

# SCIENTIFIC REPORTS



OPEN

## A symbolic network-based nonlinear theory for dynamical systems observability

Christophe Letellier<sup>1</sup>, Irene Sendiña-Nadal<sup>2,3</sup>, Ezequiel Bianco-Martinez<sup>4</sup> & Murilo S. Baptista<sup>4</sup>

Received: 6 October 2017

Accepted: 6 February 2018

Published online: 28 February 2018

When the state of the whole reaction network can be inferred by just measuring the dynamics of a limited set of nodes the system is said to be fully observable. However, as the number of all possible combinations of measured variables and time derivatives spanning the reconstructed state of the system exponentially increases with its dimension, the observability becomes a computationally prohibitive task. Our approach consists in computing the observability coefficients from a symbolic Jacobian matrix whose elements encode the linear, nonlinear polynomial or rational nature of the interaction among the variables. The novelty we introduce in this paper, required for treating large-dimensional systems, is to identify from the symbolic Jacobian matrix the minimal set of variables (together with their time derivatives) candidate to be measured for completing the state space reconstruction. Then symbolic observability coefficients are computed from the symbolic observability matrix. Our results are in agreement with the analytical computations, evidencing the correctness of our approach. Its application to efficiently exploring the dynamics of real world complex systems such as power grids, socioeconomic networks or biological networks is quite promising.

Variables spanning the state space of a dynamical system which is irreducible to a few smaller subsystems are always dependent on each other through linear and nonlinear interactions. Consequently, one may expect to be able to determine an adequate subset of variables together with their well-selected Lie derivatives to get a full observability of the underlying dynamics, that is, for distinguishing all possible states of the network<sup>1,2</sup>. With the emergence of the Science of Complexity, complex networks are more and more often considered in various fields as well exemplified by power grids<sup>3</sup>, socio-economics networks<sup>4-6</sup>, or biological systems<sup>7-10</sup>. To allow a reliable monitoring, dynamical analysis or control of these high-dimensional systems, suitable and systematic techniques are required to identify the subset of variables providing the best (if not the full) observability of their underlying dynamics. A related problem is how to unfold the whole dynamics by completing this subset of variables to reconstruct a space whose dimension is at least equal to the dimension of the original state space.

Dealing with multivariate time series, specially those produced by high-dimensional dynamical networks, is not a trivial problem<sup>11-13</sup>. Attempts to estimate network observability using symbolic techniques<sup>14,15</sup> were made to overcome the large computational times associated with the exact analytical calculations. In those approaches, a dimension reduction is performed in real time on a symbolic observability matrix until state estimation is possible from the selected measurements. However, linear and nonlinear interactions among variables are considered on an equal footing while it is strongly required to distinguish them for a reliable assessment of the observability of a system<sup>2,13</sup>. In order to tackle such a challenging task, we propose a methodological approach that will be applied to reaction networks derived from dynamical systems with appropriately large dimension to corroborate our assessments with rigorous analytical calculations, and yet provide a framework making also possible the verification of observability in networked dynamical systems. The chosen reaction networks are models of interesting biological and physical systems: the circadian oscillation in the *Drosophila* period protein, the Rayleigh-Bénard convection, and the DNA replication. They also represent nonlinear systems with increasing nonlinear complexity, commonly observed in other natural and man-made systems. Therefore, they are an appropriate subset of

<sup>1</sup>CORIA-UMR 6614 Normandie Université, Campus Universitaire du Madrillet, F-76800, Saint-Etienne du Rouvray, France. <sup>2</sup>Complex Systems Group, Universidad Rey Juan Carlos, 28933, Móstoles, Madrid, Spain. <sup>3</sup>Center for Biomedical Technology, Universidad Politécnica de Madrid, 28223, Pozuelo de Alarcón, Madrid, Spain. <sup>4</sup>Institute for Complex Systems and Mathematical Biology, SUPA, University of Aberdeen, Old Aberdeen, AB24 3UE, United Kingdom. Correspondence and requests for materials should be addressed to C.L. (email: [Christophe.Letellier@coria.fr](mailto:Christophe.Letellier@coria.fr))

nonlinear systems to serve as testbed of our approach's performance and reliability. In addition, we will show that our approach correctly identifies whether a nonlinear dynamical system is fully or only partially observable, an information not accurately obtained by previously proposed methodological approaches<sup>16–19</sup>.

Since we are dealing with dynamical systems in general, the Jacobian matrix will be used to access the nature of how variables interact, allowing us to optimize our assessment of the symbolic observability coefficients. For high-dimensional complex systems this is a quite demanding computation since the number of cases to investigate increases with the system's dimension and exact analytical computations are prohibitive. Indeed, in practice, monitoring all the variables defining the system's state is experimentally infeasible or inefficient, and it is of utmost importance to develop a methodological framework addressing the problem of targeting those variables yielding full observability. Despite several approaches have been proposed<sup>16,17</sup>, most of them neglect the nonlinear nature typically exhibited by complex systems and/or do not provide the space reconstructed from the measured variables. On the one hand, since nonlinearities are most often related to a lack of observability, linear approaches cannot properly address this problem. On the other hand, finding the appropriate combination of sensors (and time derivatives) spanning the reconstructed space is a very time demanding computational task for large dimensional systems.

Here, we adopted a nonlinear symbolic approach taking into account the nature of the interactions among variables and analyze the distribution of the linear and nonlinear load of the variables in the symbolic Jacobian matrix of the system. By means of two easy-to-implement criteria we are able to successfully identify the minimal set of variables (and their time derivatives) candidate to be measured for completing the reconstructed space. Our results are in full agreement with the analytical prediction of getting a no null determinant of the observability matrix and the technique drastically reduces the search for candidate variables, thus providing a key step to observe and model natural and man-made complex systems of large dimension.

The subsequent part of the paper is organized as follows. Section Results is devoted to illustrate how our proposed approach works considering a few large dimensional dynamical systems. Section Methods briefly introduces the observability theory, and the current challenges for the determination of a system's observability in nonlinear systems. Finally, the Discussion section provides some conclusions to this work.

## Results

On rare occasions, nonlinear systems are fully observable from just a single scalar time series<sup>20</sup> as previously investigated for many chaotic systems<sup>21–23</sup>. Since full (global) observability warrants that every distinct point of the original state space  $x \in \mathbb{R}^d$  can be univocally identified, there is a great interest to target the minimal set of variables to measure for accomplishing such a full observability condition.

As shown in Section Methods, to assess the (local or global) observability through a given measurement vector  $s$ , both a subset of  $m$  variables—sometimes designated as “sensors”<sup>24</sup>—and the Lie derivatives have to be provided. Our aim is therefore to provide a method that can indeed solve the problem of determining the minimum set of variables to measure for observing a large complex system. In order to avoid testing the rank of the observability matrix via algebraic computation, we propose to use a technique based on a symbolic computation of the observability matrix in which the terms are not explicitly expressed but only their linear, nonlinear polynomial or rational character<sup>2,13</sup>.

The general and systematic procedure developed by Bianco-Martinez and coworkers<sup>2</sup> and compiled in the Methods section, is in fact very time consuming since the computation of the observability coefficients corresponding to all the  $5.2 \cdot 10^6$  possible vectors spanning a reconstructed space of a 13-dimensional system would require intense and long computational time (more than 18 days). One optimization strategy is to reduce the number of possible combinations by identifying candidate variables that should be disregarded as members of the measurement set. A lack of observability has its origin in the existence of a singular observability manifold, a domain in the original state space where the determinant  $\text{Det } \mathcal{O}$  of the observability matrix  $\mathcal{O}$  is zero<sup>25</sup>. Let us note here that there is one very special case in which  $\text{Det } \mathcal{O} \approx 0$  and, consequently, practical problems in the state estimation may occur. This usually happens when  $\text{Det } \mathcal{O}$  depends on some parameter(s) which may be arbitrarily small<sup>26</sup>. In general, a linear system may be (rarely and practically) non-observable with a nonzero determinant of the observability matrix. By construction, a null or non-constant determinant  $\text{Det } \mathcal{O}$  is rooted in a null or a nonlinear component in the Jacobian matrix. Our technique relies precisely on tracking those nonlinear terms in the Jacobian matrix and, therefore, taking into account both linear and nonlinear interactions between variables becomes so relevant in assessing observability.

By analogy with what is done for chemical reactions<sup>27</sup>, it is possible to consider any dynamical system as a reaction network, whose associated weighted adjacency matrix is the symbolic Jacobian matrix  $\tilde{\mathcal{J}}$ . Using the terminology from graph theory, we define the linear out-strength of the node  $i$ ,  $\sigma_{\text{out}}^{\text{lin}}(i)$ , as the number of times the  $i$ th variable appears in linear terms in the governing equations, that is,

$$\sigma_{\text{out}}^{\text{lin}}(i) = \sum_{\substack{j \neq i \\ j|\tilde{\mathcal{J}}_{ji}=1}} \tilde{\mathcal{J}}_{ji} \quad (1)$$

The larger is  $\sigma_{\text{out}}^{\text{lin}}(i)$ , the higher the probability the  $i$ th variable needs not to be measured because it is related to other variables *via* linear couplings which will not induce nonlinear terms in the determinant of the observability matrix.

The situation in which  $\tilde{\mathcal{J}}_{ij} = \tilde{\mathcal{J}}_{ji} = 1$  and  $\sigma_{\text{out}}^{\text{lin}}(i) = \sigma_{\text{out}}^{\text{lin}}(j) = 1$  means that variables  $i$  and  $j$  are exclusively linearly coupled with no other variables involved. Consequently, full observability of the  $i$ th variable can only be accessed by measuring the  $j$ th variable and *vice versa*. It is thus necessary (and sufficient) to measure at least one of them because they cannot be simultaneously excluded from the set of measured variables. A second criterion to decide which variable to choose between these two is needed. The idea is built on how a variable candidate to be

non-measured is influenced by the other candidate variables. Let be  $\{x_k\}$  the set of variables candidate to be non measured with  $k \in V_{nm} \subset \{1, 2, \dots, m\}$  and  $V_{nm}$  the set of integers indexing the non-measured variables. Thus, the in-strength of the  $i$ th variable provided by the non-measured ones is defined as

$$\sigma_{in}^{nm}(i) = \sum_{\substack{k \in V_{nm} \\ k \neq i}} \tilde{J}_{ik} \tag{2}$$

where  $\tilde{J}_{ik} = 1 \equiv \bar{1}$ ,  $\tilde{J}_{ik} = \bar{2}$ , and  $\tilde{J}_{ik} = \bar{3}$ . Using this correspondence for the symbolic terms, we can assume that the larger is  $\sigma_{in}^{nm}(i)$ , the less observable through the  $i$ th variable the system is. The rationale is as follows: the more nonlinearly coupled is the  $i$ th function  $f_i(\mathbf{x})$  with the non-measured variables, the larger the degree of the determinant of the observability matrix, and the less observable the dynamics through the measurements is<sup>22</sup>. Therefore, we should preferably remove that variable with the largest non-measured in-strength  $\sigma_{in}^{nm}$ .

Therefore, the minimal set of variables to measure for reconstructing a state space with a full observability can be automatically determined from (i) the symbolic Jacobian matrix  $\tilde{\mathcal{J}}$  of the system under study, (ii) the linear out-strength  $\sigma_{out}^{lin}$ , and (iii) the in-strength  $\sigma_{in}^{nm}$  provided by the non-measured variables. Note that the knowledge of the exact functional dependence of the coupling between variables is not necessary, only its polynomial or fractional nature<sup>2</sup>. From the vector of state variables  $\mathbf{x}$ , those components  $x_i$  having the largest  $\sigma_{out}^{lin}$  are discarded as candidates to be sensors after having checked they are present at least once in the equations governing the dynamics of a sensor variable. All remaining possible embeddings  $\mathbf{s}$  that can be constructed from the final set of variables candidate to become sensors are then tested using a Matlab<sup>®</sup> algorithm and ranked according to the corresponding estimated symbolic observability coefficient  $\eta_{\mathbf{s}}$ .

In order to validate whether our proposed method is in agreement with algebraic computations, we will consider three dynamical systems of increasing dimension ( $d = 5, 9$  and  $13$ ) describing complex systems coming from biology or physics. As it is known that the presence of symmetries in the state space can affect the assesment of observability<sup>22</sup>, we will also consider the case of equivariant dynamical systems obeying  $f(\Gamma \cdot \mathbf{x}) = \Gamma \cdot f(\mathbf{x})$ , where  $\Gamma$  defines a discrete symmetry like a rotation or an inversion<sup>28</sup>.

**A 5D rational model for the circadian PER oscillations in *Drosophila*.** In our attempt to consider biological or physically motivated systems, let us start with the model

$$\begin{cases} \dot{x}_1 = \frac{v_s K_I^4}{K_I^4 + x_5^4} - \frac{v_m x_1}{K_m + x_1} \\ \dot{x}_2 = k_s x_1 - \frac{V_1 x_2}{K_1 + x_2} + \frac{V_2 x_3}{K_2 + x_3} \\ \dot{x}_3 = \frac{V_1 x_2}{K_1 + x_2} + \frac{V_4 x_4}{K_4 + x_4} - x_3 \left( \frac{V_2}{K_2 + x_3} + \frac{V_3}{K_3 + x_3} \right) \\ \dot{x}_4 = \frac{V_3 x_3}{K_3 + x_3} - x_4 \left( \frac{V_4}{K_4 + x_4} + k_1 + \frac{v_d}{K_d + x_4} \right) + k_2 x_5 \\ \dot{x}_5 = k_1 x_4 - k_2 x_5 \end{cases} \tag{3}$$

proposed by Goldbeter for the circadian oscillation in the *Drosophila* period protein<sup>29</sup>. This is a five-dimensional rational model which produces a limit cycle for the parameter values initially reported<sup>29</sup>. This system is interesting in the sense that its complexity already presents a big challenge from the analytical point of view. The corresponding symbolic Jacobian matrix reads as

$$\tilde{\mathcal{J}}_{5D} = \begin{pmatrix} \bar{1} & 0 & 0 & 0 & \bar{1} \\ \bar{1} & \bar{1} & \bar{1} & 0 & 0 \\ 0 & \bar{1} & \bar{1} & \bar{1} & 0 \\ 0 & 0 & \bar{1} & \bar{1} & 1 \\ 0 & 0 & 0 & 1 & \bar{1} \end{pmatrix}. \tag{4}$$

The symbolic observability coefficients corresponding to a univariate measurement  $s = x_i$  are  $\eta_{x_1^5} = 0.17$ ,  $\eta_{x_2^5} = 0.08$ ,  $\eta_{x_3^5} = 0.02$ ,  $\eta_{x_4^5} = 0.09$  and  $\eta_{x_5^5} = 0.30$ , where the notation  $x_i^5$  refers to the vector  $(x_i, \dot{x}_i, \ddot{x}_i, \dddot{x}_i, \dots)$  whose observability to span the state space of the original system is estimated. According to the observability coefficient values, the ranking of the variables providing better observability is

$$x_5 \triangleright x_1 \triangleright x_4 \triangleright x_2 \triangleright x_3.$$

This is in a rather good agreement with the analytical determinants

$$\left\{ \begin{aligned} \Delta_{x_1^5} &= \frac{256V_s^4 K_I^{16} k_1^3 V_3^2 K_3^2 V_1 K_1}{(K_I^4 + x_5^4)^8 (K_3 + x_3)^4 (K_1 + x_2)^2} x_5^{12}; \\ \Delta_{x_2^5}, \Delta_{x_3^5} \text{ and } \Delta_{x_4^5} &\text{ where complexity exceeds our computational abilities;} \\ \Delta_{x_5^5} &= -\frac{k_1^4 V_3^3 K_3^3 V_1^2 K_1^2 k_s}{(K_3 + x_3)^6 (K_1 + x_2)^4}, \end{aligned} \right. \tag{5}$$

since the simpler determinant (with singularity of the 10th degree) is obtained for variable  $x_5$  providing the best observability, then variable  $x_1$  is associated with a determinant with a singularity of the 14th degree, and the three variables  $x_2$ ,  $x_3$  and  $x_4$  providing the poorest observability are associated with determinants too complicated to be computed with MAPLE®. The number (125) of all possible combinations of dimension 5 is still sufficiently small for allowing a systematic computation of the corresponding symbolic observability coefficients  $\eta_s$ . Prior to carry out those computations, we conducted our *a priori* analysis to target the candidate variables to be discarded. The linear out-strengths are  $\sigma_{out}^{lin}(1) = \sigma_{out}^{lin}(4) = \sigma_{out}^{lin}(5) = 1$ , the two others being null. Among the off-diagonal terms of the symbolic Jacobian matrix which are equal to 1, we have  $J_{45} = J_{54} = 1$ , meaning that at least one of the two variables  $x_4$  and  $x_5$  has to be measured. Therefore, this suggests that the sets with the minimum number of sensors providing full observability comprise at least three variables, either  $(x_2, x_3, x_4)$  or  $(x_2, x_3, x_5)$ . In order to have a five dimensional space, these two sets have to be completed with two Lie derivatives. It turns out that only two combinations yielded full observability:  $(\eta_{x_2^2 x_3^2 x_4^2} = 1)$  and  $(\eta_{x_2^2 x_3^2 x_5^2} = 1)$  where the derivatives of  $x_2$  and  $x_4$  and of  $x_2$  and  $x_5$  are, respectively, the ones completing each set of sensors. These results are algebraically confirmed by the constant determinants of the observability matrix (or equivalently of the Jacobian matrix of the coordinate transformation between the original state space and the reconstructed space<sup>30</sup>) and equal to  $\Delta_{x_2^2 x_3^2 x_4^2} = -k_s k_2$  and  $\Delta_{x_2^2 x_3^2 x_5^2} = k_s k_1$ , respectively. The relevance of correctly choosing the derivatives is exemplified by replacing in  $s = (x_2, \dot{x}_2, x_3, x_5, \dot{x}_5)$  the derivative of the fifth variable by the derivative of the third one: the coordinate transformation  $\Phi_{x_2^2 x_3^2 x_5^2}$  yields a symbolic observability coefficient  $\eta_{x_2^2 x_3^2 x_5^2} = 0.70$ , thus reflecting a significant lack of observability. This is further supported by the corresponding determinant

$$\Delta_{x_2^2 x_3^2 x_5^2} = -\frac{k_s V_4 K_4}{(K_4 + x_4)^2} \tag{6}$$

which is now no longer constant as it depends on variable  $x_4$ . There is thus a singular observability manifold. On the other hand, if the derivative of the second variable is substituted with the derivative of the third one, the Jacobian matrix  $J$  of such a transformation  $\Phi_{x_2 x_3^2 x_5^2}$  can be rank deficient with a null determinant for some domain of the original state space.

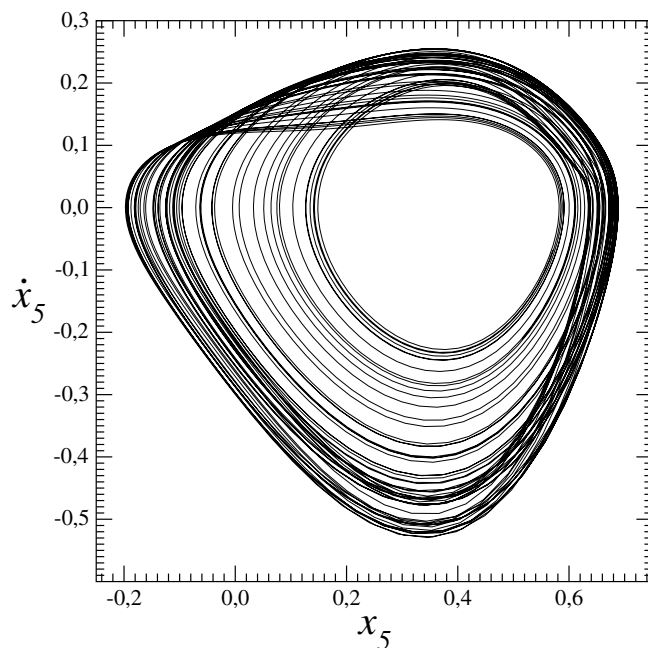
Finally, we wanted to assess the observability of the couple of variables  $\{x_1, x_5\}$ , identified as sensors of the system (3) in a previous work<sup>16</sup>. Surprisingly enough, we got  $\eta_{x_1 x_5^4} = 0.48$  and  $\eta_{x_1^2 x_3^3} = \eta_{x_1^3 x_5^2} = \eta_{x_1^4 x_5} = 0$  for all the possible 5 dimensional vectors constructed with those two variables. These symbolic observability coefficients are also in agreement with the determinants of the Jacobian matrices of the corresponding transformations,

$$\Delta_{x_1 x_5^4} = k_1^3 \frac{V_3^2 K_3^2 V_1 K_1}{(K_3 + x_3)^4 (K_1 + x_2)^2} \tag{7}$$

whose rational dependence on variables  $x_2$  and  $x_3$  defines a singular observability manifold associated with the transformation  $\Phi_{x_1 x_5^4}$ , and the other three determinants  $\Delta_{x_1^2 x_3^3} = \Delta_{x_1^3 x_5^2} = \Delta_{x_1^4 x_5} = 0$ , characterizing a rank deficient observability matrix  $\mathcal{O}$ . This is therefore a first evidence that our method to reduce the number of sensor variables correctly assesses the observability of this rather complex reaction network.

**A 9D system for the Rayleigh-Bénard convection.** Let us consider now a nine-dimensional system describing the dynamics of three-dimensional fluid cells with a square platform in a Rayleigh-Bénard convection<sup>31</sup>. It was obtained by applying a triple second-order Fourier series ansatz to the governing hydrodynamic equations. The equations read as

$$\left\{ \begin{aligned} \dot{x}_1 &= -\sigma b_1 x_1 - x_2 x_4 + b_4 x_4^2 + b_3 x_3 x_5 - \sigma b_2 x_7 \\ \dot{x}_2 &= -\sigma x_2 + x_1 x_4 - x_2 x_5 + x_4 x_5 - \frac{\sigma x_9}{2} \\ \dot{x}_3 &= -\sigma b_1 x_3 + x_2 x_4 - b_4 x_2^2 - b_3 x_1 x_5 + \sigma b_2 x_8 \\ \dot{x}_4 &= -\sigma x_4 - x_2 x_3 - x_2 x_5 + x_4 x_5 + \frac{\sigma x_9}{2} \\ \dot{x}_5 &= -\sigma b_5 x_5 + \frac{x_2^2}{2} - \frac{x_4^2}{2} \\ \dot{x}_6 &= -b_6 x_6 + x_2 x_9 - x_4 x_9 \\ \dot{x}_7 &= -b_1 x_7 - R x_1 + 2x_5 x_8 - x_4 x_9 \\ \dot{x}_8 &= -b_1 x_8 + R x_3 - 2x_5 x_7 + x_2 x_9 \\ \dot{x}_9 &= -x_9 - R x_2 + R x_4 - 2x_2 x_6 + 2x_4 x_6 + x_4 x_7 - x_2 x_8 \end{aligned} \right. \tag{8}$$



**Figure 1.** A chaotic attractor produced by the 9-dimensional dynamical network (8). Parameter values:  $\sigma = 0.5$  and  $R = 14.22$ , and rest of parameter values are listed in (10).

$m$	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$	$x_9$	$\eta$
8	1	2	1	1	1	1	1	1	—	1.00
8	1	1	2	1	1	1	1	—	1	1.00
8	1	1	1	2	1	1	1	—	1	1.00
8	1	1	1	2	1	1	1	1	—	1.00
8	2	1	1	1	1	1	—	1	1	1.00
8	1	1	—	1	1	1	1	2	1	1.00
8	1	1	—	1	1	1	2	1	1	1.00
8	—	1	1	1	1	1	2	1	1	1.00

**Table 1.** All possible subsets with  $m = 8$  measured variables and one Lie derivative (of the variable for which “2” is reported) providing a full observability of the state space associated with the 9-dimensional system (8). Those variables not affecting the full observability when not measured are highlighted in bold face.

where  $R$  is the reduced Rayleigh number and parameters  $b_i$  ( $i = 1, \dots, 6$ ) define the geometry of the square cell<sup>31</sup>. This system is equivariant under the rotation (see page 506 in Gilmore and Letellier’s book<sup>32</sup>)

$$\Gamma = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{9}$$

In fact, eight variables are symmetry-related by pairs, namely  $(x_1 - x_3)$ ,  $(x_2 - x_4)$ ,  $(x_5 - x_9)$ , and  $(x_7 - x_8)$ , and variable  $x_6$  is the single one left invariant under the symmetry (9). Up to four co-existing chaotic attractors were observed in this system<sup>31</sup>. An example of one of those chaotic attractors is shown in Fig. 1 for the  $b_i$ -values

$$\begin{cases} b_1 = 4 \frac{1 + a^2}{1 + 2a^2} & b_2 = \frac{1 + 2a^2}{2(1 + a^2)} & b_3 = 2 \frac{1 - a^2}{1 + a^2} \\ b_4 = \frac{a^2}{1 + a^2} & b_5 = \frac{8a^2}{1 + 2a^2} & b_6 = \frac{4}{1 + 2a^2}. \end{cases} \tag{10}$$

And  $a = 0.5$ . This system is irreducible in the sense that it cannot be split in lower-dimensional independent systems.

This system is an interesting example because it constitutes a highly connected reaction network for which a graphical approach as the one developed by Liu and coworkers<sup>17</sup> leads to only measure one of its nine variables to estimate its states (see the Supplementary Section S1). The variable that least influences the others (or equivalently, the one least “seen” by the rest) is  $x_6$  ( $\sigma_{\text{out}}(6) = 2$ ) since it only affects nonlinearly the derivative of  $x_9$ . From a symmetry point of view, variable  $x_6$  must be measured to recover the right symmetry property: without this variable, the reconstructed state space would be necessarily associated with an inversion symmetry (a symmetry the original system does not have).

The symbolic Jacobian matrix of system (8) is

$$\tilde{\mathcal{J}}_{9D} = \begin{pmatrix} 1 & \bar{1} & \bar{1} & \bar{1} & \bar{1} & 0 & 1 & 0 & 0 \\ \bar{1} & \bar{1} & 0 & \bar{1} & \bar{1} & 0 & 0 & 0 & 1 \\ \bar{1} & \bar{1} & 1 & \bar{1} & \bar{1} & 0 & 0 & 1 & 0 \\ 0 & \bar{1} & \bar{1} & \bar{1} & \bar{1} & 0 & 0 & 0 & 1 \\ 0 & \bar{1} & 0 & \bar{1} & 1 & 0 & 0 & 0 & 0 \\ 0 & \bar{1} & 0 & \bar{1} & 0 & 0 & 0 & 0 & \bar{1} \\ 1 & 0 & 0 & \bar{1} & \bar{1} & 1 & 1 & \bar{1} & \bar{1} \\ 0 & \bar{1} & 1 & 0 & 0 & \bar{1} & \bar{1} & \bar{1} & \bar{1} \\ 0 & \bar{1} & 0 & \bar{1} & \bar{1} & \bar{1} & \bar{1} & \bar{1} & 1 \end{pmatrix} \quad (11)$$

For this 9-dimensional system, there are 24309 possible combinations of variables and their derivatives candidates for providing full observability (see the Supplementary Section S2). Dealing with all these potential solutions is still affordable with our symbolic technique but it would take a rather long computational time (about 2 h). In order to reduce the number of combinations to test, we computed the linear out-strength  $\sigma_{\text{out}}^{\text{lin}}(i)$  of the 9 variables which are

$$\sigma_{\text{out}}^{\text{lin}}(2) = \sigma_{\text{out}}^{\text{lin}}(4) = \sigma_{\text{out}}^{\text{lin}}(5) = \sigma_{\text{out}}^{\text{lin}}(6) = 0$$

and

$$\sigma_{\text{out}}^{\text{lin}}(1) = \sigma_{\text{out}}^{\text{lin}}(3) = \sigma_{\text{out}}^{\text{lin}}(7) = \sigma_{\text{out}}^{\text{lin}}(8) = \sigma_{\text{out}}^{\text{lin}}(9) = 1.$$

These values indicate that variables  $x_2, x_4, x_5, x_6$  are necessarily to be included in the list of variables to be measured for obtaining a full observability since none of them linearly affect the dynamics of the rest. The other five are candidate variables to be removed from the measurements. Their respective in-strengths coming from the non measured variables are:

$$\sigma_{\text{in}}^{\text{nm}}(1) = \sigma_{\text{in}}^{\text{nm}}(3) = 3, \sigma_{\text{in}}^{\text{nm}}(7) = \sigma_{\text{in}}^{\text{nm}}(8) = 5, \text{ and } \sigma_{\text{in}}^{\text{nm}}(9) = 4.$$

To determine whether all the candidate variables could be removed, we checked if there are pairs of exclusive variables, that is, when two candidate variables are linearly coupled each other (one being linearly “seen” by the other). Among the off-diagonal elements  $\tilde{J}_{ij}$  equal to 1, we have  $\tilde{J}_{17} = \tilde{J}_{71}, \tilde{J}_{29}, \tilde{J}_{38} = \tilde{J}_{83}$ , and  $\tilde{J}_{49}$ . Variables  $(x_1, x_7)$  and  $(x_3, x_8)$  thus form two pairs of mutually exclusive variables. Variable  $x_9$  is the single one not involved in an exclusive pair and can be removed from the set of measured variables. To decide which variable from each pair can be safely removed, we check which variables have the largest in-strength from the candidate variables to be non measured. The comparison returns that variables  $x_7$  and  $x_8$  are the ones to be removed since  $\sigma_{\text{in}}^{\text{nl}}(7) = \sigma_{\text{in}}^{\text{nm}}(8) = 5 > \sigma_{\text{in}}^{\text{nm}}(1) = \sigma_{\text{in}}^{\text{nm}}(3) = 3$ .

The first test to assess the accuracy in selecting the minimal set of variables providing the highest observability consists in systematically investigating those combinations where all the variables are measured except one. The symbolic observability coefficients are reported in Table 1. In all cases providing full observability with just a single variable not being considered, the discarded measure matches one of the candidate variables  $x_1, x_3, x_7, x_8$  and  $x_9$  (marked in bold face in the table) confirming our preselection analysis.

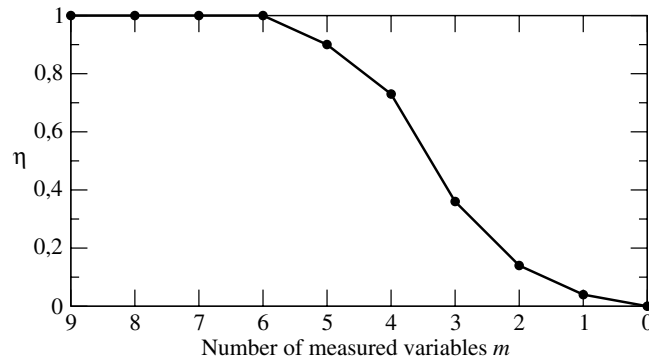
Our systematic computation of the symbolic coefficients allows us to quantify the number of times  $N_i(\eta)$  the variable  $x_i$  is not part of an embedding providing a given observability value  $\eta$ . In particular, the values of  $N_i(\eta = 1.0)$  for the variables potentially candidate to be non measured ( $N_1(1.0) = 4, N_3(1.0) = 2, N_7(1.0) = 7, N_8(1.0) = 5$ , and  $N_9(1.0) = 9$ , see the first part of Table 2), support our initial choice for not measuring  $x_7, x_8$ , and  $x_9$  but measuring  $x_1$  and  $x_3$ , since  $x_7, x_8$  and  $x_9$  seem to be less essential for providing full observability. Consequently, as long as full observability is required, our two network-based criteria correctly identify those variables whose absence from the set of sensors does not affect the full observability of the system. Indeed, when the minimal number of variables, that is,  $m = 6$ , is measured, the two possible combinations providing a full observability correspond to a space reconstructed from variables  $x_1, x_2, x_3, x_4, x_5$  and  $x_6$  (see the Supplementary Table S1). We therefore correctly assessed the best variables to measure for getting full observability with the minimum of variables. Of course, it is also possible to get full observability by measuring more than 6 variables. In that case, we searched for them among the 8 preselected variables using the linear out-strength. From the 354 possible combinations, we obtained 6 combinations using 7 measured variables and 2 with 8 measured variables. Performing a full blind search, with no preselection, from a total number of 1080 combinations with 7 or 8 measured variables, we found 2 and 6 additional combinations providing full observability, respectively. All of them are reported in the Supplementary Table S1.

$m$	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$	$x_9$	$\eta$
6	2	2	2	1	1	1	—	—	—	1.00
6	2	2	2	1	1	1	—	—	—	1.00
7	1	2	2	1	1	1	1	—	—	1.00
7	1	2	—	1	1	1	1	2	—	1.00
7	—	1	1	2	1	1	2	1	—	1.00
7	—	1	1	2	1	1	2	1	—	1.00
7	—	1	1	1	1	1	2	1	1	1.00
7	—	1	—	1	1	1	2	2	1	1.00
7	2	2	1	1	1	1	—	1	—	1.00
7	2	1	2	2	1	1	—	—	—	1.00
7	2	1	2	1	1	1	—	—	1	1.00
7	2	1	1	2	1	1	—	1	—	1.00
7	2	1	1	1	1	1	—	1	1	1.00
5	2	3	2	1	1	—	—	—	—	0.90
5	2	1	2	3	1	—	—	—	—	0.90
6	3	2	1	1	1	1	—	—	—	0.90
6	3	1	2	1	1	1	—	—	—	0.90
6	3	1	1	2	1	1	—	—	—	0.90
6	2	3	1	1	1	1	—	—	—	0.90
6	2	3	1	1	1	—	—	1	—	0.90
6	2	1	3	1	1	1	—	—	—	0.90
6	2	1	1	3	1	1	—	—	—	0.90
6	2	1	1	3	1	—	—	1	—	0.90
6	1	3	2	1	1	1	—	—	—	0.90
6	1	3	2	1	1	—	1	—	—	0.90
6	1	3	—	1	1	—	1	1	—	0.90
6	1	2	3	1	1	1	—	—	—	0.90
6	1	2	—	1	1	—	1	3	—	0.90
6	1	1	3	2	1	1	—	—	—	0.90
6	1	1	2	3	1	1	—	—	—	0.90
6	1	1	2	3	1	—	1	—	—	0.90
6	—	1	1	3	1	—	2	1	—	0.90
6	—	1	1	2	1	—	3	1	—	0.90
6	—	1	—	1	1	—	3	2	1	0.90
6	—	1	—	1	1	—	2	3	1	0.90
6	—	1	—	1	1	—	2	3	1	0.90
5	4	2	1	1	1	—	—	—	—	0.80
5	4	1	2	1	1	—	—	—	—	0.80
5	4	1	1	2	1	—	—	—	—	0.80
5	2	1	4	1	1	—	—	—	—	0.80
5	1	2	4	1	1	—	—	—	—	0.80
5	1	1	4	2	1	—	—	—	—	0.80
6	4	1	1	1	1	—	—	1	—	0.80
6	4	1	1	1	1	—	—	—	1	0.80
6	1	1	4	1	1	—	1	—	—	0.80
6	1	1	4	1	1	—	—	—	1	0.80

**Table 2.** List of the different possible combinations of measured variables and their Lie derivative orders providing a symbolic observability coefficient  $\eta \geq 0.75$  of the state space associated with the 9-dimensional system (8).

We also investigated the combinations with less than 6 measured variables and providing the largest symbolic observability coefficients whose dependency on  $m$  is shown in Fig. 2. For  $m = 5$  ( $m = 4, 3, 2$  and  $1$ ) there are 4 (2, 6, 4, and 4, respectively) combinations with  $\eta = 0.90$  ( $\eta = 0.73, 0.36, 0.14, 0.04$ , respectively). All those combinations are made up of the six preselected variables identified by solely using the symbolic Jacobian matrix and  $\sigma_{out}^{lin}$  and  $\sigma_{in}^{nm}$  to rank them.

The non-preselected variables can be involved in reconstructed vectors when a good but not a full observability is desired or when  $m > 6$  (a good observability is considered when  $\eta > 0.75$  as reported in a previous work<sup>33</sup>).



**Figure 2.** Largest symbolic observability coefficient  $\eta$  versus the number  $m$  of measured variables for the 9D Rayleigh-Bénard model (8).

For instance, this is exemplified in the middle and lower part of Table 2 with a systematic computation of the symbolic observability coefficients  $\eta$  for embeddings built from 5 or 6 measured variables. In this case, we observe that a very good observability ( $\eta = 0.90$ ) can be obtained with only 5 variables being measured, namely  $(x_1, x_2, x_3, x_4, x_5)$ . The fact that variable  $x_6$  is not included, prevents from a full observability and, in particular, its absence induces a lack of symmetry, as previously discussed. Again, by looking at the distribution of  $N_i(0.9)$  we have:  $N_1(0.9) = 5$ ,  $N_3(0.9) = 5$ ,  $N_7(0.9) = 14$ ,  $N_8(0.9) = 14$ , and  $N_9(0.9) = 20$ . Therefore, for this level of observability ( $\eta = 0.9$ ) the last three variables,  $x_7, x_8$ , and  $x_9$  can be again chosen to be removed from the set of observations. If we accept a slightly lower observability coefficient ( $\eta = 0.80$ ), other possibilities emerge in which variable  $x_6$  is systematically removed from the set of measured variables (last part of Table 2). Our two criteria are thus very efficient to detect those variables not really impacting the access to a full observability measure, but discard some possibilities offering a good (but not full until  $m \leq 6$ ) observability with severe consequences on the symmetry properties of the reconstructed attractor.

**A 13D model for DNA replication.** A third and even more challenging case is now considered, a 13-dimensional model for the DNA replication in fission yeast. Fission yeast cells are carrying two mutant genes,  $wee1^-$  and  $cdc25^{Op}$ , which initiate mitosis in eukaryotic cells before the end of their DNA replication. A second feature is that DNA synthesis can be restarted without intervening mitoses. Novak and Tyson proposed a model for cell cycle in fission yeast taking into account these two properties in *Schizosaccharomyces pombe*<sup>34</sup>. The underlying mechanisms are described by the set of thirteen differential equations

$$\begin{cases}
 \dot{x}_1 = k_1 - (k_2 + k_{wee} + k_7 x_2) x_1 + k_{25} x_8 + (k_{7r} + k_4) x_4 \\
 \dot{x}_2 = k_3 - k_4 x_2 - \frac{k_p x_2 (x_1 + \beta x_8 + \alpha x_3) m}{K_{mp} + x_2} - k_7 x_2 (x_1 + x_8) \\
 \quad - k_8 x_2 x_3 + (k_{8r} + k_6) x_9 + (k_{7r} + k_2 + k_{2'}) (x_4 + x_{10}) \\
 \dot{x}_3 = k_5 - (k_6 + k_8 x_2) x_3 + (k_{8r} + k_4) x_9 \\
 \dot{x}_4 = k_7 x_2 x_1 - (k_{7r} + k_4 + k_2 + k_{2'}) x_4 \\
 \dot{x}_5 = \frac{k_i (x_1 + \beta x_8) (1 - x_5)}{K_{mi} + 1 - x_5} - \frac{k_{ir} x_5}{K_{mir} + x_5} \\
 \dot{x}_6 = \frac{k_{u2} (x_1 + \beta x_8) (1 - x_6)}{K_{mu2} + 1 - x_6} - \frac{k_{ur2} x_6}{K_{mur2} + x_6} \\
 \dot{x}_7 = \frac{k_{wr} (1 - x_7)}{K_{mwr} + 1 - x_7} - \frac{k_w (x_1 + \beta x_8) x_7}{K_{mw} + x_7} \\
 \dot{x}_8 = k_{wee} x_1 - (k_{25} + k_2 + k_7 x_2) x_8 + (k_{7r} + k_4) x_{10} \\
 \dot{x}_9 = k_8 x_2 x_3 - (k_{8r} + k_4 + k_6) x_9 \\
 \dot{x}_{10} = k_7 x_2 x_8 - (k_{7r} + k_4 + k_2 + k_{2'}) x_{10} \\
 \dot{x}_{11} = \frac{k_u x_5 (1 - x_{11})}{K_{mu} + 1 - x_{11}} - \frac{k_{ur} x_{11}}{K_{mur} + x_{11}} \\
 \dot{x}_{12} = \frac{k_c (x_1 + \beta x_8) (1 - x_{12})}{K_{mc} + 1 - x_{12}} - \frac{k_{cr} x_{12}}{K_{mcr} + x_{12}} \\
 \dot{m} = \mu m
 \end{cases} \quad (12)$$

where  $x_1 = \text{G2K}$ ,  $x_2 = \text{R}$ ,  $x_3 = \text{G1K}$ ,  $x_4 = \text{G2R}$ ,  $x_5 = \text{IE}$ ,  $x_6 = \text{Ube2}$ ,  $x_7 = \text{Wee1}$ ,  $x_8 = \text{PG2}$ ,  $x_9 = \text{G1R}$ ,  $x_{10} = \text{PG2R}$ ,  $x_{11} = \text{Ube}$ , and  $x_{12} = \text{Cdc25}$ , are concentration variables, see Novak and Tyson<sup>34</sup> for a more detailed explanation of the meaning of these variables and values of the rate constants.



This 13-dimensional rational model is characterized by the symbolic Jacobian matrix

$$\tilde{J}_{13D} = \begin{pmatrix} \bar{1} & \bar{1} & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \bar{1} & \bar{1} & \bar{1} & 1 & 0 & 0 & 0 & \bar{1} & 1 & 1 & 0 & 0 & \bar{1} \\ 0 & \bar{1} & \bar{1} & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \bar{1} & \bar{1} & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \bar{1} & 0 & 0 & 0 & \bar{1} & 0 & 0 & \bar{1} & 0 & 0 & 0 & 0 & 0 \\ \bar{1} & 0 & 0 & 0 & 0 & \bar{1} & 0 & \bar{1} & 0 & 0 & 0 & 0 & 0 \\ \bar{1} & 0 & 0 & 0 & 0 & 0 & \bar{1} & \bar{1} & 0 & 0 & 0 & 0 & 0 \\ 1 & \bar{1} & 0 & 0 & 0 & 0 & 0 & \bar{1} & 0 & 1 & 0 & 0 & 0 \\ 0 & \bar{1} & \bar{1} & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & \bar{1} & 0 & 0 & 0 & 0 & 0 & \bar{1} & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \bar{1} & 0 & 0 & 0 & 0 & 0 & \bar{1} & 0 & 0 \\ \bar{1} & 0 & 0 & 0 & 0 & 0 & 0 & \bar{1} & 0 & 0 & 0 & \bar{1} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{13}$$

According to the linear out-strengths  $\sigma_{out}^{lin}$ , we have five candidate variables eligible to be excluded from the observations since having not null  $\sigma_{out}^{lin}$  values:

$$\sigma_{out}^{lin}(1) = \sigma_{out}^{lin}(8) = 1 < \sigma_{out}^{lin}(4) = \sigma_{out}^{lin}(9) = \sigma_{out}^{lin}(10) = 2.$$

The linear off-diagonal elements ( $\tilde{J}_{ij} = 1$ ) are  $\tilde{J}_{14}, \tilde{J}_{18}, \tilde{J}_{24}, \tilde{J}_{29}, \tilde{J}_{2,10}, \tilde{J}_{39}, \tilde{J}_{81}$ , and  $\tilde{J}_{8,10}$ . There is therefore a single pair of exclusive candidate variables, that is, variables  $x_1$  and  $x_8$  ( $\tilde{J}_{18} = \tilde{J}_{81} = 1$ ). Variables  $x_4, x_9$ , and  $x_{10}$  can therefore be safely removed from the set of measured variables and  $x_1$  and  $x_8$  can not be simultaneously removed. The in-strength  $\sigma_{in}^{nm}$  from the set of potentially non-measured variables is equal to 2 for all the candidate variables except for  $x_9$  which is  $\sigma_{in}^{nm}(9) = 0$ . Our criteria does not allow us this time to resolve the uncertainty between  $x_1$  and  $x_8$ .

Following the same procedure as with the two previous examples, we performed our *a priori* analysis by systematically computing the symbolic observability coefficients when a single variable is removed and collecting only those combinations providing either full or null observability (see Table 3). As expected, when one of the variables  $x_1, x_4, x_8, x_9$  or  $x_{10}$  is not included in the observation set of 12 variables plus one derivative, the observability is full. On the contrary, if one of the variables  $x_6, x_7, x_{11}$  and  $x_{12}$  is removed, the symbolic observability coefficient drops to zero for any possible choice of the first derivative. These variables are therefore essential and need to be measured. This is due to the fact that these variables have no out-connection other than to themselves as shown in the corresponding columns of the symbolic Jacobian matrix in Eq. (13).

Let us now validate whether it is possible to retrieve a full observability when either the set  $\{x_1, x_4, x_9, x_{10}\}$  or  $\{x_4, x_8, x_9, x_{10}\}$  are removed from the list of variables to measure. This was performed by systematically computing the symbolic observability coefficients for all the combinations reconstructing a 13-dimensional space without taking into account those two sets of variables. We found that for this DNA model, there are not too many possibilities to reconstruct a space providing full observability of the original state space (see the Supplementary Table S1). For instance, when removing two of them,  $x_8$  and  $x_9$  the reconstructed state vector  $(x_1, \dot{x}_1, x_2, \dot{x}_2, x_3, \dot{x}_3, x_4, x_5, x_6, x_7, x_{11}, x_{12}, x_{13})$  is the only one providing full observability. If we discard three variables ( $x_8, x_9$ , and  $x_{10}$ ), there are two combinations allowing for a full observability embedding:  $(x_1, \dot{x}_1, x_2, \dot{x}_2, x_3, x_4, x_5, x_6, x_7, x_{10}, x_{11}, x_{12}, x_{13})$  and  $(x_1, \dot{x}_1, x_2, x_3, \dot{x}_3, x_4, x_5, x_6, x_7, x_{10}, x_{11}, x_{12}, x_{13})$ . A systematic analysis of the symbolic observability coefficients as the number of variables are removed from the set of observations indicate that the coefficient already drops to 0.93 when  $m = 9$  variables are measured (see Fig. 3). As the estimated threshold for an optimal observability is 0.75<sup>33</sup>, it is worthless to investigate sets of size smaller than  $m = 7$ .

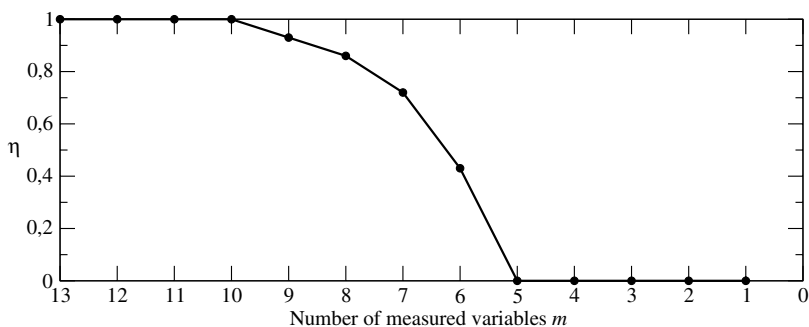
To actually check the results accounted for the symbolic observability coefficients  $\eta_s$ , we computed all the determinants  $\text{Det } \mathcal{O}_s$  corresponding to  $\eta_s = 1$  (results are reported in the Supplementary Table S1). In the 14 cases for which the symbolic observability coefficients are equal to one, the determinant  $\text{Det } \mathcal{O}_s$  was always nonzero (for the whole state space): our technique always correctly identify reconstructed vectors providing full observability of the original space. As another example, as shown in Fig. 3, full observability is never achieved for  $m = 9$  and the largest symbolic observability coefficient is 0.93, which still provides a good observability. By using the reconstructed space  $\mathbb{R}^{13}(x_1, \dot{x}_1, x_2, \dot{x}_2, x_3, \dot{x}_3, x_5, x_6, x_7, x_{10}, x_{11}, x_{12}, \dot{x}_{12}, x_{13})$ , (one of the cases reported in Table 4) the determinant

$$\text{Det } \mathcal{O} = -(k_{7r} + k_2 + k_{2p})(k_{8r} + k_4)(k_{7r} + k_4) \frac{k_c \beta (x_{12} - 1)}{K_{mc} + 1 - x_{12}} \tag{14}$$

is zero for  $x_{12} = 1$ , a singular observability manifold of first order, explaining why the observability coefficient is no longer equal to 1 but close to it. To further show how the observability coefficient decreases when  $\text{Det } \mathcal{O}_s$  vanishes for a singularity of higher degree<sup>22</sup>, we computed the determinant

Non-measured	Derivative retained	$\eta$
$x_1$	$\dot{x}_8$	1.00
$x_4$	$\dot{x}_1$ or $\dot{x}_2$	1.00
$x_8$	$\dot{x}_1$	1.00
$x_9$	$\dot{x}_2$ or $\dot{x}_3$	1.00
$x_{10}$	$\dot{x}_2$ or $\dot{x}_8$	1.00
$x_6$	$\forall \dot{x}_i$	0.00
$x_7$	$\forall \dot{x}_i$	0.00
$x_{11}$	$\forall \dot{x}_i$	0.00
$x_{12}$	$\forall \dot{x}_i$	0.00

**Table 3.** Symbolic observability coefficients when twelve (out of thirteen) variables of the DNA model (12) are measured. The derivative used for reconstructing a 13-dimensional state space is also reported.



**Figure 3.** Largest symbolic observability coefficient  $\eta$  versus the number  $m$  of measured variables for the DNA model (12).

$$\text{Det } \mathcal{O}_s = -(k_{7r} + k_2 + k_{2p})(k_{8r} + k_4)(k_{7r} + k_4) \frac{k_u k_c \beta (x_{11} - 1)(x_{12} - 1)}{(K_{mc} + 1 - x_{12})(K_\mu + 1 - x_{11})} \tag{15}$$

associated with the reconstructed space  $\mathbb{R}^{13}(x_1, \dot{x}_1, x_2, \dot{x}_2, x_3, \dot{x}_3, x_6, x_7, x_{10}, x_{11}, \dot{x}_{11}, x_{12}, \dot{x}_{12}, x_{13})$  providing a slightly smaller observability ( $\eta = 0.86$ ) in agreement with the singular observability manifold (15) of second order, defined by  $\text{Det } \mathcal{O}_s = 0$ , that is, by  $(x_{11} - 1)(x_{12} - 1) = 0$ . Due to a too large complexity, it was not possible to analytically compute the observability matrix when a single variable is measured.

As detailed in the Supplementary Section S1, Liu and coworkers’ graphical technique shows that by measuring the four variables  $x_6, x_7, x_{11}$ , and  $x_{12}$  it is possible to estimate the states of the system. Clearly, our results are in strong disagreement. At this point, it is relevant to explain why our results are so different from those reported in previous works<sup>16,17</sup>. The first reason for the discrepancy is that Liu’s algorithm uses a linear theory, only taking into account whether the  $i$ th variable participates or not in the differential equation of variable  $j$  and not concerned on how is that dependence. The latter is just equivalent to use a symbolic Jacobian matrix equal to

$$\tilde{\mathcal{J}}_{13D}^{\text{lin}} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{16}$$

and not the one defined in Eq. (13). Despite there are more than 2100 combinations with 6 or 7 measured variables resulting in full observability, there is none with a single variable. However, with a linear approach, it is still possible to show that the combinations providing full observability correctly identify the variables which must necessarily be used (Table 5), that is, variables  $x_6, x_7, x_{11}$ , and  $x_{12}$ . Nevertheless, the observability is obviously

$m$	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$	$x_9$	$x_{10}$	$x_{11}$	$x_{12}$	$x_{13}$	$\eta$
9	2	2	2	—	1	1	1	—	—	—	1	2	1	0.93
9	2	2	2	—	1	1	2	—	—	—	1	1	1	0.93
9	2	2	2	—	1	2	1	—	—	—	1	1	1	0.93
9	2	2	2	—	2	1	1	—	—	—	1	1	1	0.93
8	2	2	2	—	—	1	1	—	—	—	2	2	1	0.86
8	2	2	2	—	—	1	2	—	—	—	2	1	1	0.86
8	2	2	2	—	—	2	1	—	—	—	2	1	1	0.86
9	—	2	2	—	1	1	1	2	—	—	1	2	1	0.93
9	—	2	2	—	1	1	2	2	—	—	1	1	1	0.93
9	—	2	2	—	1	2	1	2	—	—	1	1	1	0.93
9	—	2	2	—	2	1	1	2	—	—	1	1	1	0.93
8	—	2	2	—	—	1	1	2	—	—	2	2	1	0.86
8	—	2	2	—	—	1	2	2	—	—	2	1	1	0.86
8	—	2	2	—	—	2	1	2	—	—	2	1	1	0.86
7	—	1	2	—	—	1	1	3	—	—	2	3	—	0.72
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

**Table 4.** Symbolic observability coefficients for the DNA system. The first part corresponds to the case where variables  $\{x_4, x_8, x_9, x_{10}\}$  are not measured, the middle part to the case where variables  $\{x_1, x_4, x_9, x_{10}\}$  are not. Only the cases where the symbolic coefficient is non-zero and for which only a first derivative is used (to avoid too many possibilities) are reported.

overestimated (1 compared to 0.43 with a nonlinear theory) and, in addition, this approach does not allow to select what derivatives to use for spanning the reconstructed space. Assessing the observability of a network with a linear theory thus provides very poor and misleading results.

## Discussion

The observability of a complex system refers to the property of being able to infer its whole state space by measuring the dynamics of a limited set of its variables. Determining the conditions that guarantee the full observability of a system involves testing a number of possibilities that increases exponentially with the dimension of that system and, for each case, it is required to compute the determinant of the observability matrix defining the singular observability manifold, that is, the subset of the state space that cannot be observed from the measurement<sup>25</sup>. It was shown in one of our previous works<sup>2</sup> that for a five-dimensional rational system, the analytical computation of such a determinant may already exceed the capacity of softwares like Maple<sup>®</sup> or Mathematica<sup>®</sup>. Therefore, alternative approaches to investigate large complex systems are needed. Those proposed for instance by Sedoglavic<sup>16</sup> or Liu and coworkers<sup>17</sup> remain yet unsatisfactory as discussed by Wang and coworkers<sup>35</sup>, mainly because they do not provide a method to select which Lie derivatives accompany the measured variables and, more importantly, they do not consider a nonlinear observability theory appropriate to deal with nonlinear systems, nonlinearities occurring in the node dynamics or nonlinearly coupled units.

Actually, the treatment proposed by Sedoglavic<sup>16</sup> is only probabilistic and tests local observability, not the global one. On the other hand, the graphical approach developed by Liu and co-workers<sup>17</sup> is based on a linear description of the system which can only lead, by definition, to approximated results since, as previously discussed, the lack of observability mainly originates in the location (in the fluence graph) of nonlinear terms. We here investigated the three systems considered by Liu and coworkers (see the Supplementary Section S1) and showed that, in contrast with our results, theirs are not in agreement with the analytical predictions. In our previous work<sup>2</sup>, we investigated the same five-dimensional rational system considered by Sedoglavic<sup>16</sup>. While in the latter reference, the algorithm developed by the author identifies the first variable as the one providing (in fact local) observability, it is only poorly the case when the symbolic algorithm developed in the former one is applied to the possible combinations using this variable, even combined with other variables. And what is even more questionable, it is that when  $x_1$  is combined with one of the four other variables,  $x_2, x_3, x_4$  and  $x_5$ , the largest symbolic observability coefficient is still very small, that is, 0.30, 0.18, 0.30, and 0.48, respectively.

We have shown how the efficiency of the algorithm initially proposed by Bianco-Martinez and coworkers<sup>2</sup> is improved by identifying the minimal set of measured variables providing full observability before any search for the corresponding Lie derivatives. The reduced sets of candidate variables capable of fully reconstructing large reaction networks was correctly determined by analyzing the way the variables interact, by only applying two simple criteria on the symbolic Jacobian matrix of the networked system. For the 13 DNA model, there are  $5.2 \cdot 10^6$  possible combinations to test (see the Supplementary Section S2 for the details), requiring more than 18 days of computations with a 2.5 GHz Intel Core i5 processor. With our preselection of variables, only 2870 combinations are needed to be tested lowering the computation time to about 4 min, that is, by a factor greater than 1800! These criteria reduce drastically the time spent for searching candidate variables, thus providing the grounds to observe natural and man-made complex systems.

In order to evaluate the reliability of our procedure, we computed a success rate defined as the number of times a symbolic observability coefficient equal to 1 actually corresponds to getting a constant analytical determinant of

$m$	$N$	$M_1$	$M_2$	$M_3$	$M_4$	$M_5$	$M_6$
6	230	9	87	111	62	0	<b>230</b>
7	1896	307	879	981	728	327	<b>1892</b>
$m$	$M_7$	$M_8$	$M_9$	$M_{10}$	$M_{11}$	$M_{12}$	$M_{13}$
6	<b>230</b>	9	110	63	<b>230</b>	<b>230</b>	9
7	<b>1892</b>	303	978	722	<b>1891</b>	<b>1891</b>	470

**Table 5.** Number  $N$  of combinations providing a full observability—according to a linear theory—when  $m$  variables are measured. The numbers  $M_i$  in which the  $i$ th variable is involved in a vector spanning the reconstructed state space providing a full observability are also reported. In bold, the four variables which are the most often involved.

the observability matrix. For the 42 combinations providing  $\eta = 1$ , the success rate is 100%. While we were able to identify and algebraically check all the resulting combinations for the 5D and 9D models, it was impossible for the 13D model due to the large amount of them. In the case of the 9D model, for which we obtained  $m_p = 6$  preselected variables, our procedure missed 2 out of 8 combinations with  $m = 7$ , and 4 out of 6 with  $m = 8$  (see the Supplementary Table S1). The missed combinations involve at least one variable which was not preselected and, consequently, not considered in our computations. When  $m = 6$ , some combinations are associated with a symbolic observability coefficient equal to 0.90: in that case, 36 out of 54 corresponding to this value of the symbolic observability coefficient were made up of the preselected variables. When  $m = 5 < m_p$ , 100% of the combinations (34) associated with the largest symbolic observability coefficient (0.90) involved the preselected variables. It is important to note that, when  $m \leq 6$ , all combinations providing the largest symbolic observability coefficient (see Fig. 2) are made up of the preselected variables (and are actually found). This means that the preselected variables are indeed the relevant ones for estimating the system states and, that all combinations using these variables are correctly identified. To the best of our knowledge, we have a single case for which the full observability was not detected by our procedure<sup>33</sup>: it corresponds to the rare case for which two nonlinear terms cancel each other in the computation of the determinant of the observability matrix.

Finally, as firstly reported by Parlitz and coworkers<sup>36</sup>, the observability of a system could be addressed by using delay coordinates. As shown by Gibson and coworkers<sup>37</sup>, delay coordinates are related to derivatives by a rotation and a rescaling. Consequently, any result valid for derivative coordinates (not affected by rotation and rescaling) holds for delay coordinates. The reduced sets of  $m$  measured variables ( $m < d$ ) are not dependent on the use of delay or derivative coordinates, only on the choice of the complementary coordinates to reconstruct a  $d$ -dimensional space. Therefore, the extension of the technique proposed in this work to networks of discrete time systems (discretization of continuous-time systems) and iterated maps seems to be rather straightforward according, for instance, to Sarachik and Kreindler<sup>38</sup> and to Nijmeijer<sup>39</sup>, respectively.

## Methods

**Introduction to observability theory.** Our framework to quantify the observability of a dynamical system is here introduced with some definitions. Let us consider a  $d$ -dimensional dynamical system represented by the state vector  $x \in \mathbb{R}^d$  whose components are given by

$$\dot{x}_i = f_i(x_1, x_2, x_3, \dots, x_d), \quad i = 1, 2, 3, \dots, d \quad (17)$$

where  $f_i$  is the  $i$ th component of the vector field  $f$ .

Let us introduce the vector  $s \in \mathbb{R}^m$  whose  $m$  components are the time series of measured variables given by the measurement function

$$s = h(x). \quad (18)$$

One of the formal ways to define the observability of a system is as follows<sup>40</sup>. We provide such a definition in the case where a single scalar time series is measured,  $s = h(x)$ , but a generalization to the case of  $m$  measured variables is straightforward. The dynamical system (17) is said to be *state observable* at time  $t_f$  if every initial state  $x(0)$  can be uniquely determined from the knowledge of a finite time series of the measured variable  $s(\tau)$ ,  $0 \leq \tau \leq t_f$ . In practice, it is possible to test whether the dynamical system (17) is observable through a measurement function by computing the rank of the observability matrix<sup>20</sup>, that is, the Jacobian matrix of the Lie derivatives of  $s$ . Differentiating  $s(t)$  yields

$$\dot{s}(t) = \frac{d}{dt}h(x) = \frac{\partial h}{\partial x} \dot{x} = \frac{\partial h}{\partial x} f(x) = \mathcal{L}_f h(x),$$

where  $\mathcal{L}_f h(x)$  is the Lie derivative of  $h$  along the vector field  $f$ . The  $j$ th order Lie derivative is given by

$$\mathcal{L}_f^j h(x) = \frac{\partial \mathcal{L}_f^{j-1} h(x)}{\partial x} f(x),$$

being the zero order Lie derivative the measured variable itself,  $\mathcal{L}_f^0 h(x) = h(x)$ . Therefore, the observability matrix  $\mathcal{Q}_s \in \mathbb{R}^{d \times d}$  can be written as

$$\mathcal{O}_s(\mathbf{x}) = \begin{bmatrix} dh(\mathbf{x}) \\ d\mathcal{L}_f h(\mathbf{x}) \\ \vdots \\ d\mathcal{L}_f^{d-1} h(\mathbf{x}) \end{bmatrix} \quad (19)$$

where  $d \equiv \frac{\partial}{\partial \mathbf{x}}$  and the dynamical system (17) is said to be state observable if and only if the observability matrix has full rank, that is,  $\text{rank}(\mathcal{O}_s) = d$ . Notice that, the full observability of a system is determined by the space spanned not only by the measured variables but also by their appropriate Lie derivatives<sup>1</sup>.

The observability matrix  $\mathcal{O}_s$  corresponds in fact to the Jacobian matrix of the change of coordinates  $\Phi_s: \mathbf{x} \rightarrow \mathbf{X}$  where  $\mathbf{X} \in \mathbb{R}^d$  is the reconstructed state vector from the  $m$  measured variables and their adequately chosen  $d - m$  Lie derivatives<sup>30</sup>. Let us make explicit the observability matrix  $\mathcal{O}_{x_i x_j}$  for the situation where two arbitrary variables  $x_i$  and  $x_j$  are measured directly, that is, when  $s = (h_1(x), h_2(x)) = (x_i, x_j)$ , and  $h_1$  and  $h_2$  are two measurement functions. In this case, the observability matrix reads as,

$$\mathcal{O}_{x_i x_j} = \begin{bmatrix} dh_1(\mathbf{x}) \\ d\mathcal{L}_f h_1(\mathbf{x}) \\ \vdots \\ d\mathcal{L}_f^k h_1(\mathbf{x}) \\ dh_2(\mathbf{x}) \\ d\mathcal{L}_f h_2(\mathbf{x}) \\ \vdots \\ d\mathcal{L}_f^l h_2(\mathbf{x}) \end{bmatrix} \quad (20)$$

where the order Lie derivatives  $k$  and  $l$  are such that  $k + l = d - 2$ , that is, there are  $d - 2 + 1$  possibilities for choosing  $k$  and  $l$ . According to the Takens theorem<sup>41</sup>, it is possible to increase the dimension up to  $2d_H + 1$  where  $d_H$  is ideally the Hausdorff dimension of the attractor to ensure a global diffeomorphism between the original state space and the reconstructed one, as long as the measurement function is generic. Showing that a measurement function is generic is not a trivial problem which is out of the scope of the present work. There is, therefore, no guarantee that the Takens theorem applies here. Moreover, our aim is to select the minimal set of measurements providing the best observability of the system. When a higher-dimensional reconstructed space is considered, this means that a global diffeomorphism perhaps may be obtained but it also means that the measurements provide information that is non-optimal and from which the analysis is most likely problematic and tricky<sup>42,43</sup>. Consequently, investigating higher-dimensional reconstructed spaces has a rather limited interest in the present context.

The fact the system is fully observable from the two measured variables considered in the matrix (20) depends also on the particular choice of the pair  $(k, l)$ , the numbers of successive derivatives computed from  $x_i$  and  $x_j$ , respectively. Therefore, it is crucial to specify how the measured variables and their derivatives are used to reconstruct the state space. An approach—as the ones developed in other works<sup>16–18</sup>—missing this necessary condition cannot indeed properly address the problem of full (or even good) observability.

**Symbolic observability formalism.** The procedure to calculate the symbolic observability coefficients is implemented in four steps as follows:

- i) *Construction of the symbolic Jacobian matrix ( $\tilde{\mathcal{J}}$ ).* The Jacobian matrix  $\mathcal{J}$ , composed of elements  $J_{ij}$  of the system (17) is transformed into a symbolic Jacobian matrix  $\tilde{\mathcal{J}}$  by replacing each linear element  $J_{ij}$  by 1, each non-linear polynomial element  $J_{ij}$  by  $\tilde{J}_{ij}$  and each rational element  $J_{ij}$  by  $\tilde{J}_{ij}^{-1}$  when the  $j$ th variable is present in the denominator or by 1 otherwise. This is more or less equivalent to the so-called influence (or flunce) diagram as used by Letellier and Aguirre<sup>23</sup> where linear and nonlinear coupling terms are associated with solid and dashed arrows, respectively, and as used by Liu and coworkers<sup>17</sup> where coupling terms are labelled with arrows (without distinguishing linear from nonlinear couplings). In the present approach, rational terms are distinguished from nonlinear polynomial terms since they strongly reduce the observability<sup>2</sup>.
- ii) *Construction of the symbolic observability matrix ( $\tilde{\mathcal{O}}_s$ ).* Let us consider for simplicity a univariate measurement  $s = h(x) = x_i$ . For this particular case, the first row of  $\tilde{\mathcal{O}}_s$  is just defined by the derivative of the measurement function  $dh$ , that is,  $\tilde{\mathcal{O}}_{1j} = 1$  if  $j = i$  and 0 otherwise. The second row is directly obtained from  $\tilde{\mathcal{J}}$  by copying its  $i$ th row, that is,  $\tilde{\mathcal{O}}_{2j} = \tilde{J}_{ij} \forall j$ , being  $i$  the index of the measured variable. The  $k$ th row is obtained as follows. First, each element of the  $i$ th row of the symbolic Jacobian observability matrix  $\tilde{\mathcal{J}}$  is multiplied by the corresponding  $i$ th component of the vector  $\mathbf{v} = (\tilde{\mathcal{O}}_{\ell 1}, \dots, \tilde{\mathcal{O}}_{\ell d})$  where  $\ell = k - 1$  refers to the  $(k - 1)$ th row of the symbolic observability matrix  $\tilde{\mathcal{O}}_s$ . The rules to perform the symbolic product  $\tilde{J}_{ij} \otimes v_i$  are such that<sup>2</sup>

$$\left\{ \begin{array}{l} 0 \otimes a = 0, \\ 1 \otimes a = a, \\ \bar{1} \otimes a = a \text{ for } a = \bar{1}, \bar{\bar{1}}, \\ \vdots \\ \bar{1} \otimes a = \bar{1} \text{ for } a \neq 0. \end{array} \right. \quad (21)$$

Second, the resulting symbolic Jacobian matrix  $\tilde{\mathcal{J}}'$  is thus reduced into a row where each element  $\tilde{O}_{kj} = \sum_i \tilde{J}'_{ij}$  is just the sum of the elements of the  $j$ th column according to the addition law<sup>2</sup>

$$\left\{ \begin{array}{l} 0 \oplus a = a, \\ 1 \oplus a = a \text{ for } a \neq 0, \\ \bar{1} \oplus a = a \text{ for } a \neq 0, 1, \\ \vdots \\ \bar{1} \oplus a = \bar{1}. \end{array} \right. \quad (22)$$

In the case  $m$  variables are measured, the construction of  $\tilde{O}_s$  is performed by blocks of size  $(d_i + 1) \times d_i$ , being  $d_i$  the number of derivatives of  $s_i$  and  $\sum_{i=1}^m d_i + m = d$ , and the construction of each block follows the same rules described above for univariate measures.

iii) *Computation of the symbolic observability coefficients.* The determinant of  $\tilde{O}_s$  is computed according to the symbolic product rule defined in (21) and expressed as products and addends of the symbolic terms 1,  $\bar{1}$  and  $\bar{\bar{1}}$ , whose number of occurrences are stored in variables  $N_1$ ,  $N$  and  $N$ , respectively. A special condition is required for rational systems such that, if  $N = 0$  and  $N \neq 0$  then  $N = N$ . The symbolic observability coefficient for the measurement  $s$  is then equal to

$$\eta_s = \frac{1}{D} N_1 + \frac{1}{D^2} N_1 + \frac{1}{D^3} N_{\bar{1}} \quad (23)$$

with  $D = \max(1, N_1) + N + N$  and  $0 \leq \eta_s \leq 1$ , being  $\eta_s = 1$  for a combination providing full observability.

iv) *Selecting the minimal set of variables to measure and the adequate Lie derivatives for providing a full observability.* The symbolic observability coefficients  $\eta_s$  for each one of the sets of  $m$  measured variables and their selected  $d - m$  Lie derivatives are ranked versus the decreasing value of  $\eta_s$  and increasing  $m$ . Those featuring  $\eta_s = 1$  and the smallest  $m$  can be selected as the minimal sets of variables to measure.

## References

1. Aguirre, L. A. & Letellier, C. Observability of multivariate differential embeddings. *J. Phys. A* **38**, 6311 (2005).
2. Bianco-Martinez, E., Baptista, M. S. & Letellier, C. Symbolic computations of non-linear observability. *Phys. Rev. E* **91**, 062912 (2015).
3. Pagani, G. A. & Aiello, M. The Power Grid as a complex network: A survey. *Physica A* **392**, 2688–2700 (2013).
4. Ajrouch, K. J., Blandon, A. Y. & Antonucci, T. C. Social networks among men and women: The effects of age and socioeconomic status. *J. Gerontol. B* **60**, S311–S317 (2005).
5. Moore, C., Cumming, G. S., Slingsby, J. & Grewar, J. Tracking socioeconomic vulnerability using network analysis: Insights from an avian influenza outbreak in an ostrich production network. *PLoS One* **9**, e86973 (2014).
6. Varela, L. M., Rotundo, G., Ausloos, M. & Carrete, J. Complex network analysis in socioeconomic models. *Complexity and Geographical Economics, Dynamic Modeling and Econometrics in Economics and Finance* **19**, 209–245 (2015).
7. Colcombet, J. & Hirt, H. Arabidopsis MAPKs: a complex signalling network involved in multiple biological processes. *Biochem. J.* **413**, 217–226 (2008).
8. Rubinov, M. & Sporns, O. Complex network measures of brain connectivity: Uses and interpretations. *Neuroimage* **52**, 1059–1069 (2010).
9. Raue, A., Becker, V., Klingmüller, U. & Timmer, J. Identifiability and observability analysis for experimental design in nonlinear dynamical models. *Chaos* **20**, 045105 (2010).
10. Sendiña-Nadal, I. *et al.* Unveiling protein functions through the dynamics of the interaction network. *PLoS One* **6**, e17679 (2011).
11. Allefeld, C. & Kurths, J. An approach to multivariate phase synchronization analysis and its application to event-related potentials. *Int. J. Bifurcat. Chaos* **14**, 417–426 (2004).
12. Donner, R. Multivariate analysis of spatially heterogeneous phase synchronisation in complex systems: application to self-organised control of material flows in networks. *Eur. Phys. J. B* **63**, 349–361 (2008).
13. Letellier, C. & Aguirre, L. A. Symbolic observability coefficients for univariate and multivariate analysis. *Phys. Rev. E* **79**, 066210 (2009).
14. Slutsker, I. W. & Scudder, J. M. Network observability analysis through measurement Jacobian matrix reduction. *IEEE Transactions on Power Systems* **2**(2), 331–336 (1987).
15. Bretas, N. G. & London, J. B. A. Network observability: the critical measurement identification using the symbolic Jacobian matrix. *IEEE Proceedings of the International Conference on Power System Technology* **2**, 1222–1226 (1998).
16. Sedoglavic, A. A probabilistic algorithm to test local algebraic observability in polynomial time. *J. Symb. Comput.* **33**, 735–755 (2002).
17. Liu, Y.-Y., Slotine, J.-J. & Barabási, A.-L. Observability of complex systems. *Proc. Natl. Acad. Sci. (USA)* **110**, 2460–2465 (2013).
18. Whalen, A. J., Brennan, S. N., Sauer, T. D. & Schiff, S. J. Observability and controllability of nonlinear networks: The role of symmetry. *Phys. Rev. X* **5**, 0111005 (2015).
19. Leitold, D., Vathy-Fogarassy, Á. & Abonyi, J. Controllability and observability in complex networks – the effect of connection types. *Scientific Reports* **7**, 151 (2017).
20. Hermann, R. & Krener, A. Nonlinear controllability and observability. *IEEE Trans. Autom. Control* **22**, 728–740 (1977).
21. Letellier, C., Maquet, J., Le Sceller, L., Gouesbet, G. & Aguirre, L. A. On the non-equivalence of observables in phase space reconstructions from recorded time series. *J. Phys. A* **31**, 7913–7927 (1998).
22. Letellier, C. & Aguirre, L. A. Investigating nonlinear dynamics from time series: The influence of symmetries and the choice of observables. *Chaos* **12**, 549–558 (2002).

23. Letellier, C. & Aguirre, L. A. A graphical interpretation of observability in terms of feedback circuits. *Phys. Rev. E* **72**, 056202 (2005).
24. Yang, Y., Wang, J. & Motter, A. E. Network observability transitions. *Phys. Rev. Lett.* **109**, 258701 (2012).
25. Frunzete, M., Barbot, J.-P. & Letellier, C. Influence of the singular manifold of non-observable states in reconstructing chaotic attractors. *Phys. Rev. E* **86**, 026205 (2012).
26. Aguirre, L. A., Portes, L. L. & Letellier, C. Observability and synchronization of neuron models. *Chaos* **27**, 103103 (2017).
27. Banaji, M., Donnell, P. & Baigent, S. P matrix properties, injectivity and stability in chemical reaction systems. *SIAM J. Appl. Math.* **67**, 1523–1547 (2007).
28. Letellier, C. & Gilmore, R. Covering dynamical systems: Two-fold covers. *Phys. Rev. E* **63**, 016206 (2001).
29. Goldbeter, A. A model for circadian oscillations in the *Drosophila* period protein. *Proc. Royal Soc. B* **261**, 319–324 (1995).
30. Letellier, C., Aguirre, L. A. & Maquet, J. Relation between observability and differential embeddings for nonlinear dynamics. *Phys. Rev. E* **71**, 066213 (2005).
31. Reiterer, P., Lainscsek, C., Schürer, F., Letellier, C. & Maquet, J. A nine-dimensional Lorenz system to study high-dimensional chaos. *J. Phys. A* **31**, 7121–7139 (1998).
32. Gilmore, R. & Letellier, C. *The Symmetry of Chaos* (Oxford University Press, 2007).
33. Sendiña-Nadal, I., Boccaletti, S. & Letellier, C. Observability coefficients for predicting the class of synchronizability from the algebraic structure of the local oscillators. *Phys. Rev. E* **94**, 042205 (2016).
34. Novak, B. & Tyson, J. J. Modeling the control of DNA replication in fission yeast. *Proc. Natl. Acad. Sci.* **94**, 9147–9152 (1997).
35. Wang, B., Gao, L., Gao, Y., Deng, Y. & Wang, Y. Controllability and observability analysis for vertex domination centrality in directed networks. *Sci. Rep.* **4**, 5399 (2014).
36. Parlitz, U., Schumann-Bischoff, J. & Luther, S. Local observability of state variables and parameters in nonlinear modeling quantified by delay reconstruction. *Chaos* **24**, 024411 (2014).
37. Gibson, J. F., Farmer, J. D., Casdagli, M. & Eubank, S. An analytic approach to practical state space reconstruction. *Physica D* **57**, 1–30 (1992).
38. Sarachik, P. E. & Kreindler, E. Controllability and observability of linear discrete-time systems. *International Journal of Control* **1**(5), 419–432 (1965).
39. Nijmeijer, H. Observability of autonomous discrete time non-linear systems: a geometric approach. *International Journal of Control* **36**(5), 867–874 (1982).
40. Kailath, T. *Linear Systems* (Prentice Hall, 1980).
41. Takens, F. Detecting strange attractors in turbulence. *Lecture Notes in Mathematics* **898**, 366–381 (1981).
42. Letellier, C. Estimating the Shannon entropy: recurrence plots versus symbolic dynamics. *Physical Review Letters* **96**, 254102 (2006).
43. Portes, L. L., Benda, R. N., Ugrinowitsch, H. & Aguirre, L. A. Impact of the recorded variable on recurrence quantification analysis of flows. *Physics Letters A* **378**, 2382–2388 (2014).

## Acknowledgements

EBM and MSB acknowledge the Engineering and Physical Sciences Research Council (EPSRC), grant Ref. EP/I032608/1. ISN acknowledges partial support from the Ministerio de Economía y Competitividad of Spain under project FIS2013-41057-P and from the Group of Research Excellence URJC-Banco de Santander.

## Author Contributions

C.L. and M.S.B. conceived the study. E.B.-M. and I.S.-N. prepared the MATLAB codes and made the symbolic calculations. CL performed the analytical computations. C.L., M.S.B. and I.S.-N. wrote the Manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-21967-w>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018