_{44} + a_{11}a_{22}a_{33}a_{44} \\ a_{21} \\ -a_{21}a_{33} - a_{21}a_{44} \\ a_{21}a_{33}a_{44} - a_{21}a_{34}a_{43} \end{pmatrix}^T$$

Since we will be concerned with generic local identifiability in this work, we will be using the following proposition from Meshkat et al. (2015) which follows from

the fact that the rank of the Jacobian of  $c$  evaluated at a generic point is equal to the dimension of the image of  $c$  (Harris (1992)[Prop. 14.4]).

**Proposition 2.13** (Proposition 2 from Meshkat et al. 2015) *The model  $(G, In, Out, Leak)$  is generically locally identifiable if and only if the rank of the Jacobian of  $c$  is equal to  $|E| + |Leak|$  when evaluated at a random point.*

If a model is unidentifiable, then this means that not all of the parameters can be determined (uniquely or finitely). However, we may still be interested in finding *identifiable functions of parameters* or *identifiable combinations*, and in using these functions to attempt to reparametrize the model.

**Definition 2.14** Let  $c$  be a function  $c : \mathbb{R}^{|E|+|Leak|} \rightarrow \mathbb{R}^k$ . A function  $f : \mathbb{R}^{|E|+|Leak|} \rightarrow \mathbb{R}$  is *globally identifiable* from  $c$  if there exists a function  $\Phi : \mathbb{R}^k \rightarrow \mathbb{R}$  such that  $\Phi \circ c = f$ . The function  $f$  is *locally identifiable* if there is a finitely multi-valued function  $\Phi : \mathbb{R}^k \rightarrow \mathbb{R}$  such that  $\Phi \circ c = f$ .

## 2.2 Strongly Input–Output Connected

In order to consider identifiable path/cycle models, we will be considering graphs  $G$  that have the special property of being connected and every edge is contained in a cycle or path from input to output. We call this *strongly input–output connected*:

**Definition 2.15** We say a graph  $G$  is *strongly input–output connected* if it is connected and every edge is contained in a cycle or path from input to output.

We first show that, in the case of a single output, being strongly input–output connected implies being output connectable, so if we assume the former we get output connectable and can use Theorem 2.9 with the whole matrix  $A$ . Likewise, for the case of multiple outputs, we show that being strongly connected implies being output connectable to every output.

**Proposition 2.16** (1) *Consider a model  $\mathcal{M} = (G, In, \{j\}, Leak)$ . Assume  $G$  is strongly input–output connected. Then  $G$  is output connectable.* (2) *Now consider a model  $\mathcal{M} = (G, In, Out, Leak)$ . Assume  $G$  is strongly connected. Then  $G$  is output connectable to every output.*

**Proof** Let  $\mathcal{M} = (G, In, \{j\}, Leak)$ . Assume  $G$  is strongly input–output connected, i.e., it is connected and every edge is contained in a cycle or path from input to output. Since every edge contained in a path from input to output is connected to the output, we need only consider the edges in cycles. If a vertex in a cycle coincides with a vertex on a path from input to output, then we are done. Thus, assume that there exists a cycle whose vertices do not intersect with the vertices on paths from input to output. Since the graph is connected, the cycle must be attached via a directed edge from either the cycle to a path from input to output or vice versa. But the attaching edge must also be on a path from input to output. Thus for any edge from the path to the cycle, there must be a corresponding edge from the cycle to the path. Thus the graph is output connectable.

If  $\mathcal{M} = (G, In, Out, Leak)$  and  $G$  is strongly connected, then since there is a path from each vertex to every other vertex, then  $G$  is output connectable to every output.  $\square$

**Remark** Note that in Proposition 2.16,  $G$  must be strongly connected as opposed to strongly input–output connected when  $\mathcal{M} = (G, \{i\}, Out, Leak)$ , or else not every vertex may connect to every output.

**Remark** A model that is strongly input–output connected in the case of a single output or strongly connected in the case of multiple outputs is always *structurally observable* (Godfrey and Chapman 1990), as it is output connectable to every output by Proposition 2.16.

We now show that the property of being strongly input–output connected is *almost* strongly connected, in the sense that the graph becomes strongly connected once an edge is added (if not already there) from the output to every input if  $\mathcal{M} = (G, In, \{j\}, Leak)$  or an edge is added from every output to the input if  $\mathcal{M} = (G, \{i\}, Out, Leak)$ .

**Proposition 2.17** (1) Consider a model  $\mathcal{M} = (G, In, \{j\}, Leak)$ . The model  $\mathcal{M}$  is strongly connected if an edge is added from output  $j$  to every input if and only if it is strongly input–output connected. (2) Now consider a model  $\mathcal{M} = (G, \{i\}, Out, Leak)$ . The model  $\mathcal{M}$  is strongly connected if an edge is added from every output to input  $i$  if and only if it is strongly input–output connected. Strongly connected implies strongly input–output connected.

**Proof** A model  $\mathcal{M} = (G, In, \{j\}, Leak)$  is strongly connected if and only if it is connected and every edge is contained in a cycle. Thus a model  $\mathcal{M} = (G, In, \{j\}, Leak)$  is strongly connected if a path from output  $j$  to every input is added if and only if it is connected and every edge is contained in a cycle or path from input to output, i.e., strongly input–output connected. Likewise, a model  $\mathcal{M} = (G, \{i\}, Out, Leak)$  is strongly connected if a path from every output to input  $i$  is added if and only if it is connected and every edge is contained in a cycle or path from input to output, i.e., strongly input–output connected. Additionally, if a model is strongly connected, then it is strongly input–output connected, as every edge is contained in a cycle.  $\square$

We can also examine the *minimum* number of edges in order to be either strongly connected or strongly input–output connected:

**Proposition 2.18** If  $G$  is strongly connected, the minimum number of edges is  $|V|$ . If  $G$  is strongly input–output connected for input  $i$  and output  $j$ , the minimum number of edges is  $|V| - 1$ .

**Proof** For  $G$  to be strongly connected, each vertex must have at least one incoming and one outgoing edge. Thus the minimum number of edges is  $|V|$ . If a graph is strongly input–output connected for input  $i$  and output  $j$ , then this means it becomes strongly connected if an edge is added from output  $j$  to input  $i$  (if not already there). This means the minimum number of edges is  $|V| - 1$ .  $\square$

### 2.3 Expected Number of Coefficients

We first give a result from Meshkat et al. (2015), which we have reworded to agree with the new terminology in this work and have split into two parts: Proposition 2.19 shows that the coefficient map factors through, i.e., can be written purely in terms of, cycles, self-cycles and paths, and Lemma 2.23 gives the degree of the highest-order term on the right-hand side of the input–output equation.

**Proposition 2.19** (Proposition 5 from Meshkat et al. 2015) *Let  $\mathcal{M} = (G, In, \{j\}, V)$  represent a linear compartmental model that is output connectable. The coefficient map  $c$  factors through cycles, self-cycles, and paths from input to output.*

**Proof** Let  $\mathcal{C}(G)$  be the set of all cycles in  $G$ , corresponding to a matrix  $A$ . Recall that the coefficients of the characteristic polynomial of  $A$  can be written as

$$c_i = (-1)^i \sum_{C_1, \dots, C_k \in \mathcal{C}(G)} \prod_{j=1}^k \text{sign}(C_j) a^{C_j},$$

where the sum is over all collections of vertex disjoint cycles involving exactly  $i$  edges of  $G$ , and  $\text{sign}(C) = 1$  if  $C$  is odd length and  $\text{sign}(C) = -1$  if  $C$  is even length. This means for every  $i$ , all cycles of length  $i$  appear as monomial terms in  $c_i$ , and for  $j > i$ , these cycles of length  $i$  appear as monomial products with other cycles in  $c_j$ .

By Theorem 2.9 and the fact that  $G$  is output connectable, meaning that the output-reachable subgraph of  $G$  is all of  $G$ , the input–output equation for  $y_j$  is given by:

$$\det(\partial I - A)y_j = \sum_{i \in In} (-1)^{i+j} \det(\partial I - A)_{ij} u_i. \quad (6)$$

This means the coefficients on the left-hand side factor through the cycles in  $G$ .

Let us now examine these coefficients of the  $u_i$  terms in Eq. (6). For  $i = j$ , the term  $\det(\partial I - A)_{ii}$  gives the coefficients of the characteristic polynomial for the matrix  $A_{ii}$  with row  $i$  and column  $i$  removed, thus these coefficients factor through cycles of the induced subgraph removing vertex  $i$ .

Now assume  $i \neq j$ . The characteristic polynomial of  $A$  can be determined by expanding  $\det(\partial I - A)$  along the  $i^{\text{th}}$  row. Let  $\tilde{A}$  be the matrix  $A$  with the entry  $a_{ij}$  nonzero. Then for  $i \neq j$ , taking the partial derivative of the characteristic polynomial of  $\tilde{A}$  with respect to  $a_{ij}$  precisely gives the polynomial  $\det(\partial I - A)_{ij}$ , up to a minus sign. Since the coefficients of the characteristic polynomial of  $\tilde{A}$  factor through the cycles, then taking the derivative of these coefficients with respect to  $a_{ij}$  has the effect of removing all monomial terms not involving  $a_{ij}$  and setting  $a_{ij}$  to one in the monomial terms that do involve  $a_{ij}$ . This effectively transforms all cycles involving  $a_{ij}$  to paths from the  $i^{\text{th}}$  vertex to the  $j^{\text{th}}$  vertex. Thus, each of the monomial terms are products of paths from the  $i^{\text{th}}$  vertex to the  $j^{\text{th}}$  vertex, cycles, and self-cycles. In other words, coefficients are of the form:

$$c_m = (-1)^m \sum_{P_1, \dots, P_n \in \mathcal{P}(G)} \prod_{l=1}^n \text{sign}(P_l) a^{P_l},$$

where the sum is over all collections of vertex disjoint cycles and paths from  $i$  to  $j$  involving exactly  $m$  edges of  $G$ , and  $\text{sign}(P) = 1$  if  $P$  is odd length and  $\text{sign}(P) = -1$  if  $P$  is even length.

Thus the coefficients can be factored over cycles, self-cycles, and paths from input to output. In other words, there is a polynomial map

$$\psi : \mathbb{R}^{|\mathcal{P}|} \rightarrow \mathbb{R}^k$$

which we will refer to as the path/cycle to coefficient map where  $k$  is the number of coefficients, such that  $c = \psi \circ \pi$  where  $\pi$  is the path/cycle map from Eq. 3.  $\square$

**Definition 2.20** Let the polynomial map  $\psi$  from the path/cycle space  $\mathbb{R}^{|\mathcal{P}|}$  of an output connectable model  $\mathcal{M}$  to its corresponding input–output equation coefficient space  $\mathbb{R}^k$  be defined as the *path/cycle to coefficient map*.

**Example 2.21** (Continuation of Example 2.1) Consider the model  $\mathcal{M} = (G, \{1\}, \{2\}, V)$  as described in Example 2.1. The path/cycle to coefficient map for this model is

$$\psi : \mathbb{R}^7 \rightarrow \mathbb{R}^7$$

$$(p_1, p_2, p_3, p_4, p_5, p_6, p_7) \mapsto \begin{pmatrix} -p_2 - p_3 - p_4 - p_5 \\ p_2 p_3 - p_6 + p_2 p_4 + p_3 p_4 - p_7 + p_2 p_5 + p_3 p_5 + p_4 p_5 \\ p_2 p_6 - p_2 p_3 p_4 + p_2 p_7 + p_3 p_7 - p_2 p_3 p_5 + p_5 p_6 - p_2 p_4 p_5 - p_3 p_4 p_5 \\ -p_2 p_3 p_7 - p_2 p_5 p_6 + p_2 p_3 p_4 p_5 \\ p_1 \\ -p_1 p_4 - p_1 p_5 \\ p_1 p_4 p_5 - p_1 p_7 \end{pmatrix}^T$$

Note that the composition of the path/cycle map  $\pi$  in Example 2.6 and the path/cycle to coefficient map  $\psi$  in Example 2.21 yields the coefficients of the input–output equation for  $\mathcal{M}$  as shown in Example 2.10.

We will be writing the number of coefficients in terms of the minimal distance between an input and output compartment. We define this now:

**Definition 2.22** Let  $i$  be an input compartment and let  $j$  be an output compartment,  $i \neq j$ . Let  $\mathcal{P}(i, j)$  be the set of all paths from vertex  $i$  to vertex  $j$ . Let  $l(P)$  denote the length of a path  $P \in \mathcal{P}$ . Then we can define the minimum length of all paths from vertex  $i$  to vertex  $j$  as  $\text{dist}(i, j) = \min_{P \in \mathcal{P}(i, j)} l(P)$ .

**Lemma 2.23** (Proposition 5 from Meshkat et al. 2015) *Let  $\mathcal{M} = (G, In, \{j\}, V)$  represent a linear compartmental model with  $\mathcal{M}$  output connectable. The highest-order term in  $u_i$  where  $i \in In$  on the right-hand side of the input–output equation, Eq. (6), is of degree  $|V| - 1 - \text{dist}(i, j)$ .*

**Proof** Let  $\tilde{A}$  be the matrix  $A$  with the entry  $a_{ij}$  nonzero. Let  $\mathcal{C}(\tilde{G})$  be the set of all cycles in  $\tilde{G}$ , corresponding to a matrix  $\tilde{A}$ . To determine the coefficient of the highest-order term in  $u_i$ , recall that the coefficients of the characteristic polynomial of  $\tilde{A}$  can be written as

$$c_m = (-1)^m \sum_{C_1, \dots, C_k \in \mathcal{C}(\tilde{G})} \prod_{l=1}^k \text{sign}(C_l) a^{C_l},$$

where the sum is over all collections of vertex disjoint cycles involving exactly  $i$  edges of  $\tilde{G}$ , and  $\text{sign}(C) = 1$  if  $C$  is odd length and  $\text{sign}(C) = -1$  if  $C$  is even length. This means for every  $m$ , all cycles of length  $m$  appear as monomial terms in  $c_m$ , and for  $l > m$ , these cycles of length  $m$  appear as monomial products with other cycles in  $c_l$ .

We now determine the highest-order term in  $u_i$ . Since  $\det(\partial I - A)_{ij}$  is just the partial derivative of the characteristic polynomial of  $\tilde{A}$  with respect to  $a_{ij}$ , up to a minus sign, then the right-hand side of the input–output equation for output  $y_j$  is of the form, where  $n = |V|$ :

$$\sum_{i \in In} (-1)^{i+j} \left( \frac{\partial c_1}{\partial a_{ij}} u_i^{(n-1)} + \frac{\partial c_2}{\partial a_{ij}} u_i^{(n-2)} + \frac{\partial c_3}{\partial a_{ij}} u_i^{(n-3)} + \dots + \frac{\partial c_n}{\partial a_{ij}} u_i \right)$$

We note that not all of these coefficients  $\frac{\partial c_k}{\partial a_{ij}}$  for  $k = 1, \dots, n$  are nonzero and thus we must determine the first nonzero coefficient.

Recall Definition 2.22 for the minimal distance between  $i$  and  $j$ . Let the length of the shortest cycle involving  $a_{ij}$  be of length  $\text{dist}(i, j) + 1$ , so that the length of the shortest path from  $i$  to  $j$  is of length  $\text{dist}(i, j)$ . Then the coefficient of the highest-order term in  $u_i$  is  $\partial c_{\text{dist}(i, j)+1} / \partial a_{ij}$ , which is a sum of the shortest paths (of length  $\text{dist}(i, j)$ ) from  $i$  to  $j$ . Thus it is of the form  $\sum_{P \in \mathcal{P}(i, j): l(P)=\text{dist}(i, j)} a^P$ . This means the highest-order term in  $u_i$  is of degree  $|V| - (\text{dist}(i, j) + 1) = |V| - 1 - \text{dist}(i, j)$ . □

We now give a formula for the number of coefficients of the input–output equation in the case of either single input or single output.

**Theorem 2.24** (Number of nonzero coefficients) *Let  $\mathcal{M} = (G, \{i_1, i_2, \dots, i_{|In|}\}, \{j\}, L)$  represent a linear compartmental model with  $G$  output connectable with at least  $|In \cup Out|$  leaks with  $In \cup Out \subseteq L$ . There are  $|V| + n|V| - \sum_k \text{dist}(i_k, j) + m(|V| - 1)$  nonzero coefficients where  $n = |In - Out|$  and  $m = |In \cap Out|$ . Now let  $\mathcal{M} = (G, \{i\}, \{j_1, j_2, \dots, j_{|Out|}\}, L)$  represent a linear compartmental model with  $G$  output connectable to every output with at least  $|In \cup Out|$  leaks with  $In \cup Out \subseteq L$ . There are  $|V| + n|V| - \sum_k \text{dist}(i, j_k) + m(|V| - 1)$  nonzero coefficients where  $n = |Out - In|$  and  $m = |In \cap Out|$ .*

**Proof** Assume  $\mathcal{M} = (G, \{i_1, i_2, \dots, i_{|In|}\}, \{j\}, L)$  is a linear compartmental model with  $G$  output connectable with at least  $|In \cup Out|$  leaks with  $In \cup Out \subseteq L$ . By Eq. (6), the highest degree term is  $|V|$  on the left-hand side. For the right-hand

side,  $|In \cap Out|$  is either 1 or 0. If  $|In \cap Out| = 1$ , then the highest-order term in  $u_j$  is of degree  $|V| - 1$  on the right-hand side. The highest-order term in  $u_j$  is monic, so there are  $|V| - 1$  coefficients of terms in  $u_j$ . For each  $i \in In - Out$ , the highest degree term in  $u_i$  is of order  $|V| - 1 - \text{dist}(i, j)$  on the right-hand side by Lemma 2.23. In this case, the highest-order term in  $u_i$  is not monic, so there are  $|V| - 1 - \text{dist}(i, j) + 1 = |V| - \text{dist}(i, j)$  coefficients of terms in  $u_i$  when  $i \neq j$ . Altogether, there are  $|V| + n|V| - \sum_k \text{dist}(i_k, j) + m(|V| - 1)$  nonzero coefficients where  $n = |In - Out|$  and  $m = |In \cap Out|$ .

For the case with multiple outputs, let  $\mathcal{M} = (G, \{i\}, \{j_1, j_2, \dots, j_{|Out|}\}, L)$  represent a linear compartmental model with  $G$  output connectable to every output with at least  $|In \cup Out|$  leaks with  $In \cup Out \subseteq L$ . By applying the formula for the case of single output above for each input–output equation, we obtain that there are  $|V| + n|V| - \sum_k \text{dist}(i, j_k) + m(|V| - 1)$  nonzero coefficients where  $n = |Out - In|$  and  $m = |In \cap Out|$ .

We need only show that the coefficients are nonzero (for a generic choice of parameters).

If there are leaks from every compartment, Proposition 2.19 shows that the coefficients factor through cycles, self-cycles, and paths from input to output.

Now consider the case of removing leaks. We will be substituting  $a_{ii}$  as the negative sum of all outgoing edges when  $i \notin L$ , but if  $i \in L$  then  $a_{ii}$  stays the same. Since  $In \cup Out \subseteq L$ , we have that every compartment has an outgoing edge or leak, as every vertex has an outgoing edge except for possibly the output vertex by the output connectable assumption. This means the substitution  $a_{ii}$  as the negative sum of all outgoing edges and leaks retains the  $(i, i)$  entry of  $A$  to be nonzero.

Recall by Proposition 2.19, these coefficients can be factored over cycles, self-cycles, and paths when  $L = V$ . Each coefficient, except for the highest order coefficient in  $u_{i_k}$  when  $i_k \neq j$  (which is a sum of paths from  $i_k$  to  $j$ ), must have a term involving a self-cycle. If the self-cycles in every coefficient are only from leak compartments, we are done. Otherwise, consider a coefficient that has terms involving self-cycles from non-leak compartments which we must substitute into for the case  $L \subset V$ . We want to show that the substitution of the non-leak diagonal terms as the negative sum of all outgoing edges does not cancel every term in that coefficient, so that the coefficients remain nonzero after substitution. We claim that the substitution of  $a_{kk}$  as the negative sum of all outgoing edges for  $k \notin L$  cannot create only terms that are products of cycles and paths from input to output. Since the graph must be output connectable, any cycle formed from the non-leak vertices must connect to the output. In other words, for a chain of vertices in a cycle  $k_1, k_2, \dots, k_l$ , one of these vertices must connect to the output via a path from that vertex to the output. Without loss of generality, assume it is vertex  $k_1$ . Thus the substitution of the non-leaks  $a_{k_1k_1}, a_{k_2k_2}, \dots, a_{k_lk_l}$  cannot create a single monomial term of the form  $\pm a_{k_1k_2} a_{k_2k_3} \cdots a_{k_lk_1}$ , but must also create a monomial  $\pm a_{k_1k_2} a_{k_2k_3} \cdots a_{rk_1}$  where we have substituted  $a_{k_1k_1}$  as  $-a_{k_1k_1} - a_{rk_1}$  for some vertex  $r$  that connects via a path to the output.

This monomial  $a_{k_1k_2} a_{k_2k_3} \cdots a_{rk_1}$  cannot itself be a path from input to output, as the input and output vertices have leaks and thus the corresponding diagonal terms do not get substituted. Thus, it is not a path from input to output, and thus cannot cancel with any other terms in that coefficient. □

**Remark** We note that the assumption of at least  $|In \cup Out|$  leaks with  $In \cup Out \subseteq L$  and  $G$  to be output connectable is to prevent the situation where there are no outgoing edges or leaks from a non-leak vertex and thus upon substitution of the diagonal element  $a_{ii}$  as the negative sum of all outgoing edges and leaks, it becomes zero.

**Definition 2.25** (*Expected number of coefficients*) We say a model  $\mathcal{M} = (G, \{i_1, i_2, \dots, i_{|In|}\}, \{j\}, L)$  with  $In \cup Out \subseteq L$  has the *expected number of coefficients* if there are  $|V| + n|V| - \sum_k \text{dist}(i_k, j) + m(|V| - 1)$  nonzero coefficients in the input–output equations (5) where  $n = |In - Out|$  and  $m = |In \cap Out|$ . We say a model  $\mathcal{M} = (G, \{i\}, \{j_1, j_2, \dots, j_{|Out|}\}, L)$  with  $In \cup Out \subseteq L$  has the *expected number of coefficients* if there are  $|V| + n|V| - \sum_k \text{dist}(i, j_k) + m(|V| - 1)$  nonzero coefficients in the input–output equations (5) where  $n = |Out - In|$  and  $m = |In \cap Out|$ .

### 3 Identifiable Path/Cycle Models

**Definition 3.1** We say a model  $\mathcal{M} = (G, In, Out, V)$  has a coefficient map with *expected dimension* if the dimension of the image of the coefficient map is maximal.

We will be examining a special class of models we call *identifiable path/cycle models*. This class of models is a generalization of *identifiable cycle models* as defined in Meshkat et al. (2015):

**Definition 3.2** (*Identifiable cycle models*) We say a model  $\mathcal{M} = (G, \{i\}, \{i\}, V)$  with  $G$  strongly connected is an *identifiable cycle model* if all of the independent monomial cycles in the model are locally identifiable.

**Example 3.3** The models  $(G, \{1\}, \{1\}, V)$  and  $(H, \{1\}, \{1\}, V)$  where  $G$  corresponds to a chain of exchanges  $1 \leftrightarrow 2, 2 \leftrightarrow 3$ , etc, and  $H$  corresponds to a central compartment given by compartment 1 and exchanges  $1 \leftrightarrow 2, 1 \leftrightarrow 3$ , etc, are identifiable cycle models due to Theorem 5.13 of Meshkat and Sullivant (2014). The first model is commonly called a *catenary* model and the second model is called a *mammillary* model (DiStefano 2015).

It was shown in Meshkat and Sullivant (2014) that a sufficient condition for a model to be an identifiable cycle model is that the dimension of the image of the coefficient map is  $|E| + 1$ . We now define the main object of interest in this paper, *identifiable path/cycle models* and spend the rest of this section forming analogous sufficient conditions on the dimension of the image of the coefficient map.

**Definition 3.4** (*Identifiable path/cycle models*) We say a model  $\mathcal{M} = (G, In, Out, V)$  is an *identifiable path/cycle model* if all of the independent monomial cycles and monomial paths from input to output in the model are locally identifiable and each parameter is contained in such a cycle or path.

**Remark** As identifiable path/cycle models require all of the cycles and paths from input to output in the model to be identifiable, it only makes sense to consider models that are *connected and every edge is contained in a cycle or path from input to output*, i.e. strongly input–output connected. Otherwise, an edge that is not contained in a cycle or path from input to output will not appear in the coefficient map.



**Example 3.5** (Continuation of Example 2.1) The model  $\mathcal{M} = (G, \{1\}, \{2\}, V)$  with  $G$  given by the graph  $\{ 1 \rightarrow 2, 2 \rightarrow 3, 3 \rightarrow 2, 3 \rightarrow 4, 4 \rightarrow 3 \}$  is an identifiable path/cycle model, with identifiable paths and cycles given by  $a_{11}, a_{22}, a_{33}, a_{44}, a_{21}, a_{23}a_{32}, a_{34}a_{43}$ . This can be demonstrated by writing each path and cycle as a function of the coefficients  $c_i$ , e.g., using Groebner Bases. Further justification will come from Theorem 3.13.

For models with leaks from every compartment, the dimension of the image of the coefficient map is bounded above by the number of independent paths and cycles in the graph from Proposition 2.19. We now determine what this number is. We first show that when  $G$  is output connectable for the case of single output, there are  $|E| + |In \cup Out|$  independent directed paths and undirected cycles. We then examine the case where  $G$  is strongly input–output connected so that the indicator vectors for the independent directed paths and directed cycles in the graph correspond to 0/1 vectors. We show that this number of independent paths and cycles is equal to  $|E| + |In \cup Out|$ .

We define the  $|V|$  by  $|E|$  incidence matrix  $E(G)$  as:

$$E(G)_{i,(j,k)} = \begin{cases} 1 & \text{if } i = j \\ -1 & \text{if } i = k \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

In other words,  $E(G)$  has column vectors corresponding to the edges  $j \rightarrow k \in E$  with a 1 in the  $j$ th row,  $-1$  in the  $k$ th row, and 0 otherwise. We define the *indicator vector* of a directed cycle  $C$  as the vector  $(x_s)_{s \in E}$  such that  $x_s = 1$  if  $s \in E_C$  and  $x_s = 0$  if  $s \notin E_C$ , where  $E_C$  is the set of edges associated to the directed cycle  $C$ .

We can also define the *indicator vector* of an undirected cycle  $C'$  with associated directed cycle  $C$  (reversing arrows to all point in the same direction) as the vector  $(x_s)_{s \in E}$  such that  $x_s = 1$  if  $s \in E_C$ ,  $x_s = -1$  if  $-s \in E_C$ , and  $x_s = 0$  if  $s \notin E_C$ , where  $E_C$  is the set of edges associated to the directed cycle  $C$  and  $-s$  corresponds to an edge  $s$  going in the opposite direction. In other words, if  $s$  corresponds to  $i \rightarrow j$ , then  $-s$  corresponds to  $j \rightarrow i$ .

The rank of the directed incidence matrix is well known:

**Proposition 3.6** (Proposition 4.3 of Biggs 1993) *Let  $G$  be a graph with  $|V|$  vertices,  $|E|$  edges, and  $l$  connected components. Then the rank of  $E(G)$  is  $|V| - l$ . Thus, the dimension of the kernel of  $E(G)$  is  $|E| - |V| + l$ .*

We state one final result from Meshkat and Sullivant (2014); Meshkat et al. (2015), which shows that the kernel of  $E(G)$  can be written in terms of  $|E| - |V| + 1$  directed cycles when  $G$  is strongly connected, thus the indicator vectors are 0/1 vectors:

**Proposition 3.7** (Proposition 4 of Meshkat et al. 2015) *Let  $G$  be a strongly connected graph. Then a set of  $|E| - |V| + 1$  linearly independent indicator vectors of directed cycles form a basis for the kernel of  $E(G)$ .*

In other words, this proposition shows that the space of all undirected cycles can be generated by the space of all directed cycles when  $G$  is strongly connected. We now prove a similar result in terms of cycles and paths from input to output when  $G$  is strongly input–output connected.

**Proposition 3.8** *Let  $\mathcal{M} = (G, In, Out, V)$  represent a linear compartmental model with  $G$  strongly input–output connected. Then the space of all directed paths and undirected cycles can be generated by the space of all directed paths and directed cycles and vice versa.*

**Proof** Let  $B$  have as its columns the indicator vectors of all directed paths and undirected cycles. We show that, for every undirected cycle, we can add a positive integer multiple of a directed path vector or directed cycle vector to obtain either a directed cycle or directed path from input to output. Since  $G$  is strongly input–output connected, every edge is in either a cycle or path from input to output. For every edge with a negative entry in the indicator vector of an undirected cycle, that edge either belongs to a cycle or path from input to output. If it belongs to a path from input to output, one can add a positive multiple of the path to the undirected cycle to achieve only non-negative entries corresponding to a closed path or path from input to output. If it does not belong to a path from input to output, then it belongs to a directed cycle. Thus one can add a positive multiple of the directed cycle to the undirected cycle to achieve only non-negative entries corresponding to a closed path or path from input to output. In either case, this corresponds to a multigraph with the property that the indegree of each vertex equals the outdegree of each vertex except possibly at input and output vertices. Cycles can be removed so that the result is a cycle or a path from input to output.  $\square$

**Lemma 3.9** *Let  $\mathcal{M} = (G, In, Out, V)$  represent a linear compartmental model with  $G$  output connectable if  $|Out| = 1$  or  $G$  strongly input–output connected otherwise. Then the number of independent undirected cycles and directed paths from input to output is  $|E| + |In \cup Out|$ .*

**Proof** Let the matrix  $B$  have as columns the indicator vectors of the undirected cycles and directed paths from input to output vertices. Since  $G$  is either output connectable in the single output case or strongly input–output connected, then  $G$  is certainly connected, and thus there are  $|E| - |V| + 1$  undirected cycles that form a basis for the kernel of  $E(G)$  and there must be at least one path from input to output, so  $B$  is certainly not the zero matrix. Form the product  $E(G)B$ . If column  $k$  of  $B$  corresponds to a cycle, then column  $k$  of  $E(G)B$  will be zero, and if column  $k$  of  $B$  corresponds to a path from input in  $i$  and output in  $j$ , then column  $k$  of  $E(G)B$  will have a 1 in row  $i$  and a  $-1$  in row  $j$ .

Remove the zero columns and duplicate columns (which occur when there is more than one path from an input to an output) and zero rows from this matrix  $E(G)B$  and call the resulting matrix  $M$ . We claim  $M$  is the incidence matrix of the graph where there are  $|In \cup Out|$  vertices corresponding to each input/output compartment and there is a directed edge from an input compartment to an output compartment if and only if there is a path from the corresponding input to the corresponding output in the graph  $G$ . Call this graph  $G_M$ . Note that there are only  $|In \cup Out|$  vertices in  $G_M$  because we deleted zero rows from the matrix  $E(G)B$  to obtain the matrix  $M$ , thus deleting vertices that do not correspond to inputs or outputs. This graph  $G_M$  must be connected because we assumed  $G$  is output connectable in the single output case and strongly input–output connected otherwise. Since the rank of the incidence matrix for

a connected graph is the number of vertices minus one, this means the rank of  $E(G)B$  is  $|In \cup Out| - 1$ .

Since the rank of  $E(G)B$  is equal to the rank of  $B$  minus the dimension of the column space of  $B$  intersected with the kernel of  $E(G)$ , which is exactly  $|E| - |V| + 1$  because  $B$  is generated by paths and undirected cycles and a basis for the kernel of  $E(G)$  is given by undirected cycles, then this means the rank of  $B$  is exactly  $|E| - |V| + 1 + |In \cup Out| - 1 = |E| - |V| + |In \cup Out|$ . Adding the  $|V|$  self-cycles, we obtain that the dimension of the path/cycle map is  $|V| + |E| - |V| + |In \cup Out| = |E| + |In \cup Out|$ .  $\square$

**Corollary 3.10** *Let  $\mathcal{M} = (G, In, Out, V)$  represent a linear compartmental model with  $G$  strongly input–output connected. Then the number of independent directed cycles and directed paths from input to output is  $|E| + |In \cup Out|$ .*

**Proof** If  $|Out| = 1$ , strongly input–output connected implies output connectable and if  $|Out| > 1$ , we have strongly input–output connected. The statement follows from Lemma 3.9 and Proposition 3.8 to achieve a set of  $|E| + |In \cup Out|$  independent directed cycles and directed paths from input to output.  $\square$

We now show that the dimension of the image of the coefficient map is bounded above by the number of independent paths and cycles. We will add the important assumption of either  $|In| = 1$  or  $|Out| = 1$  so that the number of distinct input–output pairs equals  $|In \cup Out| - 1$ , described in the Remark below. For  $|Out| = 1$  we can assume  $G$  is strongly input–output connected as stated in Corollary 3.10, but for the case of  $|In| = 1$  we will assume  $G$  is strongly connected in order to ensure the input–output equations are irreducible as shown in Sect. 2. For the special case where  $|In| = |Out| = 1$  and  $In = Out$ , we note that strongly input–output connected reduces to strongly connected. For the special case where  $|In| = |Out| = 1$  and  $In \neq Out$ , then strongly input–output connected is sufficient in what follows, i.e., we can take the weaker of the two conditions strongly input–output connected and strongly connected.

**Lemma 3.11** *Let  $\mathcal{M} = (G, In, Out, V)$  represent a linear compartmental model. Assume that either  $G$  is strongly input–output connected and  $|Out| = 1$  or  $G$  is strongly connected and  $|In| = 1$ . The dimension of the image of the coefficient map is bounded above by  $|E| + |In \cup Out|$ .*

**Proof** The coefficient map factors through the cycles, self-cycles, and paths from input to output from Proposition 2.19. By Corollary 3.10, the number of independent paths and cycles is  $|E| + |In \cup Out|$ . Thus the dimension of the image of the coefficient map is bounded above by  $|E| + |In \cup Out|$ .  $\square$

**Remark** We require  $|In| = 1$  or  $|Out| = 1$  so that there are either  $|Out| - |In \cap Out|$  or  $|In| - |In \cap Out|$  distinct input–output pairs, respectively, which equals  $|In \cup Out| - 1$ , the rank of  $E(G)B$  in the proof of Lemma 3.9. Example 3.12 demonstrates this.

**Example 3.12** The model  $\mathcal{M} = (G, \{1, 2\}, \{3, 4\}, V)$  shown in Fig. 2 where  $G$  is given by  $\{ 1 \rightarrow 2, 2 \rightarrow 1, 2 \rightarrow 3, 3 \rightarrow 2, 2 \rightarrow 4, 4 \rightarrow 2 \}$  has  $|E| + |In \cup Out| = 10$

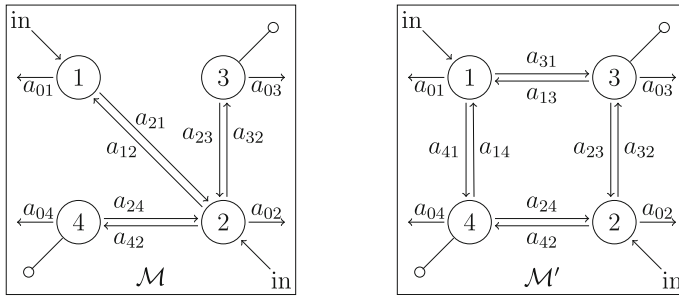


Fig. 2 The models  $\mathcal{M}$  and  $\mathcal{M}'$  from Example 3.12

independent paths and cycles but the coefficient map factors over 11 paths and cycles given by  $a_{12}a_{21}, a_{23}a_{32}, a_{24}a_{42}, a_{32}, a_{42}, a_{21}a_{32}, a_{21}a_{42}, a_{11}, a_{22}, a_{33}, a_{44}$ . The problem here is that there are only 10 parameters, but we are attempting to factor over 11 paths and cycles. This is why we require  $|In| = 1$  or  $|Out| = 1$ . However, this is a sufficient condition but not a necessary condition, as the model  $\mathcal{M}' = (G', \{1, 2\}, \{3, 4\}, V)$  also shown in Fig. 2 where  $G'$  is given by  $\{ 1 \rightarrow 3, 3 \rightarrow 1, 1 \rightarrow 4, 4 \rightarrow 1, 2 \rightarrow 3, 3 \rightarrow 2, 2 \rightarrow 4, 4 \rightarrow 2 \}$  has  $|E| + |In \cup Out| = 12$  independent paths and cycles given by  $a_{31}, a_{41}, a_{32}, a_{42}, a_{13}a_{31}, a_{14}a_{41}, a_{23}a_{32}, a_{24}a_{42}, a_{11}, a_{22}, a_{33}, a_{44}$  and the coefficient map factors over these as well.

**Remark** Notice that in Lemma 3.11, we assume that  $G$  is either strongly input–output connected or strongly connected. We have just shown that the expected dimension in this case is  $|E| + |In \cup Out|$ . But a natural question that arises is, what if we do not assume this connectedness condition on  $G$ ? Clearly, we still have that the coefficient map factors through cycles and paths. In this case there will be *at most*  $|E| + |In \cup Out|$  independent cycles and paths that appear in the coefficient map, i.e., the coefficient map may factor over fewer than  $|E| + |In \cup Out|$  independent paths and cycles. See Sect. 5.

This gives us the following theorem:

**Theorem 3.13** *Let  $\mathcal{M} = (G, In, Out, V)$  represent a linear compartmental model with either  $G$  strongly input–output connected and  $|Out| = 1$  or  $G$  strongly connected and  $|In| = 1$ . If the image of the coefficient map has dimension  $|E| + |In \cup Out|$ , then the model is an identifiable path/cycle model.*

**Proof** By Lemma 3.11, the dimension of the image of the coefficient map is bounded above by  $|E| + |In \cup Out|$ , which is also the number of independent paths and cycles. Recall a function  $f$  is locally identifiable if there is a finitely multivalued function  $\phi : \mathbb{R}^{|E|+|In \cup Out|} \rightarrow \mathbb{R}$  such that  $\phi \circ c = f$ . Let  $\pi : \mathbb{R}^{|E|+|V|} \rightarrow \mathbb{R}^{|E|+|In \cup Out|}$  be the path/cycle map from Eq. 3. Since  $c : \mathbb{R}^{|E|+|V|} \rightarrow \mathbb{R}^{|E|+|In \cup Out|}$  factors over paths and cycles, then there exists a function  $\psi : \mathbb{R}^{|E|+|In \cup Out|} \rightarrow \mathbb{R}^{|E|+|In \cup Out|}$  as defined in Definition 2.20 such that  $c = \psi \circ \pi$ . If the dimension of the image of the coefficient map is precisely  $|E| + |In \cup Out|$ , then this function  $\psi$  is locally invertible with  $\psi^{-1} = \phi$  and thus  $\pi = \phi \circ c$ . Thus the paths and cycles are identifiable.  $\square$

**Example 3.14** (Continuation of Example 2.1) The model  $\mathcal{M} = (G, \{1\}, \{2\}, V)$  from Example 2.1 can be shown to have dimension of the image of the coefficient map equal to  $|E| + |In \cup Out| = 5 + 2 = 7$ . Thus there are 7 identifiable paths and cycles given by the monomials  $a_{11}, a_{22}, a_{33}, a_{44}, a_{21}, a_{23}a_{32}, a_{34}a_{43}$ .

### 3.1 Necessary Condition for Number of Edges and the Edge Inequality

**Proposition 3.15** Let  $\mathcal{M} = (G, In, Out, V)$  represent a linear compartmental model with either  $G$  strongly input–output connected and  $|Out| = 1$  or  $G$  strongly connected and  $|In| = 1$ . If  $\mathcal{M} = (G, In, Out, V)$  is an identifiable path/cycle model, then  $|E| + |In \cup Out| \leq$  the expected number of coefficients.

**Proof** We have that the dimension of the image of the coefficient map is bounded above by the expected number of coefficients, as it cannot exceed the number of coefficients. Thus  $|E| + |In \cup Out| \leq$  the expected number of coefficients.  $\square$

**Definition 3.16** (*Edge inequality*) We say that a model has a number of edges given by the *edge inequality* if the number of edges  $|E|$  satisfies  $|E| + |In \cup Out| \leq$  the expected number of coefficients.

We can also show that the property of being strongly input–output connected is a necessary condition for having expected dimension in the case of maximal number of edges.

**Proposition 3.17** Let  $\mathcal{M} = (G, In, Out, V)$  with  $G$  output connectable and  $|Out| = 1$  represent a linear compartmental model for which the edge inequality is an equality with expected dimension  $|E| + |In \cup Out|$ . Then the graph must be strongly input–output connected.

**Proof** Assume the coefficient map has expected dimension  $|E| + |In \cup Out|$  with the maximal number of edges. This means the expected dimension is the number of coefficients. If the graph is not connected with every edge in a cycle or path from input to output, then there are parameters that do not appear in the coefficient map, and thus the coefficient map factors over fewer than  $|E| + |In \cup Out|$  independent paths and cycles. But this contradicts having expected dimension  $|E| + |In \cup Out|$ .  $\square$

**Remark** If there are fewer edges, we can still achieve expected dimension without this condition of strongly input–output connected. In other words, being strongly input–output connected is a sufficient but not necessary condition to achieve  $|E| + |In \cup Out|$  independent cycles and paths in the coefficient map. See Sect. 5.

### 3.2 Obtaining Identifiability by Removing Leaks

In this section, we show that removing all leaks except leaks from input/output compartments results in identifiability, much like the results in Meshkat et al. (2015). We will follow the same proof. Recall that for an identifiable path/cycle model, we have the coefficient map  $c : \mathbb{R}^{|V|+|E|} \rightarrow \mathbb{R}^k$ . Let  $\pi : \mathbb{R}^{|V|+|E|} \rightarrow \mathbb{R}^{|E|+|In \cup Out|}$  be the

path/cycle map from Eq. 3, that is  $\pi(A(G)) = (a^P : P \text{ is a cycle or path from input to output of } G)$ . Then Proposition 2.19 tells us that  $c$  factors through  $\pi$  without loss of dimension. Thus  $c = \psi \circ \pi$  where  $\psi$  is defined as in Definition 2.20 and the dimension of the image of  $c$  equals the dimension of the image of  $\pi$ .

Passing from a model  $(G, In, Out, V)$  to a model  $(G, In, Out, Leak)$  such that  $|Leak| = |In \cup Out|$  amounts to restricting the parameter space  $\mathbb{R}^{|V|+|E|}$  to a linear subspace  $\Lambda \subseteq \mathbb{R}^{|V|+|E|}$  of dimension  $|E| + |Leak|$  and we would like the image of  $\Lambda$  under the coefficient map  $c$  to have dimension  $|E| + |Leak|$ . Since  $c$  factors through the path/cycle map  $\pi$  it suffices to prove that the image of  $\Lambda$  under  $\pi$  has dimension  $|E| + |Leak|$ .

**Lemma 3.18** *Let  $G = (V, E)$  be a directed graph with corresponding identifiable path/cycle model  $(G, In, Out, V)$ . Assume that either  $G$  is strongly input–output connected and  $|Out| = 1$  or  $G$  is strongly connected and  $|In| = 1$ . Consider a model  $(G, In, Out, L)$  where  $In \cup Out \subseteq L$ . Let  $\pi : \mathbb{R}^{|V|+|E|} \rightarrow \mathbb{R}^{|E|+|In \cup Out|}$  denote the path/cycle map. Let  $\Lambda \subseteq \mathbb{R}^{|V|+|E|}$  be the linear space satisfying*

$$\Lambda = \{A \in \mathbb{R}^{|V|+|E|} : a_{ii} = - \sum_{j, j \neq i} a_{ji} \text{ for all } i \notin L\}.$$

*Then the dimension of the image of  $\Lambda$  under the map  $\pi$  is  $|E| + |In \cup Out|$ .*

**Proof** Since  $\Lambda$  is a linear space, we just consider the natural map from  $\mathbb{R}^{|E|+|L|} \rightarrow \mathbb{R}^{|E|+|In \cup Out|}$  which maps to the path/cycle space. To show that the dimension of the image of this map is correct, we consider the Jacobian of this map and show that it has full rank.

Note that the rows corresponding to the  $|L|$  self-cycles are linearly independent so we focus on the  $|E| + |In \cup Out| - |L|$  by  $|E|$  submatrix ignoring those rows and columns, which we will call  $J$ . Arrange the matrix so that the first  $|E| - |V| + |In \cup Out|$  rows correspond to the paths and cycles of  $G$  and the last  $|V| - |L|$  rows correspond to the non-leak diagonal elements. Let the first  $|E| - |V| + |In \cup Out|$  rows be called  $A$  and the last  $|V| - |L|$  rows be called  $B$ . Clearly the rows of  $A$  are linearly independent by Lemma 3.11. The rows of  $B$  are linearly independent since they are in triangular form since each involves distinct parameters. To show that the full set of  $|E| + |In \cup Out| - |L|$  rows are linearly independent, we need only show that the row space of  $A$  and the row space of  $B$  intersect only in the origin.

To prove that  $J$  generically has maximal possible rank, it is enough to show that there is some point where the evaluation of  $J$  at said point yields the maximal rank. We choose the point where we set all the edge parameters  $a_{ij} = 1$  for all  $j \rightarrow i \in E$ . This specialization yields that the row space of  $A$  is exactly the path/cycle space of the graph  $G$ , i.e., all of the weightings on the edges of the graph where the indegree equals the outdegree of every vertex in a cycle and every vertex except the first and last in a path from input to output. Also, we have that the matrix  $B$  which has dimension  $(|V| - |L|) \times (|E|)$ , which consists of the rows corresponding to the vertices in  $V \setminus L$ , and the (negated) row corresponding to vertex  $i$  has a one for an edge  $i' \rightarrow j'$  if and only if  $i = i'$ , with all other entries zero.

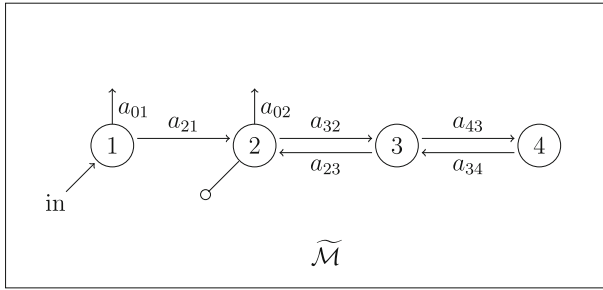


Fig. 3 Graph for model corresponding to  $\tilde{\mathcal{M}}$  in Example 3.20

Since  $A$  spans the path/cycle space of  $G$ , each element in the row space of  $A$  corresponds to a weighting on the edges of  $G$  where the total weight of all incoming edges at a vertex  $i$  equals the total weight of all outgoing edges at vertex  $i$  except at input or output vertices  $In \cup Out$ . On the other hand, we claim that the only vector in the row span of  $B$  with the same property is the zero vector. To show this, let  $b_i$  be the row vector associated to some vertex  $i$ . Note that a vector in the row span of  $B$  will have zero weight on any of the outgoing edges of vertices in  $In \cup Out$ .

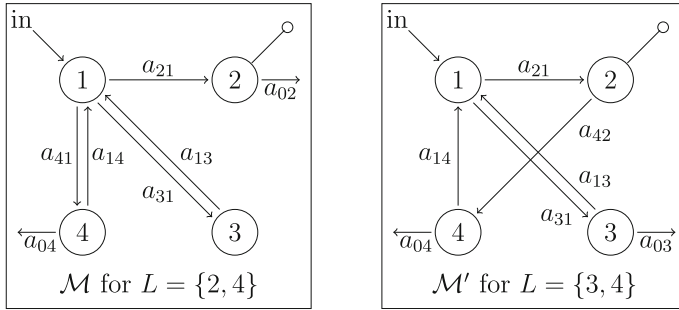
In order for the indegree to equal the outdegree, we would need to include a  $b_j$  with an edge pointing toward vertex  $i$ . Continuing in this way, we can only stop when we have included an input or output vertex since the indegree need not equal the outdegree for those vertices. However, this contradicts the fact that a vector in the row span of  $B$  will have zero weight on any of the outgoing edges of vertices in  $In \cup Out$ .  $\square$

**Theorem 3.19** (Removing leaks) *Let  $\mathcal{M} = (G, In, Out, V)$  represent a linear compartmental model. Assume that either  $G$  is strongly input–output connected and  $|Out| = 1$  or  $G$  is strongly connected and  $|In| = 1$ . Assume it is an identifiable path/cycle model. Then, the corresponding model  $\tilde{\mathcal{M}} = (G, In, Out, L)$  where  $In \cup Out \subseteq L$  for any such  $L$  has expected dimension. In particular, if  $L = In \cup Out$ , then  $\tilde{\mathcal{M}}$  is locally identifiable.*

**Proof** By Lemma 3.18 and the comments preceding it we know that the image of the restricted parameter space under the path/cycle map  $\pi$  has dimension  $|E| + |In \cup Out|$ , which is equal to the dimension of the image of the full parameter space under the path/cycle map. Since, for an identifiable path/cycle model, the dimension of the image of the coefficient map  $c$  is  $|E| + |In \cup Out|$ , this must be the same for the restricted model. In particular, if  $|L| = |In \cup Out|$ , then the model has  $|E| + |In \cup Out|$  parameters, hence it is locally identifiable.  $\square$

**Example 3.20** (Continuation of Example 2.1) The model  $\tilde{\mathcal{M}} = (G, \{1\}, \{2\}, \{1, 2\})$  obtained from the model in Example 2.1 by removing two leaks and leaving the leaks in the input and output compartments is locally identifiable (see Fig. 3).

**Remark** We note that, while  $L = In \cup Out$  is sufficient in Theorem 3.19, it is certainly not necessary, as there are other possible configurations of  $|L| = |In \cup Out|$  leaks that also result in identifiability. The next example demonstrates this.



**Fig. 4** On the left is the graph corresponding to  $\mathcal{M}$  with leak set  $L = \{2, 4\}$ , and on the right is the graph corresponding to  $\mathcal{M}'$  with leak set  $L = \{3, 4\}$  from Example 3.21

**Example 3.21** Consider the model  $\mathcal{M} = (G, \{1\}, \{2\}, L)$  shown in Fig. 4 where  $|L| = |In \cup Out| = 2$  and  $G$  is given by the edges  $\{1 \rightarrow 2, 1 \rightarrow 3, 3 \rightarrow 1, 1 \rightarrow 4, 4 \rightarrow 1\}$ . The identifiable models are the ones where  $L = \{2, 4\}, \{2, 3\}, \{1, 2\}$  and the unidentifiable models have  $L = \{3, 4\}, \{1, 4\}, \{1, 3\}$ .

We note that while  $2 \in L$  appears to be sufficient for identifiability in this model, we can consider another model given by  $\mathcal{M}' = (G', \{1\}, \{2\}, L)$  also shown in Fig. 4 where  $|L| = |In \cup Out| = 2$  and  $G'$  is given by the edges  $\{1 \rightarrow 2, 3 \rightarrow 1, 4 \rightarrow 1, 1 \rightarrow 3, 2 \rightarrow 4\}$ . The identifiable models are the ones where  $L = \{3, 4\}, \{2, 3\}, \{1, 4\}, \{1, 2\}$  and the unidentifiable models have  $L = \{2, 4\}, \{1, 3\}$ . This shows the pattern of identifiability depends on the graph structure itself and not just the placement of inputs and outputs.

### 3.3 Sufficient Condition for Identifiable Path/Cycle Model

We now give a sufficient condition for a model to be an identifiable path/cycle model with 1 input and 1 output. This sufficient condition is analogous to the sufficient condition from Meshkat and Sullivant (2014) of inductively strongly connected for models with input and output in the same compartment. In fact, Theorem 3.22 reduces to Theorem 5.13 of Meshkat and Sullivant (2014) if the input and output compartments are the same.

**Theorem 3.22** *Let  $\mathcal{M} = (G, \{i\}, \{j\}, V)$  represent a linear compartmental model with  $G$  strongly input–output connected and  $|E| = 2|V| - (\text{dist}(i, j) + 2)$ . If  $\mathcal{M} = (G, \{i\}, \{j\}, V)$  has no path from compartment  $j$  to compartment  $i$  but becomes inductively strongly connected if an edge from compartment  $j$  to compartment  $i$  is added, then  $\mathcal{M}$  is an identifiable path/cycle model.*

Before we prove Theorem 3.22, we define a graph structure which will be useful in the proof.

**Definition 3.23** (Definition 5.6 from Meshkat and Sullivant 2014) A chain of cycles is a graph  $H$  which consists of a sequence of directed cycles that are attached to each other in a chain by joining at the vertices.



**Theorem 3.24** *Suppose  $\mathcal{M}' = (G', \{i\}, \{j\}, V)$  represents a linear compartmental model with  $G'$  strongly input–output connected and  $|E| = 2|V| - (\text{dist}(i, j) + 2)$ . Suppose too that  $\mathcal{M}'$  has expected dimension and also that  $\mathcal{M}'$  has no path from  $j$  to  $i$  and becomes inductively strongly connected if the edge from  $j$  to  $i$  is added. Then, if  $G$  is a new graph obtained from  $G'$  by adding a vertex  $n$  and two edges  $k \rightarrow n$  and  $n \rightarrow l$  such that  $G$  has a chain of cycles containing either  $i$  and  $n$  or  $j$  and  $n$ , then the model  $\mathcal{M} = (G, \{i\}, \{j\}, V \cup \{n\})$  has expected dimension.*

Recall that we can induce a *weight order* on a polynomial ring  $\mathbb{K}[x_1, \dots, x_n]$  for some weight vector  $\omega \in \mathbb{Q}^n$  where the weight of a monomial  $x_1^{\alpha_1} \cdots x_n^{\alpha_n}$  is  $\omega \cdot \alpha$  where  $\alpha = (\alpha_1, \dots, \alpha_n)$ . We can then define the *initial forms* of a polynomial  $f$  as  $\text{in}_\omega(f)$  to be the sum of all terms of  $f$  whose monomial has the highest weight with respect to said  $\omega$ .

Now, if we define the coefficient map associated to the graph  $G$  as  $\phi_G : \mathbb{R}^{|V|+|E|} \rightarrow \mathbb{R}^k$ , then we can also consider the pull-back of said map defined as  $\phi_G^* : \mathbb{K}[c, d] \rightarrow \mathbb{K}[a]$  where  $c, d$  correspond to the coefficients of the left and right-hand side of the input–output equation respectively, and  $a$  corresponds to the parameters found in compartmental matrix  $A$ . Now define  $\phi_{G,\omega}$  to be the initial parameterization defined as the parameterization with pullback  $\phi_{G,\omega}^*$  where  $\phi_{G,\omega}^*(f) = \text{in}_\omega(f)$  for a given weight  $\omega$ .

**Lemma 3.25** (Corollary 5.9, Meshkat and Sullivant 2014) *Let  $\phi^* : \mathbb{K}[x] \rightarrow \mathbb{K}[y]$  be a  $\mathbb{K}$ -algebra homomorphism and  $\omega \in \mathbb{Q}^m$  a weight vector, then*

$$\dim(\text{im}\phi_\omega) \leq \dim(\text{im}\phi).$$

We will use Lemma 3.25 in the following way. We want to compute the dimension of the image of a polynomial parametrization  $\phi$ . We know for other reasons an upper bound  $d$  on this dimension. We have a weight vector  $\omega$  where we can compute the dimension of the image of the polynomial parametrization  $\phi_\omega$ , and we show it is equal to  $d$ . Then, by Lemma 3.25, we know that the dimension of the image of  $\phi$  must be  $d$ .

**Proof of Theorem 3.24** Suppose  $\mathcal{M}' = (G', \{i\}, \{j\}, V)$  is a linear compartmental model with expected dimension such that  $G'$  is strongly input–output connected,  $|E| = 2|V| - (\text{dist}(i, j) + 2)$ . Also suppose that if we add the edge from  $j$  to  $i$ , the new graph becomes inductively strongly connected. Note that by Theorem 2.24, the input–output equation of the model  $\mathcal{M}'$  has  $2|V| - \text{dist}(i, j)$  nonzero, non-monic coefficients. Let  $|V| = n - 1$  and  $m = |E|$ .

Define  $\phi_G : \mathbb{R}^{|V|+|E|} \rightarrow \mathbb{R}^{2|V|-\text{dist}(i,j)}$ , to be the coefficient map associated to a graph  $G$  as above with corresponding pull-back  $\phi_G^*$ . Choose weight  $\omega$  as follows:

$$\omega_{uv} = \begin{cases} 0 & \text{if } (u, v) = (n, n) \\ \frac{1}{2} & \text{if } (u, v) = (n, k) \text{ or } (l, n) \\ 1 & \text{otherwise.} \end{cases}$$

Recall that for each coefficient, the corresponding polynomial function is homogeneous in terms of the parameters. Also, recall that the left-hand side coefficients are

generated by cycles of the corresponding graph, while the right-hand side coefficients are generated by products of cycles of the corresponding graph along with paths from the input to output.

Applying this weight to the polynomial coefficients has the effect of removing any monomial containing a cycle which is incident to compartment  $n$ , in all coefficients except for the lowest order terms in both  $c$  and  $d$ . Note that in the case of  $c_n$  and  $d_{n-1}$ , each of the monomials in the sum will have a cycle incident to  $n$ , meaning that each of them has the same weight.

More explicitly, in terms of the pull-back maps we have in all cases that

$$\begin{aligned} \phi_{G,\omega}^*(c_i) &= \phi_{G'}^*(c_i) \quad i = 1, \dots, n - 1 \\ \phi_{G,\omega}^*(d_i) &= \phi_{G'}^*(d_i) \quad i = \text{dist}(i, j), \dots, n - 2 \\ \phi_{G,\omega}^*(c_n) &= \phi_G^*(c_n) \\ \phi_{G,\omega}^*(d_{n-1}) &= \phi_G^*(d_{n-1}). \end{aligned}$$

Thus,  $\phi_{G,\omega}$  agrees with  $\phi_{G'}$  everywhere except for the highest order coefficients on either side of the input–output equation, in which case  $\phi_{G,\omega}$  matches  $\phi_G$ . This implies that the Jacobian matrix corresponding to  $\phi_{G,\omega}$  defined as  $J(\phi_{G,\omega})$  (whose generic rank yields the dimension of the image of the map), has the form

$$J(\phi_{G,\omega}) = \begin{pmatrix} J(\phi_{G'}) & 0 \\ * & C \end{pmatrix}$$

where  $J(\phi_{G'})$  is the  $(2n - (\text{dist}(i, j) + 2)) \times (n + m - 3)$  Jacobian matrix of  $\phi_{G'}$  and  $C$  is the  $2 \times 3$  matrix

$$C = \begin{pmatrix} \frac{\partial c_n}{\partial a_{nn}} & \frac{\partial c_n}{\partial a_{ln}} & \frac{\partial c_n}{\partial a_{nk}} \\ \frac{\partial d_{n-1}}{\partial a_{nn}} & \frac{\partial d_{n-1}}{\partial a_{ln}} & \frac{\partial d_{n-1}}{\partial a_{nk}} \end{pmatrix}$$

where  $l$  and  $k$  are the nodes to which the added node  $n$  has an edge to and from respectively.

Note that we assume that the model corresponding to  $G'$  has expected dimension, hence  $J(\phi_{G'})$  has rank  $2(n - 1) - \text{dist}(i, j)$ . Since  $J(\phi_G)$  is lower block triangular, to show that it has rank  $2n - \text{dist}(i, j)$ , we need only show that  $C$  has generic rank 2.

Let  $H$  be a chain of cycles in  $G$  defined as  $s_2, \dots, s_t$  in order such that  $s_2$  is a cycle containing either the input or the output and  $s_t$  is the cycle containing the node  $n$ . Also, define  $s_1$  to be one of the shortest paths from  $i$  to  $j$ .

Now we will choose entries for the matrix  $A$  such that the matrix  $C$  has rank 2. First, let all diagonal elements of  $A$  be 1, i.e.,  $a_{kk} = 1$  for all  $k = 1, \dots, n$ . Also, let  $a_{uv} = 0$  for all edges  $v \rightarrow u \notin H$ . For all edges in  $H$ , for each cycle  $s_i$ , choose the edge weights so that the product of edges' weights is equal to  $(-1)^{\ell(s_i)-1}$ , that is so that the product of the edges in the cycle is equivalent to the sign of the cycle. For  $s_1$ , choose edge weights so that the product of the edges' weights is also equal to  $(-1)^{\ell(s_1)-1}$ .

First, consider the entry  $\frac{\partial c_n}{\partial a_{nn}}$ . The only nonzero monomials appearing here will arise from taking products of the cycles  $s_2, \dots, s_{t-1}$ , since the cycle  $s_t$  cannot be involved, as we only consider the elements of the sum of  $c_n$  with  $a_{nn}$  as a factor. Also,  $s_1$  is not a cycle, hence cannot be part of any of the  $c_i$ . Since each cycle touches its two neighboring cycles, and no other cycles, and in the expansion we expand over all products of nontouching cycles that cover all  $n$  vertices, we get that the number of monomials will equal the number of subsets of  $\{2, \dots, t - 1\}$  with no adjacent elements. By Lemma 3.26, this is exactly  $F_t$ .

Now consider the entry  $\frac{\partial d_{n-1}}{\partial a_{nn}}$ , which will arise from taking products of  $s_1, s_3, \dots, s_{t-1}$ , since we must have  $s_1$ , hence cannot have  $s_2$ , and by similar reasoning above cannot have  $s_t$ . Thus, we get that the number of monomials are the number of nonadjacent subsets of  $\{s_3, \dots, s_{t-1}\}$ , hence we get the Fibonacci number  $F_{t-1}$ .

In the case of the entry  $\frac{\partial c_n}{\partial a_{ln}}$  or equivalently  $\frac{\partial c_n}{\partial a_{nk}}$ , we must use the cycle  $s_t$  prohibiting us from using  $s_{t-1}$ , and again must not use  $s_1$ . This yields that the number of monomials is the number of nonadjacent subsets of  $s_2, \dots, s_{t-2}$ , i.e., the Fibonacci number  $F_{t-1}$ .

Finally, when considering entry  $\frac{\partial d_{n-1}}{\partial a_{ln}}$  or equivalently  $\frac{\partial d_{n-1}}{\partial a_{nk}}$ , we must use  $s_1$  and cycles  $s_t$ , hence cannot use cycles  $s_2$  or  $s_{t-1}$ . This means that the number of monomials will be the number of nonadjacent subsets of  $s_3, \dots, s_{t-2}$ , i.e., the Fibonacci number  $F_{t-2}$ .

Thus, the submatrix  $C$  will have the form

$$C = \begin{pmatrix} F_t & F_{t-1} & F_{t-1} \\ F_{t-1} & F_{t-2} & F_{t-2} \end{pmatrix}.$$

The classical identity of Fibonacci number  $F_t F_{t-2} - F_{t-1}^2 = (-1)^{t-1}$  yields that this matrix has full rank. hence, the Jacobian of  $\phi_{G,\omega}^*$  has full rank.

Note that the upper bound for the number of coefficients of the input–output equation, i.e., the upper bound on the dimension of the image of the coefficient map is  $2n - \text{dist}(i, j)$  via Theorem 2.24. Thus, because  $\dim(\text{im}(\phi_{G,\omega}^*)) \leq \dim(\text{im}(\phi_G^*))$  by Lemma 3.25, and  $\dim(\text{im}(\phi_G^*))$  is bounded above by  $2n - \text{dist}(i, j)$ , we have that  $\dim(\text{im}(\phi_G^*)) = 2n - \text{dist}(i, j)$  as desired.  $\square$

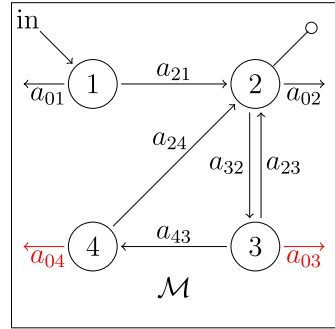
**Lemma 3.26** *The number of subsets  $S$  of  $\{1, 2, \dots, n\}$  such that  $S$  contains no pair of adjacent numbers is the  $n + 2$ -nd Fibonacci number,  $F_{n+2}$  which satisfies the recurrence  $F_0 = 0, F_1 = 1$ , and  $F_{n+1} = F_n + F_{n-1}$ .*

We can now prove Theorem 3.22.

**Proof of Theorem 3.22** By Theorem 3.24 and the inductive nature of inductively strongly connected graphs, it suffices to show that every inductively strongly connected graph beginning with a cycle between  $i$  and  $j$  has a chain of cycles containing the vertices  $i$  and  $n$  (or, analogously, a chain of cycles containing the vertices  $j$  and  $n$ ).

We prove this by induction on  $n$ . Since  $G$  is inductively strongly connected if the edge from  $j$  to  $i$  is added, there is a nontrivial cycle  $c$  that passes through the vertex  $n$ . If  $c$  contains  $i$ , we are done. Otherwise, let  $q$  be the smallest vertex appearing in  $c$ , and let  $G'$  be the induced subgraph on  $\{i, j, \dots, q\}$ . By induction,  $G'$  has a chain of

**Fig. 5** (Color Figure Online) The model  $\mathcal{M}$  corresponds to the above graph with all four leaks, while the graph  $\mathcal{M}'$  has the same graph with only the black leaks, that is leaks in compartments 1 and 2, all from Example 3.28



cycles containing  $i$  and  $q$ . Attaching  $c$  to  $H$  gives a chain of cycles in  $G$  containing  $i$  and  $n$ . A similar argument can be applied to give a chain of cycles in  $G$  containing  $j$  and  $n$ . □

Theorem 3.22 is not only useful as a sufficient condition for an identifiable path/cycle model, but it is also useful as a means to start with an identifiable path/cycle model and then remove leaks to obtain identifiability:

**Corollary 3.27** *Let  $\mathcal{M} = (G, \{i\}, \{j\}, V)$  represent a linear compartmental model with  $G$  strongly input–output connected and  $|E| = 2|V| - (\text{dist}(i, j) + 2)$ . If  $\mathcal{M} = (G, \{i\}, \{j\}, V)$  has no path from compartment  $j$  to compartment  $i$  but becomes inductively strongly connected if the edge from compartment  $j$  to compartment  $i$  is added, then  $\mathcal{M}' = (G, \{i\}, \{j\}, \{i, j\})$  is locally identifiable, i.e. removing all but two leaks in the input/output compartments.*

**Proof** This follows from Theorem 3.22 and Theorem 3.19. □

**Example 3.28** The model  $\mathcal{M} = (G, \{1\}, \{2\}, V)$  shown in Fig. 5 with  $G$  given by the edges  $\{1 \rightarrow 2, 2 \rightarrow 3, 3 \rightarrow 4, 4 \rightarrow 2, 3 \rightarrow 2\}$  is an identifiable path/cycle model by Theorem 3.22. Thus, the model  $\mathcal{M}' = (G, \{1\}, \{2\}, \{1, 2\})$  where we remove all but two leaks from the input/output compartments is locally identifiable.

We also note that, as cycles with input/output in the same compartment were shown to have expected dimension in Meshkat and Sullivant (2014) (see Proposition 5.4), paths from input to output can be shown to have expected dimension as well.

**Proposition 3.29** *Let  $\mathcal{M} = (G, \{1\}, \{|V|\}, V)$  be a linear compartmental model with  $G$  given by a path from input 1 to output  $|V|$  with  $|V| - 1$  edges. Then  $\mathcal{M}$  is an identifiable path/cycle model and the model  $\widetilde{\mathcal{M}} = (G, \{1\}, \{|V|\}, \{1, |V|\})$  is locally identifiable.*

**Proof** Let  $n = |V|$ . Assume  $\mathcal{M} = (G, \{1\}, \{n\}, V)$  is a linear compartmental model with  $G$  given by a path from input 1 to output  $n$  with  $n - 1$  edges. Recall the coefficients on the left hand side of the input–output equation are given by the characteristic polynomial of  $A$ , which is:

$$(\lambda - a_{11})(\lambda - a_{22}) \cdots (\lambda - a_{nn})$$

Since the roots of a polynomial can be determined from its coefficients, then all of  $a_{11}, a_{22}, \dots, a_{nn}$  are locally identifiable. Since the degree of the highest-order term on the right-hand side of the input–output equation is  $n - 1 - \text{dist}(1, n)$  by Lemma 2.23, this reduces to zero so the right-hand side is  $a_{n,n-1} \cdots a_{32}a_{21}u_n$ . Thus the monomial path  $a_{n,n-1} \cdots a_{32}a_{21}$  is identifiable. This means the dimension of the image of the coefficient map is  $|E| + |In \cup Out| = n - 1 + 2 = n + 1$  which is the number of paths and cycles, thus the model is an identifiable path/cycle model. By Theorem 3.19, the model  $\widetilde{\mathcal{M}} = (G, \{1\}, \{n\}, \{1, n\})$  is locally identifiable.  $\square$

In Meshkat and Sullivant (2014), it was shown in Proposition 5.5 that if a model  $\mathcal{M} = (G, \{1\}, \{1\}, V)$  has expected dimension, then the model  $\mathcal{M} = (G', \{1'\}, \{1'\}, V)$  also has expected dimension, where  $G'$  is the new graph obtained from  $G$  by adding a new vertex  $1'$  and an exchange  $1 \rightarrow 1', 1' \rightarrow 1$  and making  $1'$  the new input–output node. We show an analogous result now:

**Proposition 3.30** *Let  $\mathcal{M} = (G, \{1\}, \{j\}, V)$  be a linear compartmental model that has expected dimension where  $|V| = n$  and  $j \neq 1$ . Let  $G'$  be a new graph obtained from  $G$  by adding a set of new vertices  $n + 1, n + 2, \dots, n + k$  and a set of edges  $n + 1 \rightarrow 1, n + 2 \rightarrow n + 1, \dots, n + k \rightarrow n + (k - 1)$  and making  $n + k$  the new input node. Then the model  $\widetilde{\mathcal{M}} = (G', \{n + k\}, \{j\}, V \cup \{n + 1, n + 2, \dots, n + k\})$  also has expected dimension.*

**Proof** Let  $A$  be the full matrix associated to the graph  $G'$  where the first  $k$  rows and  $k$  columns correspond to the added path from compartment  $n + k$  to compartment 1,  $A_k$  be the matrix where the first  $k$  rows and first  $k$  columns have been deleted (and, hence associated to the graph  $G$ ), and let  $E_G$  be the edges of the graph  $G$ . We assume that the dimension of the image of the map  $c$  associated to the model  $\mathcal{M} = (G, \{1\}, \{j\}, V)$  is  $|E_G| + |In \cup Out| = |E_G| + 2$ , and we want to show that for the model  $\widetilde{\mathcal{M}} = (G', \{n + k\}, \{j\}, V \cup \{n + 1, n + 2, \dots, n + k\})$  we get  $|E_{G'}| + 2 = |E_G| + k + 2$ , as we are adding  $k$  new edges.

The input–output equation for the model  $\widetilde{\mathcal{M}} = (G', \{n + k\}, \{j\}, V \cup \{n + 1, n + 2, \dots, n + k\})$  is:

$$\det(\partial I - A)y_j = \det(\partial I - A)_{1,j+k}u_{n+k}$$

where  $\det(\partial I - A) = (\partial - a_{n+1,n+1}) \cdots (\partial - a_{n+k,n+k}) \det(\partial I - A_k)$  and  $\det(\partial I - A)_{1,j+k} = a_{1,n+1}a_{n+2,n+1} \cdots a_{n+k,n+k-1} \det(\partial I - A_k)_{1j}$ . The input–output equation for the model  $\mathcal{M} = (G, \{1\}, \{j\}, V)$  is:

$$\det(\partial I - A_k)y_j = \det(\partial I - A_k)_{1j}u_1$$

For notational ease, write  $\partial$  as  $\lambda$ ,  $|V| = n$ ,  $p = (a_{n+1,n+1}, \dots, a_{n+k,n+k})$ , and  $q = a_{1,n+1}a_{n+2,n+1} \cdots a_{n+k,n+k-1}$ . Note that  $p$  is the vector of new self-cycles and  $q$  can be interpreted as the added monomial path from compartment  $n + k$  to compartment 1. We can write  $\det(\lambda I - A_k)$  as:

$$\lambda^n + c_1\lambda^{n-1} + \cdots + c_{n-1}\lambda + c_n$$

and we can write  $\det(\lambda I - A_k)_{1j}$  as

$$d_1\lambda^{n-1} + d_2\lambda^{n-2} + \dots + d_{n-1}\lambda + d_n.$$

Thus  $\det(\lambda I - A) = (\lambda - a_{n+1,n+1}) \cdots (\lambda - a_{n+k,n+k}) \det(\lambda I - A_k)$  can be written as (up to a minus sign):

$$\begin{aligned} &\lambda^{n+k} + (c_1 - S_1(p))\lambda^{n+k-1} + (c_2 - c_1S_1(p) + S_2(p))\lambda^{n+k-2} \\ &\quad + (c_3 - c_2S_1(p) + c_1S_2(p) - S_3(p))\lambda^{n+k-3} \\ &\quad + \dots + (c_k - c_{k-1}S_1(p) + c_{k-2}S_2(p) - \dots - S_k(p))\lambda^n \\ &\quad + \dots + (c_n - c_{n-1}S_1(p) + c_{n-2}S_2(p) - \dots - c_{n-k}S_k(p))\lambda^k \\ &\quad + (-c_nS_1(p) + \dots + c_{n-k+2}S_{k-1}(p) - c_{n-k+1}S_k(p))\lambda^{k-1} \\ &\quad + \dots + (-c_nS_{k-2}(p) + c_{n-1}S_{k-1}(p) - c_{n-2}S_k(p))\lambda^2 \\ &\quad + (c_nS_{k-1}(p) - c_{n-1}S_k(p))\lambda - c_nS_k(p) \end{aligned} \tag{8}$$

where  $S_1(p), \dots, S_k(p)$  are the  $k$  elementary symmetric polynomials in the parameter vector  $p$ . Here we assumed  $n > k$ , but an analogous formula follows for the case of  $n \leq k$ .

We will refer to the non-constant coefficients of  $\det(\lambda I - A)$  as  $C_1, \dots, C_{n+k}$ . Note that these are by assumption identifiable. Likewise,

$$\det(\lambda I - A)_{1,j+k} = qd_1\lambda^{n-1} + qd_2\lambda^{n-2} + \dots + qd_{n-1}\lambda + qd_n. \tag{9}$$

We will refer to the coefficients of  $\det(\lambda I - A)_{1,j+k}$  as  $D_1, \dots, D_n$ . Note that these are by assumption identifiable. We must now show that if the mapping given by  $(c_1, \dots, c_n, d_1, \dots, d_n)$  has expected dimension, then the new mapping  $(C_1, \dots, C_{n+k}, D_1, \dots, D_n)$  also has expected dimension. Since the parameters in  $p$  are roots of the polynomial  $\det(\lambda I - A)$ , then this means these parameters can be written in terms of the coefficients  $(C_1, \dots, C_{n+k})$ , which are identifiable, and thus the parameters in  $p$  are identifiable. This means each of the elementary symmetric polynomials  $S_1(p), \dots, S_k(p)$  are identifiable. Since  $C_1$  and  $S_1(p)$  are identifiable from Eq. 8, then  $c_1$  can be recovered from the first coefficient from Eq. 8. Likewise, since  $C_2, c_1, S_1(p)$ , and  $S_2(p)$  are identifiable, then  $c_2$  can be recovered from the second coefficient of Eq. 8. Continuing in this fashion, we can recover  $c_1, \dots, c_n$ , i.e.  $c_1, \dots, c_n$  are identifiable. This means the dimension of the image of  $(C_1, \dots, C_{n+k})$  is  $k$  more than the dimension of the image of  $(c_1, \dots, c_n)$ . Since the coefficients of Eq. 9 are just the coefficients  $d_1, \dots, d_n$  scaled by  $q$ , which contains disjoint parameters from the parameters in the coefficients  $c_1, \dots, c_n, d_1, \dots, d_n$ , then the dimension of the image of  $(D_1, \dots, D_n)$  is the same as the dimension of the image of  $(d_1, \dots, d_n)$ . The parameters in  $p$  do not appear in  $(D_1, \dots, D_n)$ , thus combining the maps  $(C_1, \dots, C_{n+k})$  and  $(D_1, \dots, D_n)$  we have that the dimension of the image of the new map  $(C_1, \dots, C_{n+k}, D_1, \dots, D_n)$  must be  $k$  more than the dimension of the image of  $(c_1, \dots, c_n, d_1, \dots, d_n)$  due to the identifiability of the parameters in  $p$ .

Thus the model  $\widetilde{\mathcal{M}} = (G', \{n + k\}, \{j\}, V \cup \{n + 1, n + 2, \dots, n + k\})$  has dimension of the image of the coefficient map equal to  $|E_G| + k + 2$ , i.e.,  $k$  more than  $\mathcal{M} = (G, \{1\}, \{j\}, V)$ . □

### 4 Classification of All Identifiable Models that are Strongly Input–Output Connected with 1 Output or Strongly Connected with 1 Input and Leaks in Input/Output Compartments

Theorem 4.1 gives necessary conditions for strongly input–output connected models with 1 output or strongly connected models with 1 input with leaks in input/output compartments to be identifiable, namely that they must be identifiable path/cycle models when all the leaks are added to the model.

#### 4.1 Necessary Conditions for Identifiability

**Theorem 4.1** (Adding leaks) *Let  $\mathcal{M} = (G, In, Out, L)$  represent a linear compartmental model with  $|L| = |In \cup Out|$  and either  $G$  strongly input–output connected and  $|Out| = 1$  or  $G$  strongly connected and  $|In| = 1$  which we assume has expected dimension, i.e., has dimension of the image of the coefficient map equal to  $|E| + |In \cup Out|$ . Then, the corresponding model with an additional leak  $\widetilde{\mathcal{M}} = (G, In, Out, L \cup \{k\})$  also has expected dimension. Thus the model  $\widetilde{\mathcal{M}} = (G, In, Out, Leak)$  where  $L \subseteq Leak$  and  $|Leak| \leq |V|$  also has expected dimension.*

**Proof** Suppose  $\mathcal{M} = (G, In, Out, L)$  is a linear compartmental model with either  $G$  strongly input–output connected and  $|Out| = 1$  or  $G$  strongly connected and  $|In| = 1$  and  $|L| = |In \cup Out|$  which we assume has expected dimension, i.e., has dimension of the image of the coefficient map equal to  $|E| + |In \cup Out|$ .

Note that because we assume that  $\mathcal{M}$  has expected dimension and  $|L|$  leaks, this implies that the Jacobian of the coefficient map has the expected number of coefficients as the number of rows and  $(|E| + |In \cup Out|)$  columns with full rank  $|E| + |In \cup Out|$ . Note too that the addition of the  $|V| - |L|$  parameters from the leaks being added to the model  $\mathcal{M}$  will not increase the number of coefficients in the resulting input–output equation, as the number of coefficients is the maximal amount by Theorem 2.24.

Therefore, the Jacobian of the coefficient map of the model  $\widetilde{\mathcal{M}} = (G, In, Out, V)$  generated by forcing every compartment in  $\mathcal{M}$  to have a leak, has the same number of rows but now  $(|E| + |V|)$  columns. The dimension of the image of the coefficient map is bounded above by the number of cycles and paths when there are  $|V|$  leaks, which is  $|E| + |In \cup Out|$ . Thus adding  $|V| - |L|$  leaks to a  $|L|$ -leak model cannot increase the dimension of the image of the coefficient map above  $|E| + |In \cup Out|$  if it has already achieved that dimension with  $|L|$  leaks.

Note then that if we consider the specialization generated by substituting zero for every added leak, and consider the submatrix of said Jacobian with expected number of coefficients as the number of rows and  $(|E| + |In \cup Out|)$  columns generated by the  $(|E| + |In \cup Out|)$  columns corresponding to the edges and leaks in  $L$ , we have exactly

the Jacobian of the coefficient map of  $\mathcal{M}$ , which we know is full rank. Therefore, we have that the Jacobian of the coefficient map of  $\widetilde{\mathcal{M}}$  is also full rank, implying that the model has expected dimension.

The same argument applies for adding any number of leaks up to  $|V|$  total leaks.  $\square$

**Remark** We note that while Theorems 3.19 and 4.1 assume opposite operations of adding or subtracting leaks, we have the condition in Theorem 3.19 that  $In \cup Out \subseteq Leak$ , while in Theorem 4.1 only  $|Leak| = |In \cup Out|$  is assumed, i.e., only the number and not the placement of leaks matters.

## 4.2 Necessary and Sufficient Conditions for Identifiability

Combining Theorem 4.1 with Theorem 3.19, we now come to the main result of this section and obtain the following necessary and sufficient conditions for identifiable models:

**Corollary 4.2** *Let  $\mathcal{M} = (G, \{i\}, \{i\}, \{k\})$  represent a linear compartmental model with  $G$  strongly connected and let  $\widetilde{\mathcal{M}} = (G, \{i\}, \{i\}, V)$  be the corresponding model with a leak in every compartment.  $\mathcal{M}$  is locally identifiable if and only if  $\widetilde{\mathcal{M}}$  is an identifiable cycle model.*

**Proof** This follows from combining Theorems 4.1 and 1 of Meshkat et al. (2015) (also written as Theorem 9.3).  $\square$

**Corollary 4.3** *Let  $\mathcal{M} = (G, In, Out, L)$  represent a linear compartmental model with and  $L = In \cup Out$  and assume that either  $G$  is strongly input–output connected and  $|Out| = 1$  or  $G$  is strongly connected and  $|In| = 1$ . Let  $\widetilde{\mathcal{M}} = (G, In, Out, V)$  be the corresponding model with a leak in every compartment.  $\mathcal{M}$  is locally identifiable if and only if  $\widetilde{\mathcal{M}}$  is an identifiable path/cycle model.*

**Proof** This follows from combining Theorems 4.1 and 3.19.  $\square$

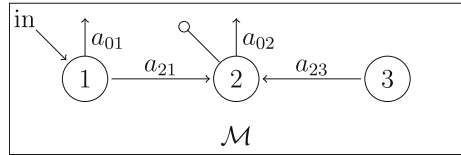
**Remark** Corollary 4.3 gives us a complete classification of all identifiable models that are strongly input–output connected with 1 output or strongly connected with 1 input and leaks in input/output compartments. We note that this class of models has the very special property of being *dimension-preserving* when leaks are added or subtracted from non-input/output compartments, up to a point. To demonstrate that this special dimension-preserving property when removing leaks is not always the case, we revisit Example 3.21.

**Example 4.4** Recall Example 3.21 where we had the model  $\mathcal{M} = (G, \{1\}, \{2\}, L)$  where  $|L| = |In \cup Out| = 2$  and  $G$  is given by the edges  $\{1 \rightarrow 2, 1 \rightarrow 3, 3 \rightarrow 1, 1 \rightarrow 4, 4 \rightarrow 1\}$ . The identifiable models are the ones where  $L = \{2, 4\}, \{2, 3\}, \{1, 2\}$  and the unidentifiable models have  $L = \{3, 4\}, \{1, 4\}, \{1, 3\}$ , so removing leaks from output compartments is not dimension-preserving for this example.

In the next section, we give a conjecture about removing leaks from non-input/output compartments in the general output connectable case for models with one output.



**Fig. 6** The graph corresponding to model  $\mathcal{M}$  from Example 5.3



### 5 Other Expected Dimension Results

We first show that there are at most  $|E| + |In \cup Out|$  independent paths and cycles appearing in the coefficient map  $c$  if we relax the condition of strongly input–output connected to *output connectable* instead:

**Proposition 5.1** *Let  $\mathcal{M} = (G, In, Out, V)$  represent a linear compartmental model with  $G$  output connectable. Assume that  $|Out| = 1$ . Then there are at most  $|E| + |In \cup Out|$  independent paths and cycles in the coefficient map  $c$ .*

**Proof** If  $G$  is strongly input–output connected (or strongly connected) then we have already shown in Lemma 3.11 that the coefficient map factors through  $|E| + |In \cup Out|$  independent paths and cycles. If  $G$  is output connectable but not strongly input–output connected, then there may be fewer than  $|E| + |In \cup Out|$  independent directed paths and directed cycles because there are  $|E| - |V| + |In \cup Out|$  independent directed paths and undirected cycles by Lemma 3.9. Since the coefficient map factors over the directed paths and directed cycles, then there are at most  $|E| + |In \cup Out|$  independent paths and cycles in the coefficient map  $c$ . □

This means the expected dimension is now the number of independent directed paths and directed cycles, which may be less than  $|E| + |In \cup Out|$ .

We can relax the connectedness conditions in Theorem 4.1 to *output connectable* instead and still obtain statements about expected dimension, although now the models with a full set of leaks are **not** identifiable path/cycle models.

**Theorem 5.2** *Let  $\mathcal{M} = (G, In, Out, L)$  represent a linear compartmental model with  $G$  output connectable and  $|L| = |In \cup Out|$  and  $|Out| = 1$  which we assume has dimension of the image of the coefficient map equal to  $|E| + |In \cup Out|$ . Then, the corresponding model with a leak in every compartment  $\tilde{\mathcal{M}} = (G, In, Out, V)$  also has dimension of the image of the coefficient map equal to  $|E| + |In \cup Out|$ .*

**Proof** The proofs mirrors the one in Theorem 4.1. □

**Example 5.3** The model  $\mathcal{M} = (G, \{1\}, \{2\}, \{1, 2\})$  shown in Fig. 6 where  $G$  is the graph given by  $\{1 \rightarrow 2, 3 \rightarrow 2\}$  is output connectable and has dimension of the image of the coefficient map equal to  $|E| + 2 = 4$ , thus it is locally identifiable. By Theorem 5.2, the model  $\tilde{\mathcal{M}} = (G, \{1\}, \{2\}, V)$  also has dimension of the image of the coefficient map equal to  $|E| + 2 = 4$ . Thus the identifiable functions are  $a_{11}, a_{22}, a_{33}, a_{21}$ . Note that it is *not* an identifiable path/cycle model because the parameter  $a_{23}$  does not appear in the coefficient map (as it is not strongly input–output connected).

This result shows that if a model has its dimension of the image of the coefficient map is equal to  $|E| + |In \cup Out|$ , then adding leaks alone maintains the dimension of the

**Fig. 7** The graph corresponding to model  $\mathcal{M}$  from Example 5.5

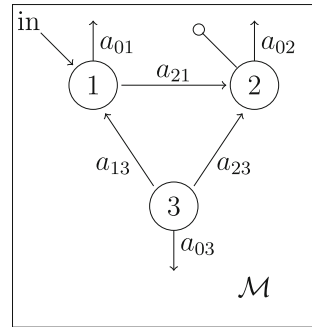


image of the coefficient map. This result is perhaps more useful for its contrapositive, i.e., if a model with leaks from every compartment does not have dimension  $|E| + |In \cup Out|$  for  $c$ , then no amount of removing leaks up to a certain point ( $|L| = |In \cup Out|$ ) can attain identifiability.

**Corollary 5.4** Let  $\tilde{\mathcal{M}} = (G, In, Out, V)$  represent a linear compartmental model with  $G$  output connectable and  $|Out| = 1$  which does **not** have the dimension of the image of the coefficient map equal to  $|E| + |In \cup Out|$ . Then, the corresponding model  $\mathcal{M} = (G, In, Out, L)$  with  $|L| = |In \cup Out|$  is **not** locally identifiable.

**Example 5.5** The model  $\mathcal{M} = (G, \{1\}, \{2\}, V)$  shown in Fig. 7 where  $G$  is the graph given by  $\{1 \rightarrow 2, 3 \rightarrow 2, 3 \rightarrow 1\}$  is output connectable and has dimension of the image of the coefficient map *not* equal to  $|E| + 2 = 5$ , but equal to 4 instead. By Corollary 5.4, the model  $\tilde{\mathcal{M}} = (G, \{1\}, \{2\}, L)$  where  $|L| = 2$  is thus *not* locally identifiable.

These results also give us insight into models that have dimension of the image of the coefficient map equal to  $|E| + |In \cup Out|$  but are not strongly input–output connected. We can show that the self-cycles are always identifiable:

**Theorem 5.6** Let  $\mathcal{M} = (G, In, Out, V)$  represent a linear compartmental model with  $G$  output connectable and  $|Out| = 1$ . If  $\mathcal{M}$  has dimension of the image of the coefficient map equal to  $|E| + |In \cup Out|$ , then the self-cycles  $a_{11}, \dots, a_{nn}$  are locally identifiable.

**Proof** Since the coefficient map always factors over  $a_{11}, \dots, a_{nn}$ , this means the self-cycles are locally identifiable. □

Finally, we give a conjecture on removing leaks from non-input/output compartments for output connectable models and prove this conjecture in a special case:

**Conjecture 5.7** Let  $\mathcal{M} = (G, In, Out, V)$  represent a linear compartmental model. Assume that  $G$  is output connectable and  $|Out| = 1$ . Assume the dimension of the image of the coefficient map is  $k$ . Then, the corresponding model  $\tilde{\mathcal{M}} = (G, In, Out, L)$  where  $In \cup Out \subseteq L$  also has dimension of the image of its coefficient map as  $k$ .

In other words, we conjecture that this property of being dimension-preserving applies to all output connectable models. However, if the dimension to begin with is not maximal, the dimension-preserving property will not lead to identifiability. We give a proof of this conjecture in the special case where the dimension of the image of the coefficient map is  $|E| + |In \cup Out|$ :

**Theorem 5.8** *Let  $\mathcal{M} = (G, In, Out, V)$  represent a linear compartmental model with  $G$  output connectable and  $|Out| = 1$ . If  $\mathcal{M}$  has dimension of the image of the coefficient map equal to  $|E| + |In \cup Out|$ , then the corresponding model  $\tilde{\mathcal{M}} = (G, In, Out, L)$  where  $In \cup Out \subseteq L$  also has dimension of the image of its coefficient map as  $|E| + |In \cup Out|$ . In particular, if  $L = In \cup Out$ , then  $\tilde{\mathcal{M}}$  is locally identifiable.*

To prove Theorem 5.8, we give a variation of Lemma 3.18 and then a variation of the proof of Theorem 3.19.

**Lemma 5.9** *Let  $G = (V, E)$  be a directed graph with corresponding model  $(G, In, Out, V)$ . Assume that  $G$  is output connectable and  $|Out| = 1$ . Consider a model  $(G, In, Out, L)$  where  $In \cup Out \subseteq L$ . Let  $\pi : \mathbb{R}^{|V|+|E|} \rightarrow \mathbb{R}^{|E|+|In \cup Out|}$  denote the path/cycle map. Let  $\Lambda \subseteq \mathbb{R}^{|V|+|E|}$  be the linear space satisfying*

$$\Lambda = \{A \in \mathbb{R}^{|V|+|E|} : a_{ii} = - \sum_{j, j \neq i} a_{ji} \text{ for all } i \notin L\}.$$

*If the dimension of the image of  $\pi$  is  $|E| + |In \cup Out|$ , then the dimension of the image of  $\Lambda$  under the map  $\pi$  is  $|E| + |In \cup Out|$ .*

**Proof** Removing the assumption of strongly input–output connected from Lemma 3.18 means that we cannot guarantee there are  $|E| + |In \cup Out|$  independent directed paths and directed cycles. However, if we assume the dimension of the image of  $\pi$  is  $|E| + |In \cup Out|$ , then the rest of the proof follows that of Lemma 3.18.  $\square$

**Proof of Theorem 5.8** By Lemma 5.9 we know that the image of the restricted parameter space under the path/cycle map  $\pi$  has dimension  $|E| + |In \cup Out|$ , which is equal to the dimension of the image of the full parameter space under the path/cycle map. Since the dimension of the image of the coefficient map  $c$  is  $|E| + |In \cup Out|$ , this must be the same for the restricted model. In particular, if  $|L| = |In \cup Out|$ , then the model has  $|E| + |In \cup Out|$  parameters, hence it is locally identifiable.  $\square$

**Example 5.10** (Example 5.3 revisited) The model  $\tilde{\mathcal{M}} = (G, \{1\}, \{2\}, V)$  where  $G$  is the graph given by  $\{1 \rightarrow 2, 3 \rightarrow 2\}$  is output connectable and has dimension of the image of the coefficient map equal to  $|E| + 2 = 4$ , thus the model given by  $\mathcal{M} = (G, \{1\}, \{2\}, \{1, 2\})$  is locally identifiable.

Combining Theorems 5.2 and 5.8, we get the following necessary and sufficient conditions:

**Corollary 5.11** *Let  $\mathcal{M} = (G, In, Out, L)$  represent a linear compartmental model with and  $L = In \cup Out$  and assume that  $G$  is output connectable and  $|Out| = 1$ . Let*

$\widetilde{\mathcal{M}} = (G, In, Out, V)$  be the corresponding model with a leak in every compartment.  $\mathcal{M}$  is locally identifiable if and only if  $\widetilde{\mathcal{M}}$  has dimension of the image of the coefficient map as  $|E| + |In \cup Out|$ .

**Remark** Corollary 5.11 shows that this dimension-preserving property when adding or removing leaks from non-input/output compartments also holds in the output connectable case when the dimension of the image of the coefficient map is  $|E| + |In \cup Out|$ .

## 6 Necessary Conditions for Identifiable Models Based on Model Structure

Outside of checking the conditions in Theorem 3.22, i.e., if a model is an inductively strongly connected model if edges from output to input are added, checking if a model is an identifiable path/cycle model amounts to checking the dimension of the image of the coefficient map, and thus cannot be ascertained by simply examining the graph of the model. However, it is possible to provide necessary conditions for identifiable models and identifiable path/cycle models based on the graph itself, and thus can be used to rule out identifiability.

**Theorem 6.1** Let  $\mathcal{M} = (G, In, Out, L)$  represent a linear compartmental model with either  $G$  strongly input–output connected and  $|Out| = 1$  or  $G$  strongly connected and  $|In| = 1$ . If  $|L| > |In \cup Out|$ , then  $\mathcal{M}$  is unidentifiable.

**Proof** If the number of parameters  $|E| + |L| > |E| + |In \cup Out|$ , where  $|E| + |In \cup Out|$  is the maximal dimension by Lemma 3.11, then the model is unidentifiable. This reduces to  $|L| > |In \cup Out|$ .  $\square$

We can now make some statements about necessary conditions in the case of the maximal amount of edges.

**Theorem 6.2** Let  $\mathcal{M} = (G, \{i\}, \{i\}, \{k\})$  represent a linear compartmental model with  $G$  strongly connected and  $2|V| - 2$  edges. If  $\mathcal{M}$  is locally identifiable (or equivalently,  $\widetilde{\mathcal{M}} = (G, \{i\}, \{i\}, V)$  is an identifiable cycle model), it must have an exchange.

**Proof** From Proposition 5.3 of Meshkat and Sullivant (2014), we know that  $G$  must have an exchange in order for  $\widetilde{\mathcal{M}} = (G, \{i\}, \{i\}, V)$  to be an identifiable cycle model. Thus, if  $G$  does not have an exchange,  $\widetilde{\mathcal{M}}$  is not an identifiable cycle model and thus  $\mathcal{M} = (G, \{i\}, \{i\}, \{k\})$  is not an identifiable model.  $\square$

**Theorem 6.3** Let  $\mathcal{M} = (G, \{i\}, \{j\}, L)$  represent a linear compartmental model with  $G$  strongly input–output connected,  $\text{dist}(i, j) = 1$ , and  $2|V| - (\text{dist}(i, j) + 2)$  edges and  $|L| = |In \cup Out|$ . If  $\mathcal{M}$  is locally identifiable (or equivalently,  $\widetilde{\mathcal{M}} = (G, \{i\}, \{j\}, V)$  is an identifiable path/cycle model), it must have an edge from  $i$  to  $j$  (i.e., a path of length  $\text{dist}(i, j)$ ).

**Proof** If there is no path from  $i$  to  $j$ , then the coefficient of the highest-order term on the right-hand side of the input–output equation is zero and there would be fewer than

$2|V| - \text{dist}(i, j)$  coefficients. But there are  $2|V| - (\text{dist}(i, j) + 2)$  edges, so if  $\mathcal{M}$  is locally identifiable, then it has expected dimension  $|E| + 2 = 2|V| - \text{dist}(i, j)$ , which is impossible if there are fewer than  $2|V| - \text{dist}(i, j)$  coefficients.  $\square$

We can also have an analogous necessary condition in the case of fewer than  $2|V| - (\text{dist}(i, j) + 2)$  edges.

**Theorem 6.4** *Let  $\mathcal{M} = (G, \{i\}, \{j\}, L)$  with  $i \neq j$  represent a linear compartmental model with  $G$  strongly input–output connected and  $2|V| - (k + 2)$  edges where  $k \geq 1$  and  $|L| = |\text{In} \cup \text{Out}|$ . If  $\mathcal{M}$  is locally identifiable (or equivalently,  $\widehat{\mathcal{M}} = (G, \{i\}, \{j\}, V)$  is an identifiable path/cycle model), it must have a path from  $i$  to  $j$  of length at most  $k$ .*

**Proof** The coefficient of the highest-order term on the right-hand side of the input–output equation is a sum of shortest paths from input to output of length  $\text{dist}(i, j)$  and this must be nonzero for  $\mathcal{M}$  to have expected dimension. Since there are  $2|V| - \text{dist}(i, j)$  nonzero coefficients and expected dimension is  $|E| + 2$ , this means  $k$  is at least  $\text{dist}(i, j)$ . So there must be a path from  $i$  to  $j$  of length at most  $k$ .  $\square$

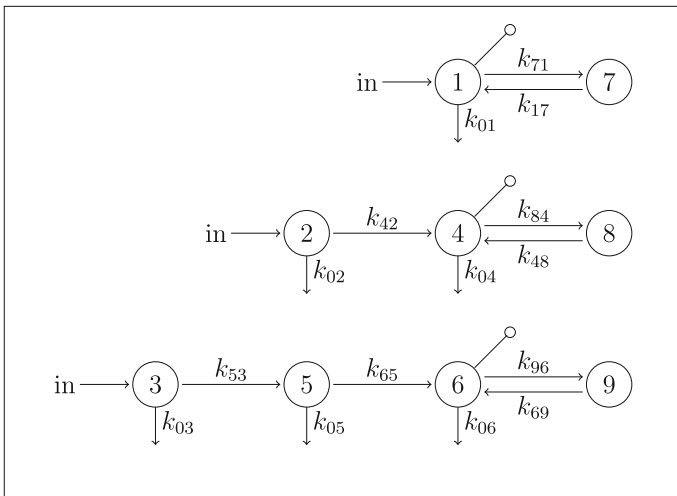
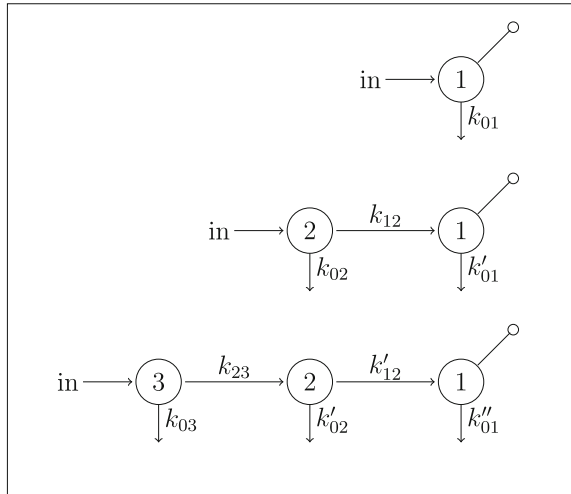
## 7 Examples

We now provide some real-world examples that fall into the categories of models considered in this paper. In particular, we obtain identifiability or unidentifiability results in Examples 7.1 and 7.2 without any symbolic computation, i.e., purely based on the graph structure alone.

**Example 7.1** Consider Example 13.6 from DiStefano (2015) on HIV vaccine development (Part 1). Three models are considered that fall into the category of path models with leaks from every compartment as in Proposition 3.29, shown in Fig. 8. The top model corresponds to Experiment 1, the middle model corresponds to Experiment 2, and the bottom model corresponds to Experiment 3. It is clear that the Experiment 1 model is identifiable. Using Proposition 3.29, we can easily obtain that the Experiment 2 model is identifiable. By Proposition 3.29, the Experiment 3 model has expected dimension and is thus unidentifiable with identifiable functions given by the paths and cycles (where the “self-cycles” have been expanded out):  $k_{23}k'_{12}$ ,  $-k_{03} - k_{23}$ ,  $-k'_{02} - k'_{12}$ ,  $-k''_{01}$ .

**Example 7.2** Consider Example 13.16 from DiStefano (2015) on HIV vaccine development (Part 3). The models from Example 7.1 are amended by adding on exchanges to the output compartments, shown in Fig. 9, and the numbering scheme has changed to agree with DiStefano (2015). The Experiment 1 model is identifiable by Theorem 3.19 as it is an identifiable cycle model (it is inductively strongly connected) with a single leak in the input/output compartment. The Experiment 2 model is almost inductively strongly connected and thus is identifiable by Corollary 3.27. A variation on the Experiment 3 model with leaks from all compartments can be shown to be an identifiable path/cycle model by a direct calculation, thus removing the leak from

**Fig. 8** The three models considered in Example 7.1



**Fig. 9** The three models considered in Example 7.2

compartment 9 retains the dimension by Theorem 3.19, which means the model in Experiment 3 is unidentifiable. Alternatively, one can apply Proposition 3.30 to a variation on the model in Experiment 2 with leaks from every compartment (which has

expected dimension) and thus obtain that the variation on the model in Experiment 3 with leaks from every compartment also has expected dimension. Now removing the leak from compartment 9 retains the dimension by Theorem 3.19, and thus the model in Experiment 3 is unidentifiable. The identifiable functions are given by the paths and cycles (where the “self-cycles” have been expanded out):  $k_{53}k_{65}$ ,  $k_{69}k_{96}$ ,  $-k_{03} - k_{53}$ ,  $-k_{05} - k_{65}$ ,  $-k_{06} - k_{96}$ ,  $-k_{69}$ .

## 8 Computations

In the table below we outline the number of graphs with  $n$  vertices and  $m$  edges that have the expected dimension with input in  $i$  and output in  $j$ , assuming leaks from every compartment:

The number of strongly connected graphs and the number of strongly input–output connected graphs with different input/output configurations is noted.

We then computed the number of models with expected dimension for 4 notable cases: the case of identical single input and single output with a strongly connected graph  $G$  (as in Meshkat and Sullivant 2014), the case of single input but multiple outputs with a strongly connected graph  $G$ , the case of distinct single input and single output with a strongly input–output connected graph  $G$ , and the case of single output but multiple inputs with a strongly input–output connected graph  $G$ . Due to restrictions on the number of edges, not all cases are possible, and those are labeled “NA.”

## 9 Construction of Identifiable Models

In this section, we consider the special case of single input and single output in the same compartment with  $G$  strongly connected and  $Leak = V$ , as in Meshkat and Sullivant (2014), i.e.  $\mathcal{M} = (G, \{i\}, \{i\}, V)$  with  $G$  strongly connected. Since  $Leak$  is assumed to be  $V$  and input/output are assumed to be the same vertex, we can just discuss the graph  $G$  in what follows.

In Meshkat and Sullivant (2014), Theorem 5.7 gives a way of constructing a new model with expected dimension from a smaller model with expected dimension by adding an incoming and outgoing edge to a chain of cycles (See Definition 3.23):

**Theorem 9.1** (Theorem 5.7 of Meshkat and Sullivant 2014) *Let  $G'$  be a graph that has the expected dimension with  $n - 1$  vertices. Let  $G$  be a new graph obtained from  $G'$  by adding a new vertex and two edges  $k \rightarrow n$  and  $n \rightarrow l$  and such that  $G$  has a chain of cycles containing both 1 and  $n$ . Then  $G$  has the expected dimension.*

In Baaijens and Draisma (2016), the authors strengthened this result to allow for adding loops of any length, not just length two:

**Proposition 9.2** (Proposition 4.14 of Baaijens and Draisma 2016) *Let  $G = (V, E)$  on  $n - 1$  vertices be a graph with the expected dimension. Construct  $G'$  from  $G$  by adding new vertices  $n_1, \dots, n_s$  and edges  $k \rightarrow n_1$ ,  $n_s \rightarrow l$ , and  $n_i \rightarrow n_{i+1}$  for  $i = 1, \dots, s - 1$  where  $k, l \in V$  are vertices of  $G$ . Then  $G'$  has the expected dimension.*

$(n, m)$	Total	Strongly connected	$i = 1, j = 1$	$i = 1, j = 2, 3$	Strongly input-output connected $i = 1, j = 2$	$i = 1, j = 2$	Strongly input-output connected $i = 1, 3, j = 2$	$i = 1, 3, j = 2$
(3,2)	15	NA	NA	NA	1	1	3	3
(3,3)	20	2	2	2	7	4	10	8
(3,4)	15	9	3	3	11	NA	12	4
(4,3)	220	NA	NA	NA	2	2	7	7
(4,4)	495	6	6	6	37	25	72	59
(4,5)	792	84	54	62	193	70	267	167
(4,6)	924	316	166	118	445	NA	518	184
(4,7)	792	492	NA	86	565	NA	603	96
(5,4)	4845	NA	NA	NA	6	6	24	24
(5,5)	15,504	24	24	24	222	162	518	432
(5,6)	38,760	720	576	600	2470	1288	4130	1110
(5,7)	77,520	6440	4052	4030	13,004	3154	17,708	1552
(5,8)	125,970	26,875	9565	10,336	40,126	NA	48,277	17,113
(5,9)	167,960	65,280	NA	15,984	82,159	NA	91,658	20,272
(5,10)	184,756	105,566	NA	9841	120,202	NA	128,003	10,689



We will use Proposition 4.14 from Baaijens and Draisma (2016) combined with Theorem 1 of Meshkat et al. (2015) to form identifiable models. In other words, we will use Proposition 4.14 to construct identifiable cycle models and then use Theorem 1 to eliminate all but one leak to form an identifiable model. We state Theorem 1 here:

**Theorem 9.3** (Theorem 1 from Meshkat et al. 2015) *Let  $M$  be an identifiable cycle model. If the model is changed to have exactly one leak, then the resulting model is locally identifiable.*

**Algorithm 9.4** (Construction of identifiable models with  $In = Out = \{1\}$  and one leak  $|L| = 1$ )

- (1) Begin with  $(G, \{1\}, \{1\}, \{1\})$  where  $V = \{1\}$  and  $E = \emptyset$ .
- (2) Construct  $G'$  from  $G$  by adding new vertices  $n_1, \dots, n_s$  and edges  $1 \rightarrow n_1$ ,  $n_s \rightarrow 1$ , and  $n_i \rightarrow n_{i+1}$  for  $i = 1, \dots, s - 1$  where  $k, l \in V$  are vertices of  $G$  and adding leaks from every new vertex.
- (3) Repeat Step 2 by starting at some vertex  $n_i$  and ending at some vertex  $n_j$  for  $n_i, n_j \in \{1, n_1, \dots, n_s\}$  and adding leaks from every new vertex.
- (4) Continue adding edges, vertices, and leaks as described in Steps 2 and 3.
- (5) Remove all leaks except one leak.

**Theorem 9.5** *Let  $\mathcal{M} = (G, \{1\}, \{1\}, \{k\})$  be a model constructed from Algorithm 9.4. The model  $\mathcal{M}$  is identifiable.*

**Proof** By Proposition 9.2 the model  $\mathcal{M}$  is an identifiable cycle model and by Theorem 9.3 the model with only one leak is identifiable.  $\square$

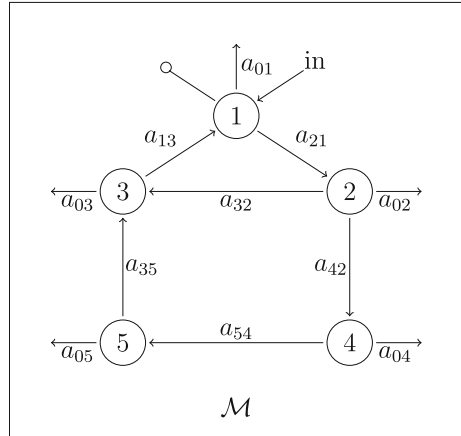
**Example 9.6** The model  $\mathcal{M} = (G, \{1\}, \{1\}, V)$  shown in Fig. 10 where  $G$  is given by the edges  $\{1 \rightarrow 2, 2 \rightarrow 3, 3 \rightarrow 1, 2 \rightarrow 4, 4 \rightarrow 5, 5 \rightarrow 3\}$  is an identifiable cycle model by Proposition 9.2. Thus we can remove all leaks except one, e.g.,  $\widetilde{\mathcal{M}} = (G, \{1\}, \{1\}, \{5\})$  and the resulting model is identifiable. We note that this model is *not* inductively strongly connected, thus Proposition 9.2 does expand upon the results in Meshkat et al. (2015) to construct identifiable models.

Note that the authors in Baaijens and Draisma (2016) only considered *identifiable cycle models*. We suspect there may be a similar result to Proposition 9.2 for the case of identifiable path/cycle models.

## 10 Conclusion and Future work

In this work, we have defined identifiable path/cycle models and found sufficient conditions for obtaining them. We have also formed necessary and sufficient conditions for identifiable models with certain graph properties. Most importantly, we have demonstrated this notion of maximal dimension of the image of the coefficient map in terms of the number of edges and the number of inputs and outputs and have shown that the only identifiable models with certain graph properties are models where the dimension of the image of the coefficient map is preserved to be maximal while adding or

**Fig. 10** The graph corresponding to  $\mathcal{M}$  from Example 9.6



subtracting leaks up to a point. This shows that identifiable models of this form have a special dimension-preserving property even when they are no longer identifiable from added leaks. We hope that this motivates the investigation of other classes of models that have this same dimension-preserving property.

As mentioned in the introduction, one of the main uses of identifiable functions of parameters is to find identifiable reparametrizations over those functions of parameters. In Meshkat and Sullivant (2014), necessary and sufficient conditions for an identifiable scaling reparametrization were found and an algorithm to find such a scaling reparametrization was discussed for the special case of single input/single output in the same compartment,  $G$  strongly connected, and leaks from every compartment. We end this work with the following conjecture to generalize these necessary and sufficient conditions for the case of multiple inputs or multiple outputs:

**Conjecture 10.1** *Let  $\mathcal{M} = (G, In, Out, V)$  represent a linear compartmental model with  $G$  strongly input–output connected and  $|Out| = 1$  or  $G$  strongly connected and  $|In| = 1$ . An identifiable scaling reparametrization exists if and only if  $\mathcal{M} = (G, In, Out, V)$  is an identifiable path/cycle model.*

While the models described in Corollary 5.11 also have the dimension of the image of the coefficient map as  $|E| + |In \cup Out|$ , we note that these models will not have an identifiable scaling reparametrization since these models do not necessarily have every parameter appearing in a cycle or path from input to output. However, it may be possible to reparametrize those models another way, e.g., scaling and adding. We leave these problems of identifiable reparametrizations of models with maximal dimension of the image of their coefficient map for future work.

**Acknowledgements** This project began at an AIM workshop on “Identifiability problems in systems biology,” and the authors thank AIM for providing financial support and an excellent working environment. The authors also thank Seth Sullivant, Anne Shiu, Gleb Pogudin, and two referees for their helpful comments. Cashous Bortner was partially supported by the US National Science Foundation (DMS 1615660). Nicolette Meshkat was partially supported by the Clare Boothe Luce fellowship from the Henry Luce Foundation and the US National Science Foundation (DMS 1853525).

## References

- Baaijens JA, Draisma J (2016) On the existence of identifiable reparametrizations for linear compartment models. *SIAM J Appl Math* 76(4):1577–1605
- Bellman R, Åström KJ (1970) On structural identifiability. *Math Biosci* 7(3–4):329–339
- Bellu G, Saccomani MP, Audoly S, D’Angiò L (2007) DAISY: a new software tool to test global identifiability of biological and physiological systems. *Comput Methods Prog Biol* 88(1):52–61
- Biggs N (1993) Algebraic graph theory. Cambridge University Press, Cambridge
- Bressloff PC, Taylor JG (1993) Compartmental-model response function for dendritic trees. *Biol Cybern* 70(2):199–207
- DiStefano JJ III (2015) Dynamic systems biology modeling and simulation. Academic Press, New York
- Gerberding S, Obatake N, Shiu A (2020) Identifiability of linear compartmental models: the effect of moving inputs, outputs, and leaks. *Linear Multilinear Algebra* 0:1–22
- Glad ST (1990) Differential algebraic modelling of nonlinear systems. Birkhäuser Boston, Boston, pp 97–105
- Godfrey KR, Chapman MJ (1990) Identifiability and indistinguishability of linear compartmental models. *Math Comput Simul* 32:273–295
- Gross E, Harrington H, Meshkat N, Shiu A (2019) Linear compartmental models: input–output equations and operations that preserve identifiability. *SIAM J Appl Math* 79(4):1423–1447
- Harris J (1992) Algebraic geometry. Graduate Texts in Mathematics. Springer, Berlin
- Hong H, Ovchinnikov A, Pogudin G, Yap C (2020) Global identifiability of differential models. *Commun Pure Appl Math* 73(9):1831–1879
- Ljung L, Glad T (1994) On global identifiability for arbitrary model parametrizations. *Automatica* 30(2):265–276
- Meshkat N, Rosen Z, Sullivant S (2018) Algebraic tools for the analysis of state space models. In: The 50th anniversary of Gröbner bases, *Adv. Stud. Pure Math.*, vol 77. Math. Soc. Japan, Tokyo, pp 171–205
- Meshkat N, Sullivant S (2014) Identifiable reparametrizations of linear compartment models. *J Symb Comput* 63:46–67
- Meshkat N, Sullivant S, Eisenberg M (2015) Identifiability results for several classes of linear compartment models. *Bull Math Biol* 77(8):1620–1651
- Ollivier F (1990) *Le Problème de l’Identifiabilité Structurelle Globale: Étude Théorique, Méthodes Effectives et Bornes de Complexité*. PhD thesis, École Polytechnique
- Ovchinnikov A, Pogudin G, Thompson P (2022) Input–output equations and identifiability of linear ode models. *IEEE Trans Automat Contr*. <https://doi.org/10.1109/TAC.2022.3145571>
- Pohjanpalo H (1978) System identifiability based on the power series expansion of the solution. *Math Biosci* 41(1):21–33
- Saccomani MP, Audoly S, D’Angiò L (2003) Parameter identifiability of nonlinear systems: the role of initial conditions. *Automatica* 39(4):619–632
- Soderstrom T, Stoica P (1989) System identification. Prentice-Hall, Hoboken
- Vajda S, Godfrey KR, Rabitz H (1989) Similarity transformation approach to identifiability analysis of nonlinear compartmental models. *Math Biosci* 93(2):217–248
- Villaverde AF (2019) Observability and structural identifiability of nonlinear biological systems. *Complexity* 2019:8497093
- Walter E, Lecourtier Y (1982) Global approaches to identifiability testing for linear and nonlinear state space models. *Math Comput Simul* 24(6):472–482

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.