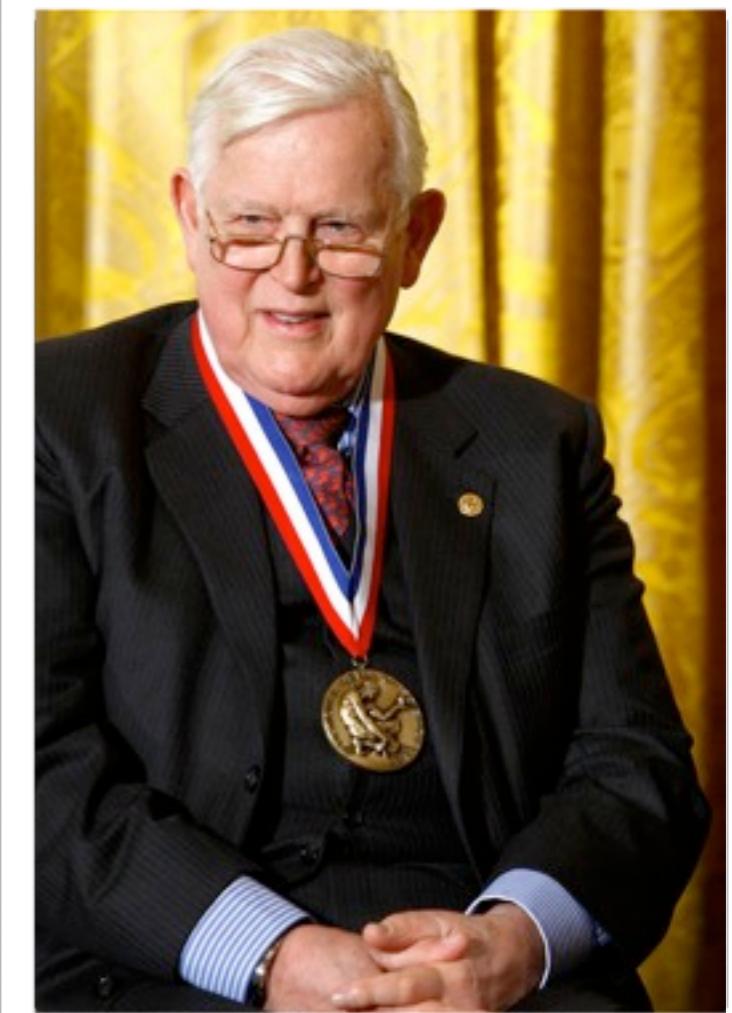


Observability of Complex Biological Systems

Yang-Yu Liu

*Center for Complex Network Research @ Northeastern University
Center for Cancer Systems Biology @ Dana-Farber Cancer Institute*



J.S.I.A.M. CONTROL
Ser. A, Vol. 1, No. 2
Printed in U.S.A., 1963

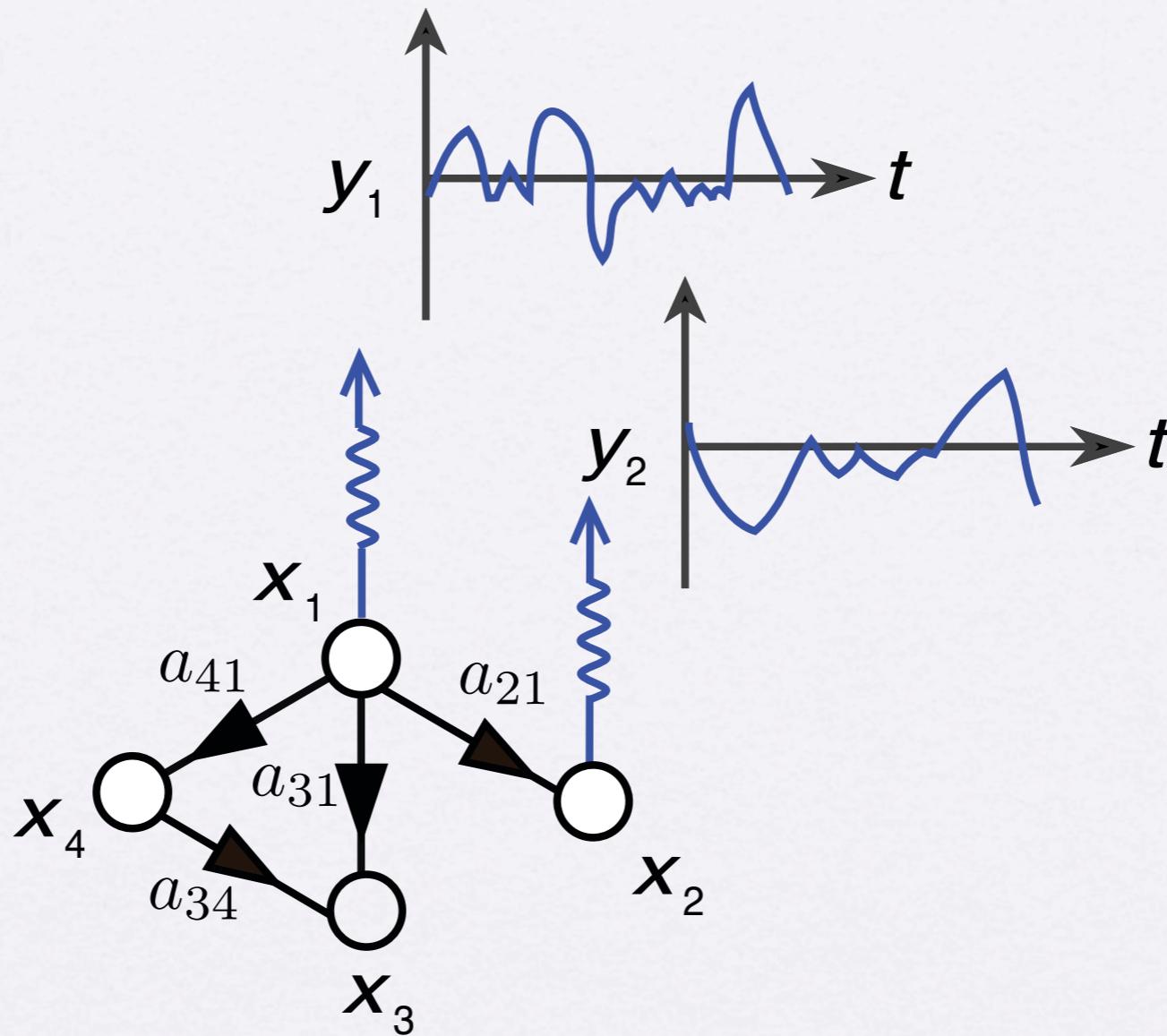
MATHEMATICAL DESCRIPTION OF LINEAR DYNAMICAL SYSTEMS*

R. E. KALMAN†

Abstract. There are two different ways of describing dynamical systems: (i) by means of state variables and (ii) by input/output relations. The first method may be regarded as an axiomatization of Newton's laws of mechanics and is taken to be the basic definition of a system.

It is then shown (in the linear case) that the input/output relations determine only one part of a system, that which is completely observable and completely controllable. Using the theory of controllability and observability, methods are given for calculating irreducible realizations of a given impulse-response matrix. In particular, an explicit procedure is given to determine the minimal number of state variables necessary to realize a given transfer-function matrix. Difficulties arising from the use of reducible realizations are discussed briefly.





Observable: the system's current state can be determined in finite time using only the system's outputs, e.g., the measurements of a few nodes.

Motivation

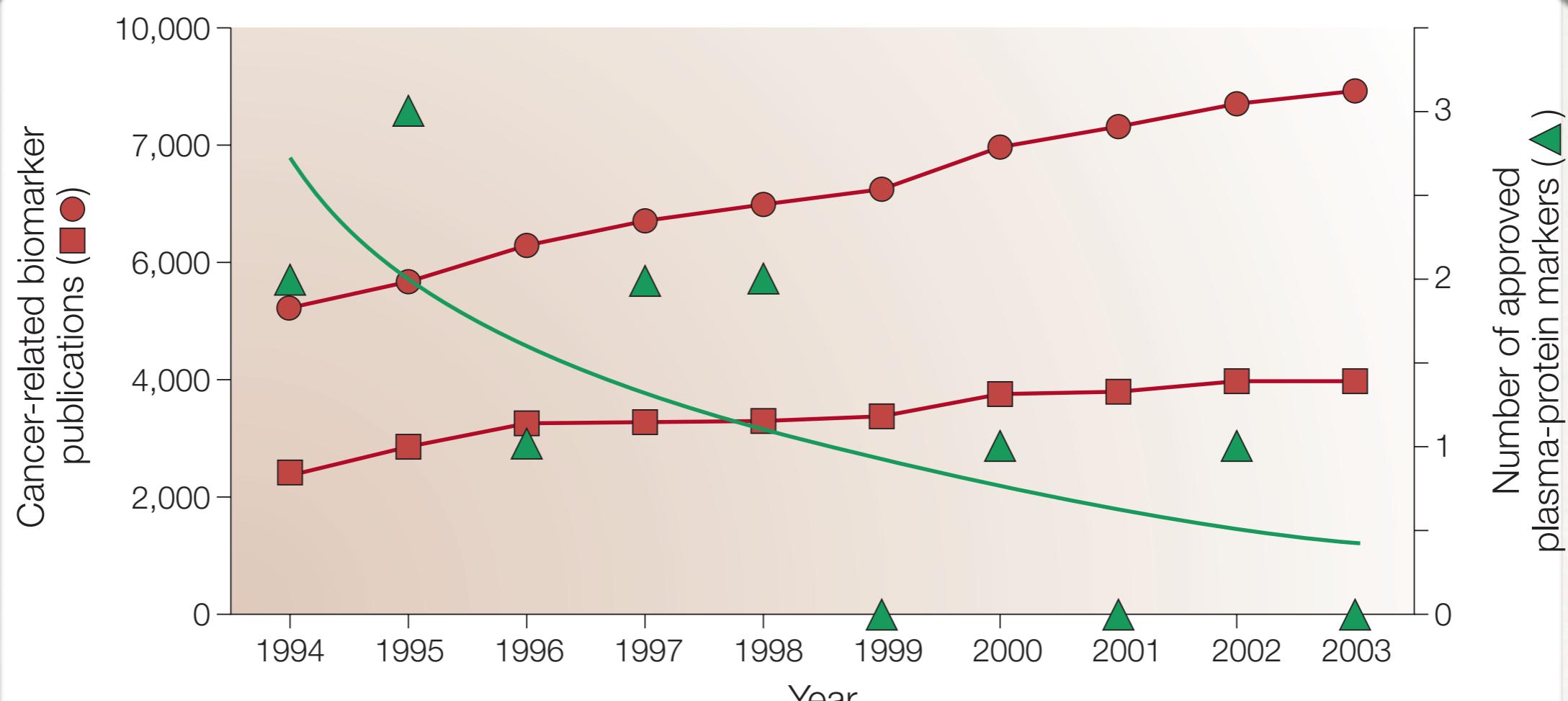
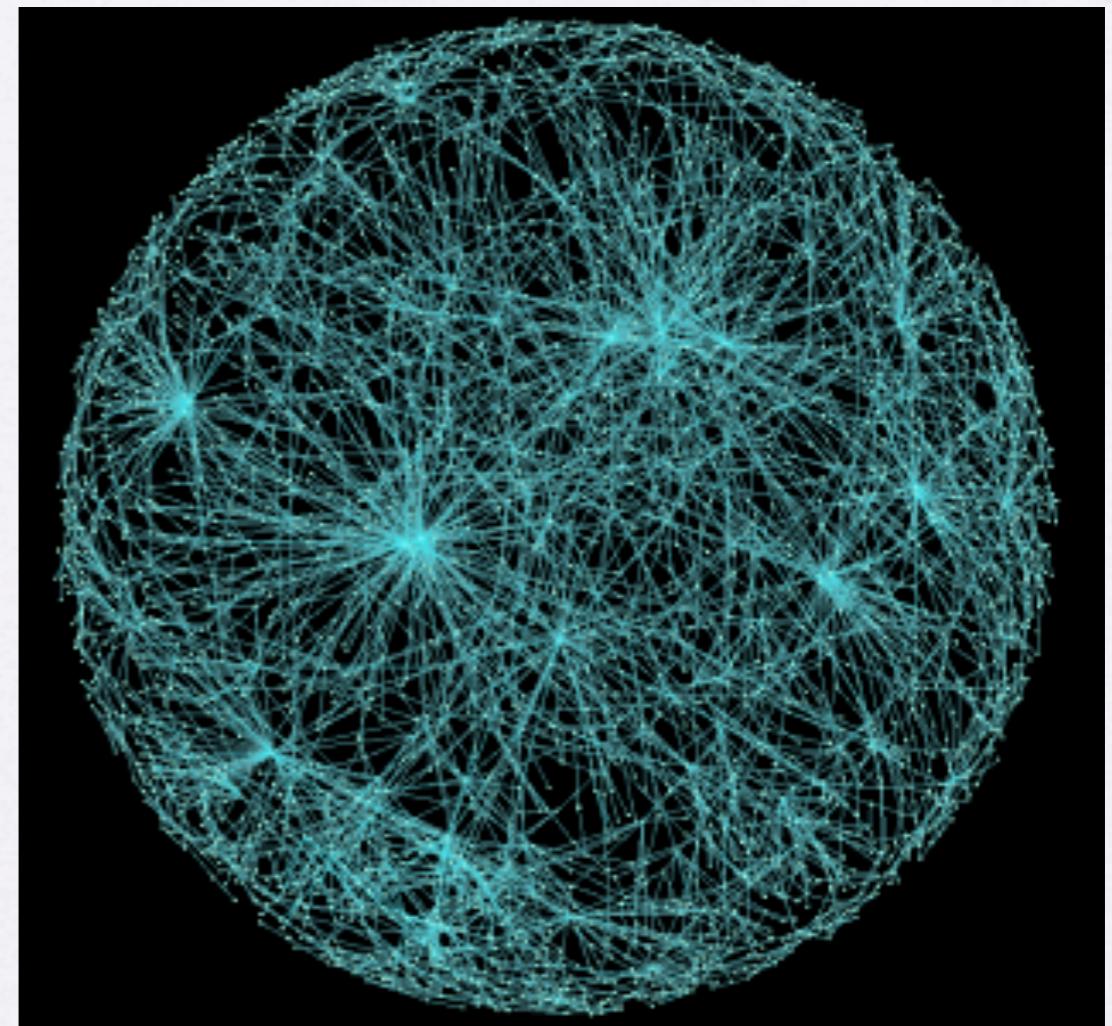


Figure 1 | Numbers of publications on biomarkers and FDA approval of biomarkers.

Despite the increasing rates of publications on biomarkers, the number of FDA-approved biomarkers is decreasing.

Our Approach

- **Observability analysis** of a mathematically well-defined biological system: biochemical reaction system, e.g., metabolic networks.
- **Key assumption:** Components, which if monitored can be used to infer the whole state of the system, are sensors or biomarkers.



Use control theory to guide rational biomarker design

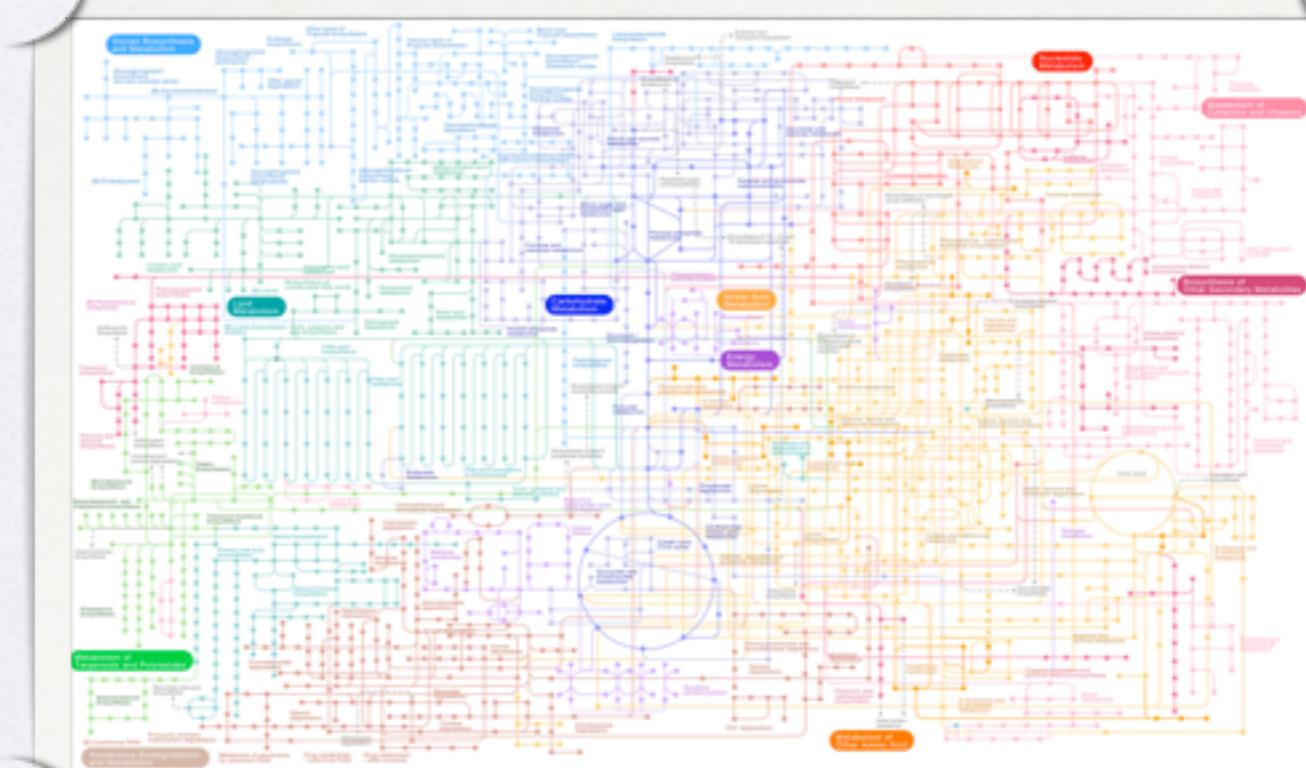
Questions

For a general nonlinear dynamical system:

- * What's the minimum number of sensor nodes (N_s) ?
- * How to efficiently identify them?

For metabolic network:

How to find the minimum set of metabolites so that measuring their concentrations over time will enable us to infer all other metabolites' concentrations?



Linear System

- Linear Time-Invariant System

$$\dot{\mathbf{x}}(t) = \mathbf{A} \mathbf{x}(t) + \mathbf{B} \mathbf{u}(t)$$

$$\mathbf{y}(t) = \mathbf{C} \mathbf{x}(t)$$

$\mathbf{x}(t) \in \mathbb{R}^{N \times 1}$: state vector.

$\mathbf{u}(t) \in \mathbb{R}^{M \times 1}$: input vector ($M \leq N$).

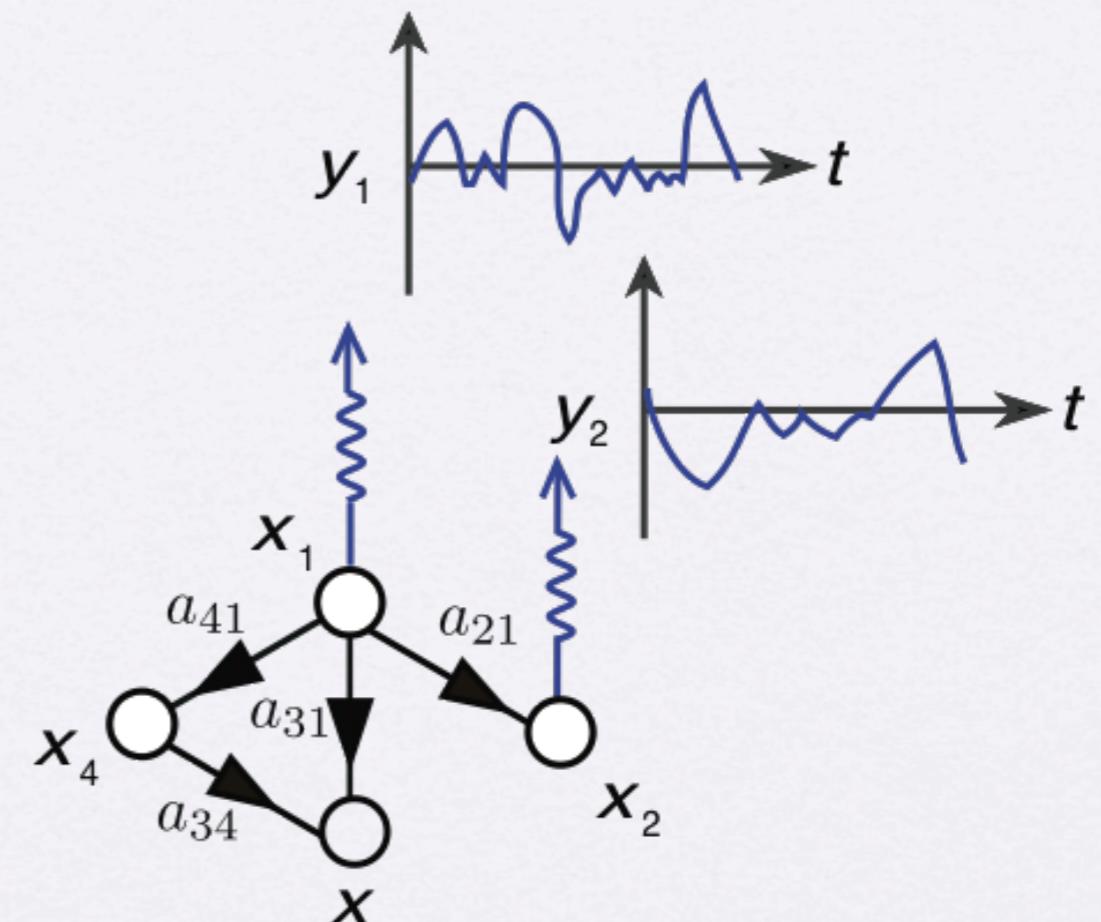
$\mathbf{y}(t) \in \mathbb{R}^{K \times 1}$: output vector ($K \leq N$).

$\mathbf{A} \in \mathbb{R}^{N \times N}$: state matrix

(weighted wiring diagram).

$\mathbf{B} \in \mathbb{R}^{N \times M}$: input matrix

$\mathbf{C} \in \mathbb{R}^{N \times K}$: output matrix



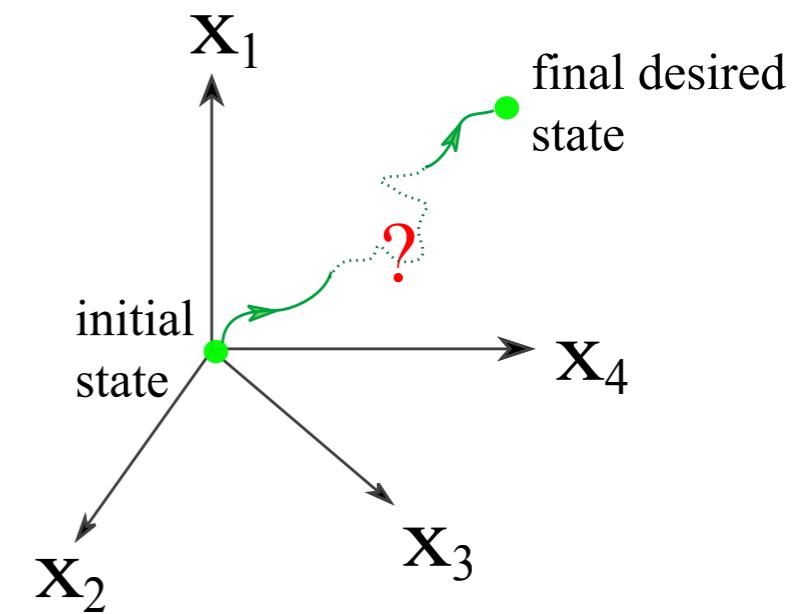
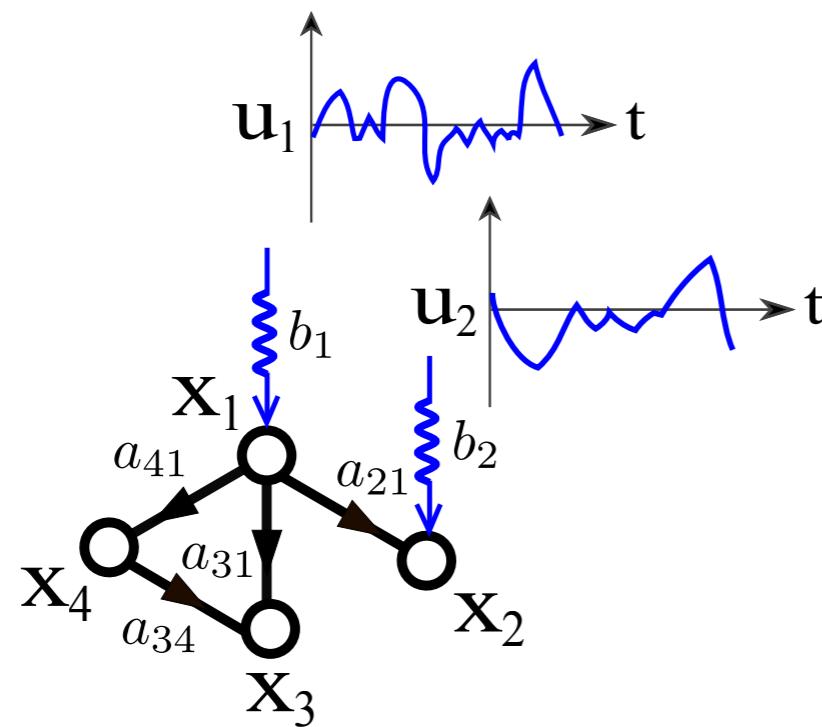
- Kalman's Rank Condition:

The LTI system is observable iff its observability matrix has full rank.

$$\boxed{\text{rank } \mathbf{O} = N}$$

$$\mathbf{O} = [\mathbf{C}^T, \mathbf{A}^T \mathbf{C}^T, (\mathbf{A}^2)^T \mathbf{C}^T, \dots, (\mathbf{A}^{N-1})^T \mathbf{C}^T]^T$$

Controllability



Controllable: the system can be driven from any initial state to any desired final state in finite time.

Duality: Observability & Controllability

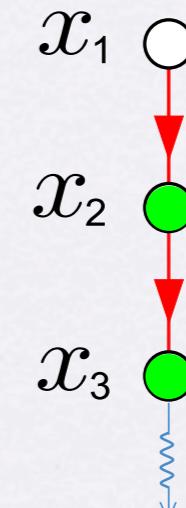
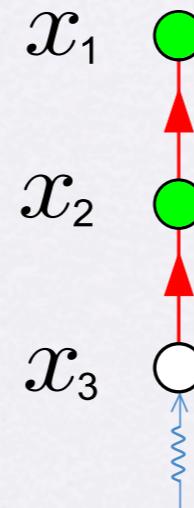
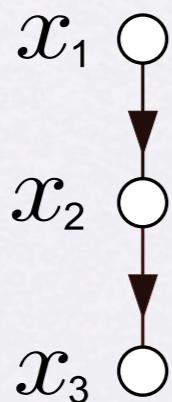
The system

$$\begin{cases} \dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \\ \mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) \end{cases}$$

is **observable** iff the **dual system**

$$\begin{cases} \dot{\mathbf{x}}(t) = \mathbf{A}^T\mathbf{x}(t) + \mathbf{C}^T\mathbf{u}(t) \\ \mathbf{y}(t) = \mathbf{B}^T\mathbf{x}(t) \end{cases}$$

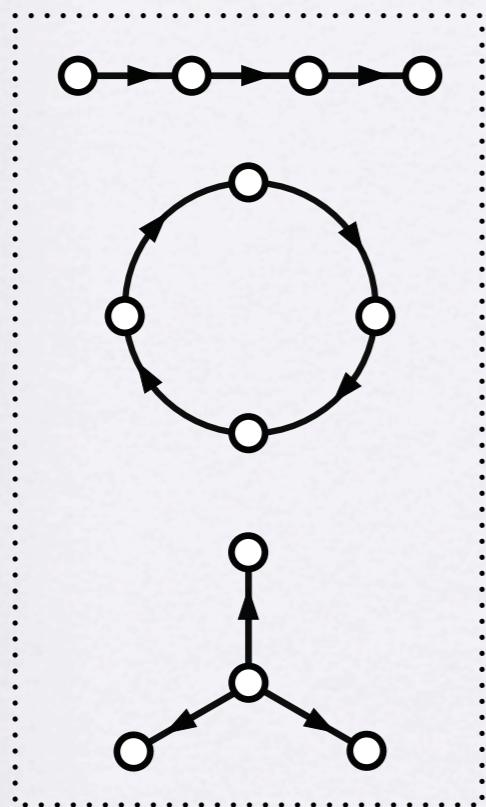
is **controllable**.



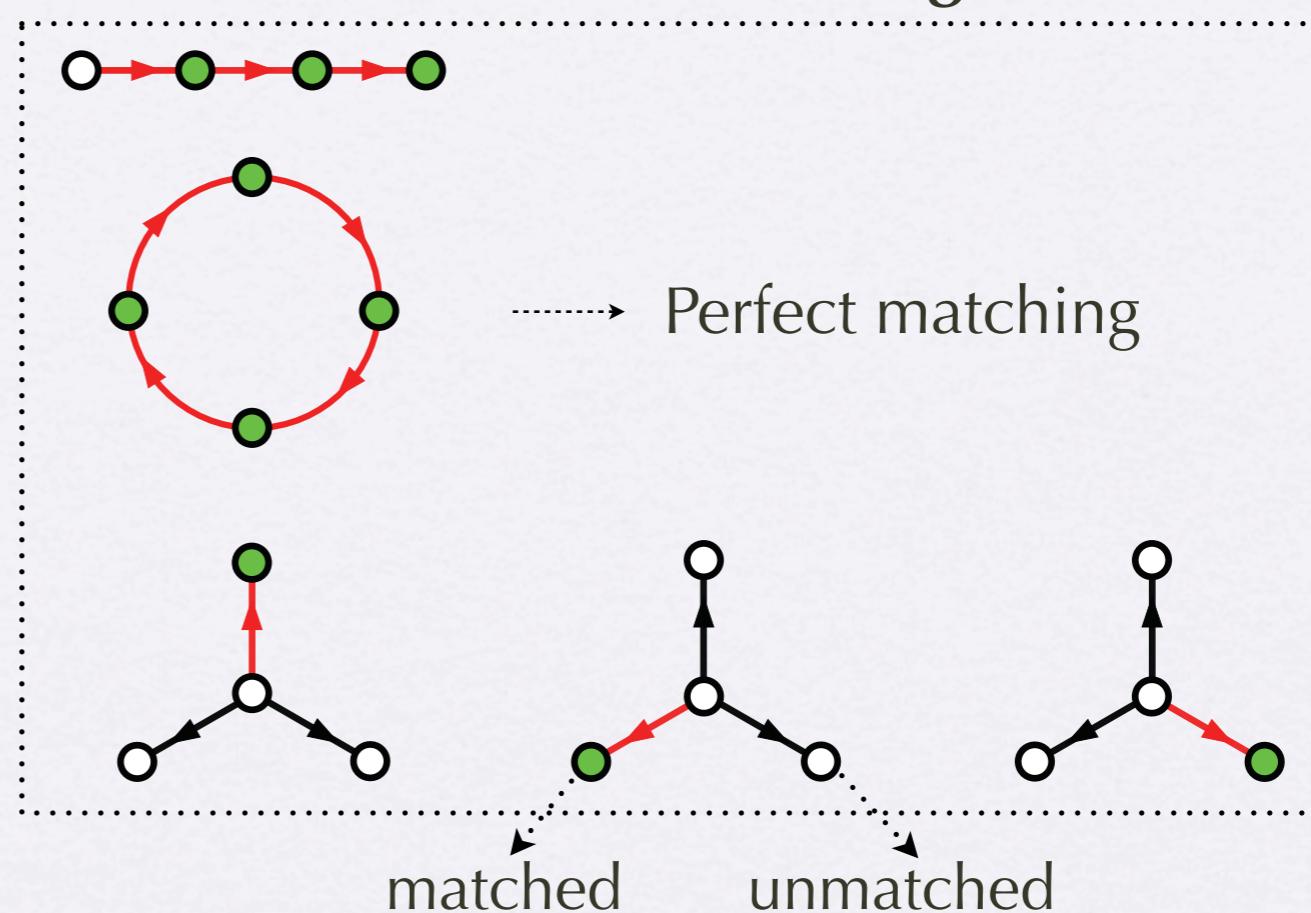
Minimum Inputs Theorem

Matching : a set of directed edges **without common heads or tails**.

Directed network



Maximum matching



Minimum Inputs Theorem

Driver nodes = unmatched nodes

$$N_D = \max\{1, N_{\text{unmatched}}\}$$

nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

TAMING COMPLEXITY

The mathematics of network control – from cell biology to cellphones PAGES 158 & 167

ANTIBIOTICS
A SPOONFUL OF SUGAR
Carbohydrate 'helpers' boost antibacterial efficacy PAGE 216

POLICY
WHO NEEDS CHANGE
A World Health Organization for today's world PAGE 143

GEOLOGY
WAITING FOR VESUVIUS
What to do when Europe's most deadly volcano blows PAGE 140

NATURE.COM/NATURE
12 May 2011

\$10.00 US \$12.99 CAN 19>
0 71486 03070 6

ARTICLE

doi:10.1038/nature10011

Controllability of complex networks

Yang-Yu Liu^{1,2}, Jean-Jacques Slotine^{3,4} & Albert-László Barabási^{1,2,5}

The ultimate proof of our understanding of natural or technological systems is reflected in our ability to control them. Although control theory offers mathematical tools for steering engineered and natural systems towards a desired state, a framework to control complex self-organized systems is lacking. Here we develop analytical tools to study the controllability of an arbitrary complex directed network, identifying the set of driver nodes with time-dependent control that can guide the system's entire dynamics. We apply these tools to several real networks, finding that the number of driver nodes is determined mainly by the network's degree distribution. We show that sparse inhomogeneous networks, which emerge in many real complex systems, are the most difficult to control, but that dense and homogeneous networks can be controlled using a few driver nodes. Counterintuitively, we find that in both model and real systems the driver nodes tend to avoid the high-degree nodes.

According to control theory, a dynamical system is controllable if, with a suitable choice of inputs, it can be driven from any initial state to any desired final state within finite time^{1–3}. This definition agrees with our intuitive notion of control, capturing an ability to guide a system's behaviour towards a desired state through the appropriate manipulation of a few input variables, like a driver prompting a car to move with the desired speed and in the desired direction by manipulating the pedals and the steering wheel. Although control theory is a mathematically highly developed branch of engineering with applications to electric circuits, manufacturing processes, communication systems^{4–6}, aircraft, spacecraft and robots^{7,8}, fundamental questions pertaining to the controllability of complex systems emerging in nature and engineering have resisted advances. The difficulty is rooted in the fact that two independent factors contribute to controllability: each with its own layer of unknown: (1) the system's architecture, represented by the network encapsulating which components interact with each other; and (2) the dynamical rules that capture the time-dependent interactions between the components. Thus, progress has been possible only in systems where both layers are well mapped, such as the control of synchronized networks^{9–10}, small biological circuits¹¹ and rate control for communication networks^{4–6}. Recent advances towards quantifying the topological characteristics of complex networks^{12–16} have shed light on factor (1), prompting us to wonder whether some networks are easier to control than others and how network topology affects a system's controllability. Despite some pioneering conceptual work^{17–23} (Supplementary Information, section II), we continue to lack general answers to these questions for large weighted and directed networks, which most commonly emerge in complex systems.

Network controllability

Most real systems are driven by nonlinear processes, but the controllability of nonlinear systems is in many aspects structurally similar to that of linear systems³, prompting us to start our study using the canonical linear, time-invariant dynamics

$$\frac{dx(t)}{dt} = Ax(t) + Bu(t) \quad (1)$$

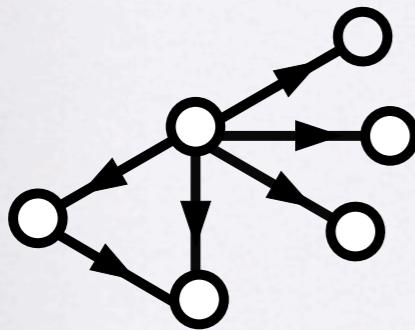
where the vector $x(t) = (x_1(t), \dots, x_N(t))^T$ captures the state of a system of N nodes at time t . For example, $x_i(t)$ can denote the amount

¹Center for Complex Network Research and Departments of Physics, Computer Science and Biology, Northeastern University, Boston, Massachusetts 02115, USA. ²Center for Cancer Systems Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. ³Nonlinear Systems Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ⁴Department of Mechanical Engineering and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ⁵Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA.

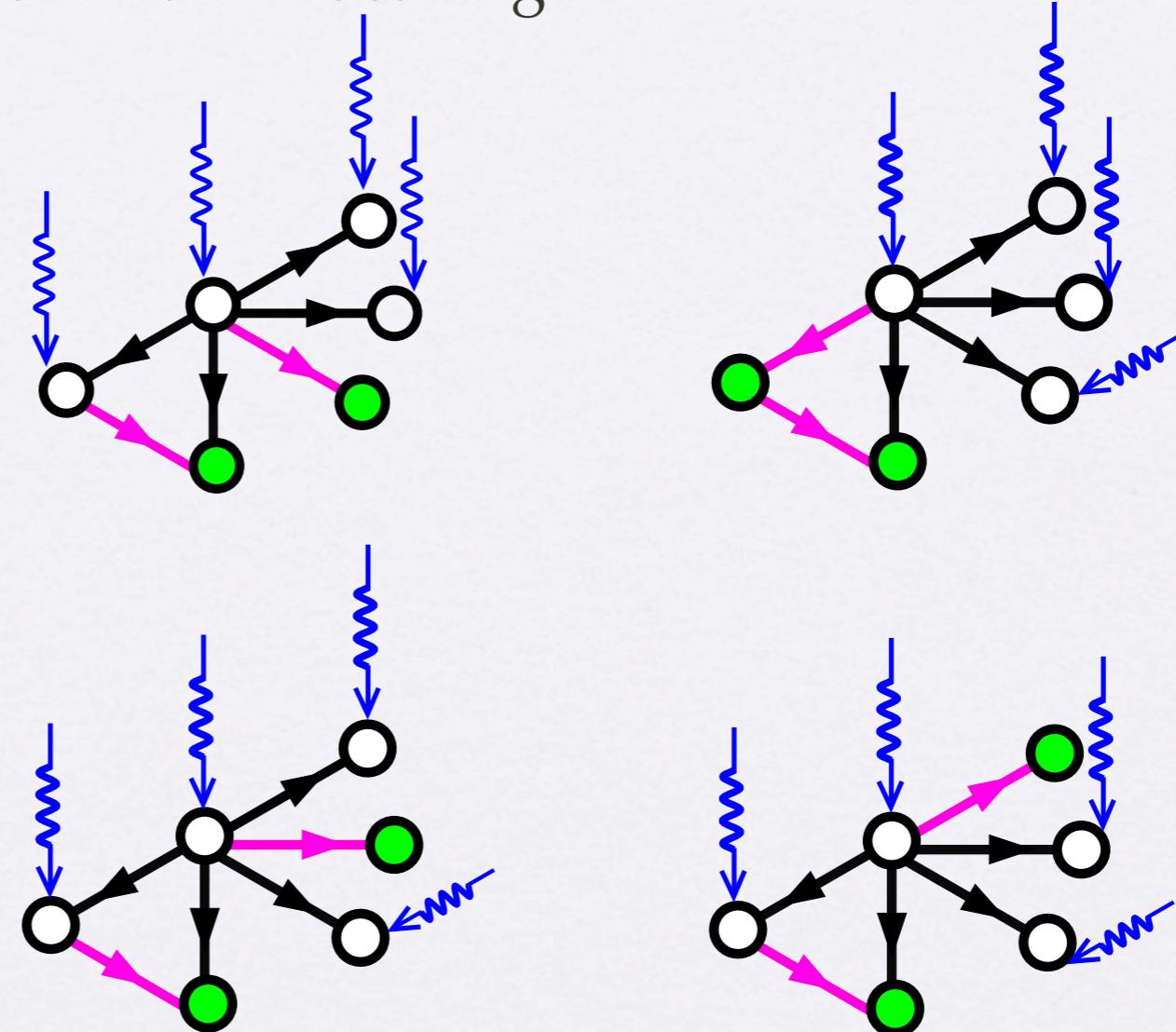
00 MONTH 2011 | VOL 000 | NATURE | 1

Methodology

network



maximum matching



Brute-force search:
 $(2^N - 1)$ combinations,
 $\sim 10^{30}$ for $N=100$. **Hopeless!**



Hopcroft-Karp Algorithm
Time complexity $O(N^{1/2} L)$,
Fast even for $N \sim 10^6$

Linear System

Driver nodes of the transposed network are sensor nodes for the original network.

Yet, most complex systems are nonlinear:

$$\begin{cases} \dot{\mathbf{x}}(t) = \mathbf{f}(t, \mathbf{x}(t), \mathbf{u}(t)) \\ \mathbf{y}(t) = \mathbf{h}(t, \mathbf{x}(t), \mathbf{u}(t)) \end{cases}$$

Nonlinear System

The observability of a rational system is determined by the dimension of the space spanned by gradients of the Lie-derivatives of its output functions.

$$L := \frac{\partial}{\partial t} + \sum_{i=1}^N f_i \frac{\partial}{\partial x_i} + \sum_{j \in \mathbb{N}} \sum_{l=1}^K u_l^{(j+1)} \frac{\partial}{\partial u_l^{(j)}}$$

The nonlinear system is algebraically observable if and only if the $NM \times N$ Jacobian matrix has full rank.

$$\text{rank } \mathcal{J} = \text{rank} \begin{bmatrix} \frac{\partial L_f^0 h_1}{\partial x_1} & \frac{\partial L_f^0 h_1}{\partial x_2} & \dots & \frac{\partial L_f^0 h_1}{\partial x_N} \\ \dots & \dots & \dots & \dots \\ \frac{\partial L_f^0 h_M}{\partial x_1} & \frac{\partial L_f^0 h_M}{\partial x_2} & \dots & \frac{\partial L_f^0 h_M}{\partial x_N} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial L_f^{N-1} h_1}{\partial x_1} & \frac{\partial L_f^{N-1} h_1}{\partial x_2} & \dots & \frac{\partial L_f^{N-1} h_1}{\partial x_N} \\ \dots & \dots & \dots & \dots \\ \frac{\partial L_f^{N-1} h_M}{\partial x_1} & \frac{\partial L_f^{N-1} h_M}{\partial x_2} & \dots & \frac{\partial L_f^{N-1} h_M}{\partial x_N} \end{bmatrix} = N$$

Difficulties

- Parameters (e.g., rate constants): usually unknown.

- Dynamics: highly nonlinear.

- Rank condition: hard to check for large system

* Jacobian matrix J is of dimension $NM \times N$

* Symbolic calculation

- If brute-force search: $(2^N - 1)$ combinations.

$$\begin{cases} \dot{x}_1 = -k_1x_1x_2x_3 \\ \dot{x}_2 = -k_1x_1x_2x_3 \\ \dot{x}_3 = -k_1x_1x_2x_3 \\ \dot{x}_4 = +k_1x_1x_2x_3 - k_2x_4 + k_3x_5 \\ \dot{x}_5 = +k_2x_4 - k_3x_5 \\ \dot{x}_6 = +k_1x_1x_2x_3 \\ \dot{x}_7 = +k_4x_8x_9 - k_5x_7 + k_6x_{10}x_{11} \\ \dot{x}_8 = -k_4x_8x_9 + k_5x_7 + k_6x_{10}x_{11} \\ \dot{x}_9 = -k_4x_8x_9 + k_5x_7 \\ \dot{x}_{10} = +k_1x_1x_2x_3 - k_6x_{10}x_{11} \\ \dot{x}_{11} = -k_6x_{10}x_{11} \end{cases}$$

$$\text{rank } \mathcal{J} = \text{rank} \begin{bmatrix} \frac{\partial L_f^0 h_1}{\partial x_1} & \frac{\partial L_f^0 h_1}{\partial x_2} & \dots & \frac{\partial L_f^0 h_1}{\partial x_N} \\ \dots & \dots & \dots & \dots \\ \frac{\partial L_f^0 h_M}{\partial x_1} & \frac{\partial L_f^0 h_M}{\partial x_2} & \dots & \frac{\partial L_f^0 h_M}{\partial x_N} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial L_f^{N-1} h_1}{\partial x_1} & \frac{\partial L_f^{N-1} h_1}{\partial x_2} & \dots & \frac{\partial L_f^{N-1} h_1}{\partial x_N} \\ \dots & \dots & \dots & \dots \\ \frac{\partial L_f^{N-1} h_M}{\partial x_1} & \frac{\partial L_f^{N-1} h_M}{\partial x_2} & \dots & \frac{\partial L_f^{N-1} h_M}{\partial x_N} \end{bmatrix} = N$$

$$\binom{N}{1} + \binom{N}{2} + \dots + \binom{N}{N} = 2^N - 1$$

Solution

Graphical approach based on inference diagram

Draw a directed link $x_i \rightarrow x_j$ if x_j appears in x_i 's differential equation.

$$\dot{x}_i = f(\dots, x_j, \dots)$$

$$x_i \rightarrow x_j$$



Observability of complex systems

Yang-Yu Liu^{a,b,c,d,e}, Jean-Jacques Slotine^{f,g,h}, and Albert-László Barabási^{a,b,c,d,i,1}

^aCenter for Complex Network Research and Departments of ^bPhysics, ^cComputer Science, and ^dBiology, Northeastern University, Boston, MA 02115; ^eCenter for Cancer Systems Biology, Dana-Farber Cancer Institute, Boston, MA 02115; ^fNonlinear Systems Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139; Departments of ^gMechanical Engineering and ^hBrain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139; and ⁱDepartment of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115

Edited by Giorgio Parisi, University of Rome, Rome, Italy, and approved December 26, 2012 (received for review September 6, 2012)

A quantitative description of a complex system is inherently limited by our ability to estimate the system's internal state from experimentally accessible outputs. Although the simultaneous measurement of all internal variables, like all metabolite concentrations in a cell, offers a complete description of a system's state, in practice experimental access is limited to only a subset of variables, or sensors. A system is called observable if we can reconstruct the system's complete internal state from its outputs. Here, we adopt a graphical approach derived from the dynamical laws that govern a system to determine the sensors that are necessary to reconstruct the full internal state of a complex system. We apply this approach to biochemical reaction systems, finding that the identified sensors are not only necessary but also sufficient for observability. The developed approach can also identify the optimal sensors for target or partial observability, helping us reconstruct selected state variables from appropriately chosen outputs, a prerequisite for optimal biomarker design. Given the fundamental role observability plays in complex systems, these results offer avenues to systematically explore the dynamics of a wide range of natural, technological and socioeconomic systems.

algebraic observability | biochemical reactions | control theory

The internal variables of a complex system are rarely independent of each other, as the interactions between the system's components induce systematic interdependences between them. Hence, a well-selected subset of variables can contain sufficient information about the rest of the variables, allowing us to reconstruct the system's complete internal state, making the system observable. To address observability in quantitative terms, we focus on systems whose dynamics can be described by the generic state-space form

$$\dot{\mathbf{x}}(t) = \mathbf{f}(t, \mathbf{x}(t), \mathbf{u}(t)), \quad [1]$$

where $\mathbf{x}(t) \in \mathbb{R}^N$ represents the complete internal state of the system (e.g., the concentrations of all metabolites in a cell), and the input vector $\mathbf{u}(t) \in \mathbb{R}^K$ captures the influence of the environment. Observing the system means that we monitor a set of variables $\mathbf{y}(t) \in \mathbb{R}^M$ that depend on the time t , the system's internal state $\mathbf{x}(t)$, and the external input $\mathbf{u}(t)$,

$$\mathbf{y}(t) = \mathbf{h}(t, \mathbf{x}(t), \mathbf{u}(t)). \quad [2]$$

Observability requires us to establish a relationship between the outputs $\mathbf{y}(t)$, the state vector $\mathbf{x}(t)$, and the inputs $\mathbf{u}(t)$ in a manner that we can uniquely infer the system's complete initial state $\mathbf{x}(0)$. The observability criteria can be formulated algebraically for dynamical systems consisting of polynomial or rational expressions (1, 2) stating that [1] is observable if the Jacobian matrix $\mathcal{J} = [\mathcal{J}_{ij}]_{NM \times N}$ has full rank,

$$\text{rank } \mathcal{J} = N, \quad [3]$$

where $\mathcal{J}_{ij} = \frac{\partial L_f}{\partial x_i} \Big|_{x=0}^{x_j} h_{(j-1)\% M+1}$, the Lie derivatives $L_f = \frac{\partial}{\partial t} + \sum_{i=1}^N f_i \frac{\partial}{\partial x_i} + \sum_{j \in \mathbb{N}} \sum_{l=1}^K u_l^{(j+1)} \frac{\partial}{\partial u_l^{(j)}}$, $[x]$ is the largest integer not greater than x , and $\%$ is the modulo operation (SI Text,

section 1). For a linear time-invariant dynamic system (3, 4), $\dot{\mathbf{x}}(t) = \mathbf{A} \mathbf{x}(t) + \mathbf{B} \mathbf{u}(t)$ and $\mathbf{y}(t) = \mathbf{C} \mathbf{x}(t)$, \mathcal{J} reduces to the observability matrix $\mathcal{O} = [\mathbf{C}^T, (\mathbf{C} \mathbf{A})^T, \dots, (\mathbf{C} \mathbf{A}^{N-1})^T]^T$.

To simplify the observability analysis, we assume that we can monitor a selected subset of state variables, i.e., $\mathbf{y}(t) = (\dots, x_i(t), \dots)^T$, which we call sensors. Observability of complex systems can then be posed as follows: Identify the minimum set of sensors from whose measurements we can determine all other state variables. Whereas [3] offers a formal answer to the observability issue in the context of small engineered systems, it has notable practical limitations for natural and complex systems. First, it can only confirm (or deny) if a specific sensor set can be used to observe a system, without telling us how to select it. Second, a brute-force search for a minimum sensor set requires us to inspect via [3] of about 2^N sensor combinations, a computationally prohibitive task for large complex systems. Third, the rank test of the Jacobian matrix via symbolic computation is computationally limited to small systems (5). Hence, the fundamental and the practically useful question of identifying the minimum set of sensors through which we can observe a large complex system remains unsolved.

To resolve these limitations, one can exploit the dynamic interdependence of the system's components through a graphical representation, a common approach used in structured system theory (6–10). This procedure consists of the following steps:

- i) Inference diagram: We draw a directed link $x_i \rightarrow x_j$ if x_j appears in x_i 's differential equation (i.e., if $\frac{\partial f_i}{\partial x_j}$ is not identically zero), implying that one can collect information on x_i by monitoring x_j as a function of time. Because the constructed network captures the information flow in inferring the state of individual variables, we call it an inference diagram (Fig. 1C). By flipping the direction of each edge, the procedure recovers the system digraph encountered in structured systems theory (11–13).
- ii) Strongly connected component (SCC) decomposition: We decompose the inference diagram into a unique set of maximal SCCs, which are the largest subgraphs chosen such that there is a directed path from each node to every other node in the subgraph (14). The SCCs of the inference diagram of Fig. 1C are surrounded by dashed circles. Note that each node in a SCC contains information pertaining to all other nodes within the SCC.
- iii) Sensor node selection: We call root SCCs those SCCs that have no incoming edges (shaded circles in Fig. 1C). We choose at least one node from each root SCC to ensure observability of the whole system. For example, the inference

Author contributions: Y.-Y.L., J.-J.S., and A.-L.B. designed research; Y.-Y.L., J.-J.S., and A.-L.B. performed research; Y.-Y.L. contributed analytic tools; Y.-Y.L. analyzed data; and Y.-Y.L. and A.-L.B. wrote the paper.

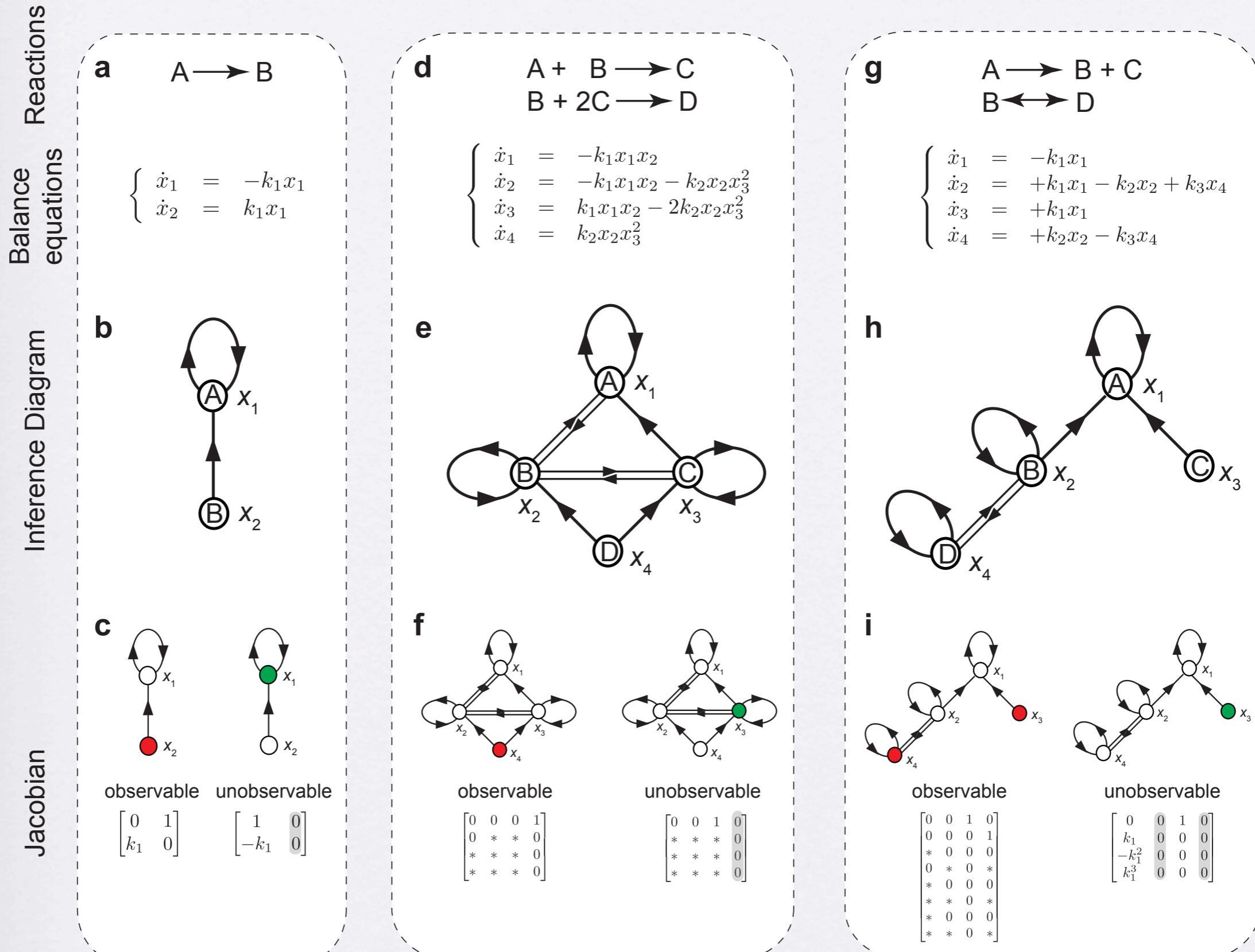
The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

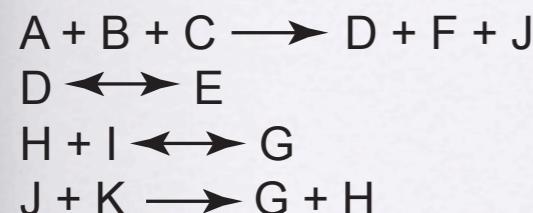
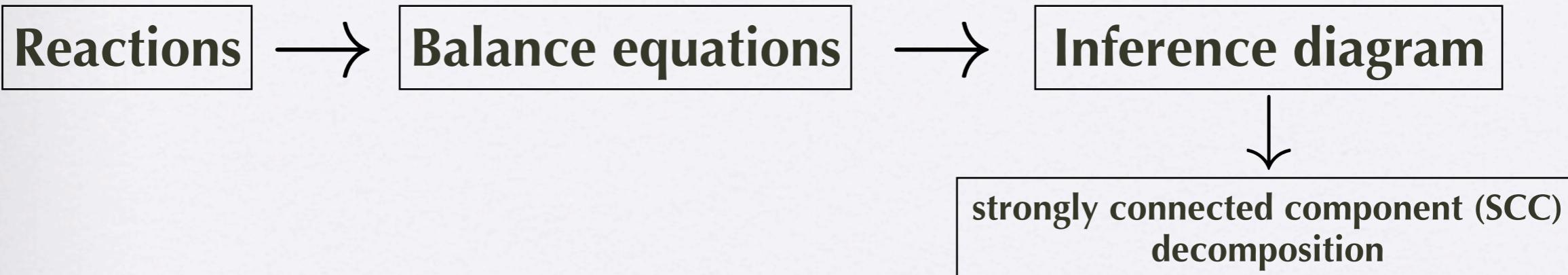
¹To whom correspondence should be addressed. E-mail: alb@neu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1215508110/DCSupplemental.

Graphical Approach



Chemical reaction system



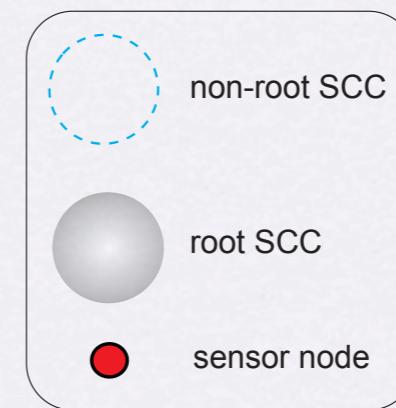
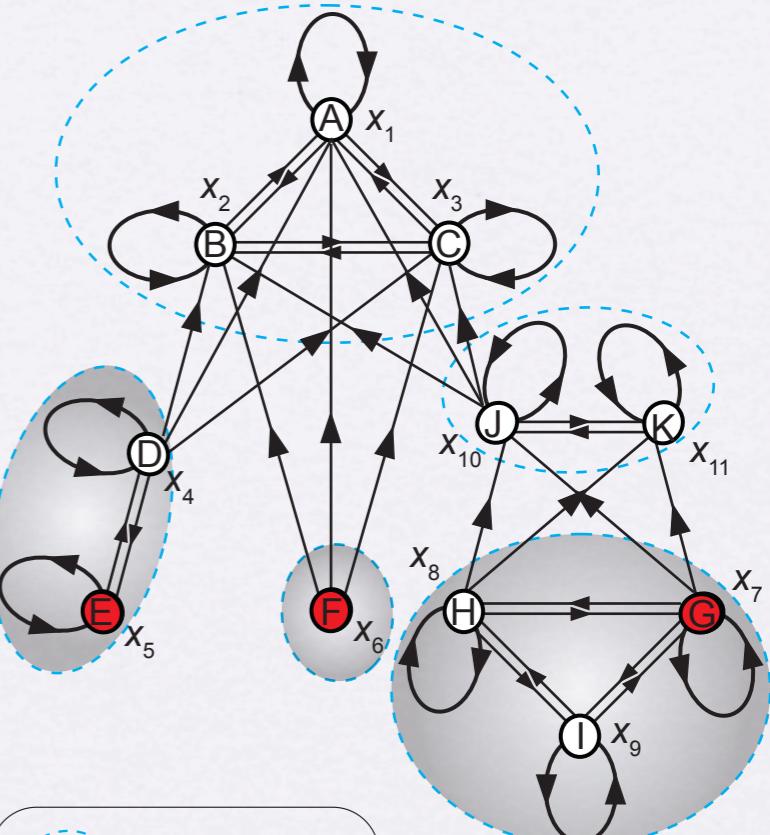
$$\left\{
 \begin{array}{lcl}
 \dot{x}_1 & = & -k_1 x_1 x_2 x_3 \\
 \dot{x}_2 & = & -k_1 x_1 x_2 x_3 \\
 \dot{x}_3 & = & -k_1 x_1 x_2 x_3 \\
 \dot{x}_4 & = & +k_1 x_1 x_2 x_3 - k_2 x_4 + k_3 x_5 \\
 \dot{x}_5 & = & +k_2 x_4 - k_3 x_5 \\
 \dot{x}_6 & = & +k_1 x_1 x_2 x_3 \\
 \dot{x}_7 & = & +k_4 x_8 x_9 - k_5 x_7 + k_6 x_{10} x_{11} \\
 \dot{x}_8 & = & -k_4 x_8 x_9 + k_5 x_7 + k_6 x_{10} x_{11} \\
 \dot{x}_9 & = & -k_4 x_8 x_9 + k_5 x_7 \\
 \dot{x}_{10} & = & +k_1 x_1 x_2 x_3 - k_6 x_{10} x_{11} \\
 \dot{x}_{11} & = & -k_6 x_{10} x_{11}
 \end{array}
 \right.$$

1. The number of sensors is lower bounded by the number of root-SCCs.

$$N_S \geq N_{r\text{SCC}}$$

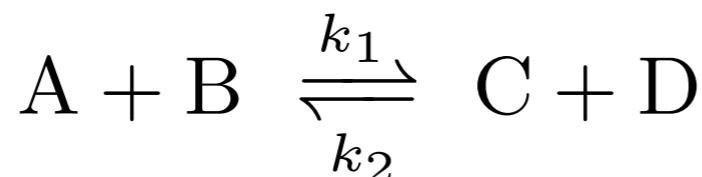
2. Monitoring one sensor node for each root-SCC is often sufficient for observability.

$$N_S = N_{r\text{SCC}} \text{ in most cases.}$$

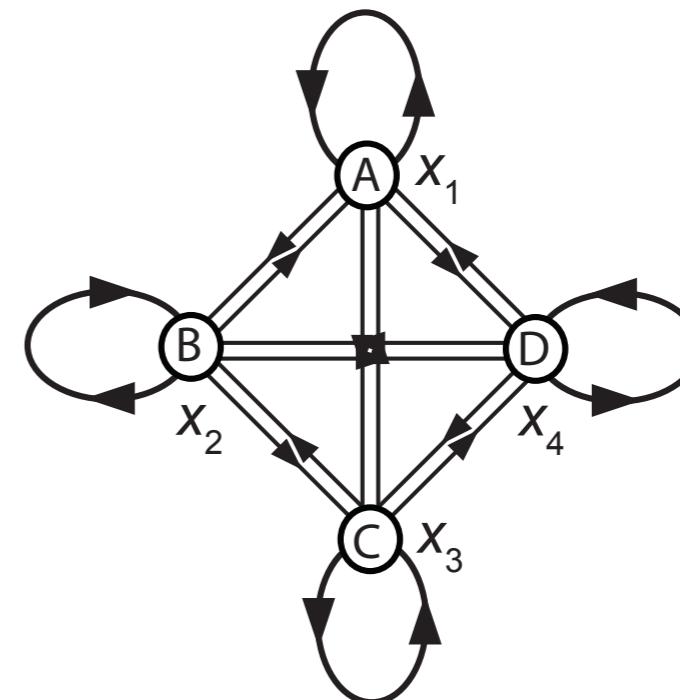


Chemical reaction system

1. Most systems are observable by monitoring one sensor for each root-SCC.
2. Counterexamples exist when there are isolated reversible reactions.



$$\left\{ \begin{array}{lcl} \dot{x}_1 & = & -k_1 x_1 x_2 + k_2 x_3 x_4 \\ \dot{x}_2 & = & -k_1 x_1 x_2 + k_2 x_3 x_4 \\ \dot{x}_3 & = & +k_1 x_1 x_2 - k_2 x_3 x_4 \\ \dot{x}_4 & = & +k_1 x_1 x_2 - k_2 x_3 x_4. \end{array} \right.$$

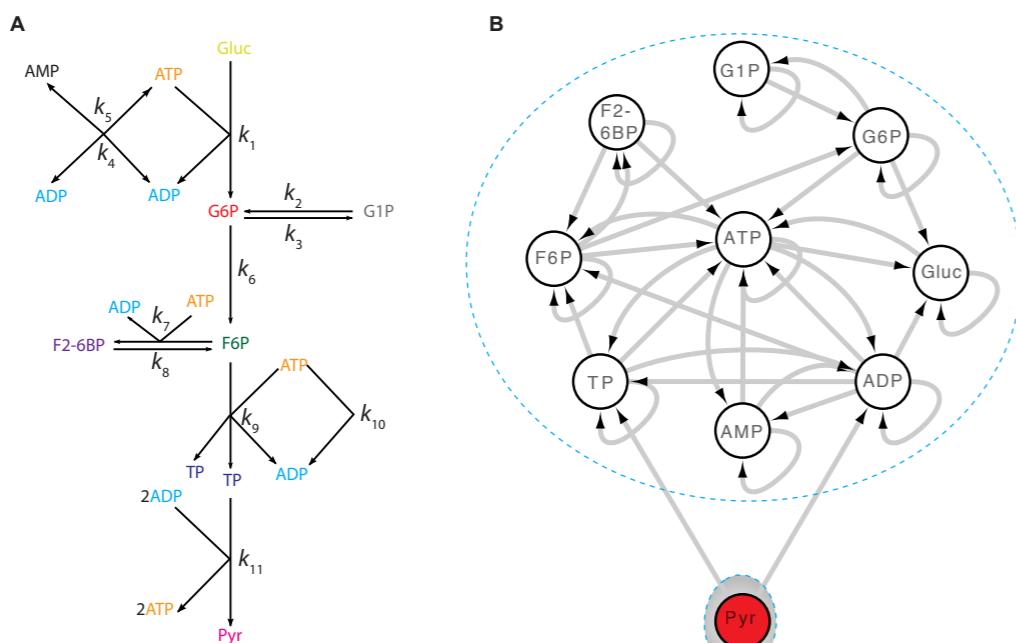


Symmetries in the state variables cause the system unobservable.
Nontrivial Lie subalgebra of model's symmetries lets the inputs and the outputs of the system be invariant ==> unobservable.

Small Biochemical Reaction Systems

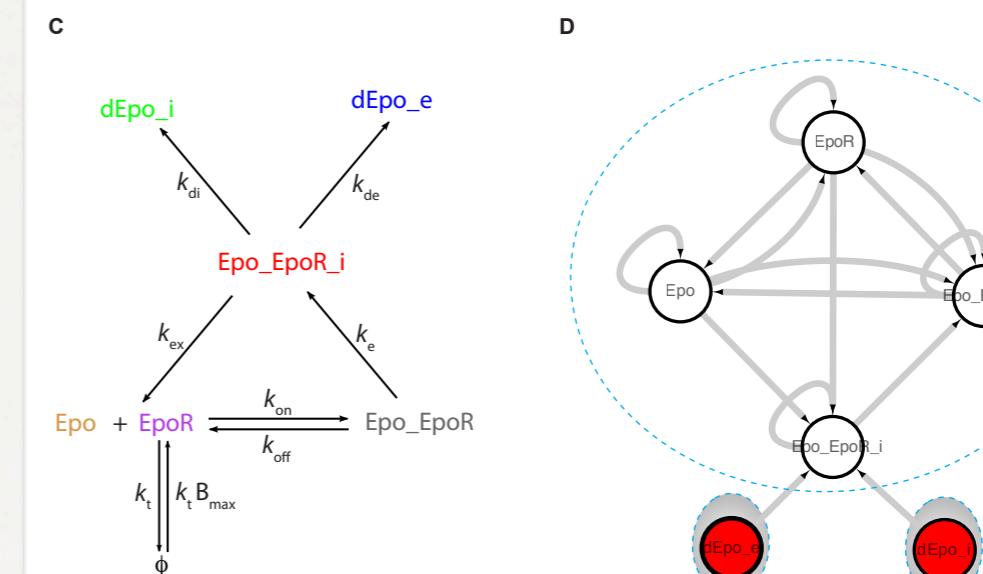
Simplified glycolytic pathway

Heinrich & Schuster, The Regulation of Cellular Systems (1996)



Model for ligand binding

Raue et al. Chaos (2010)



Small Biochemical Reaction Systems

Model for cell cycle control in fission yeast

Novak & Tyson, PNAS (1997)

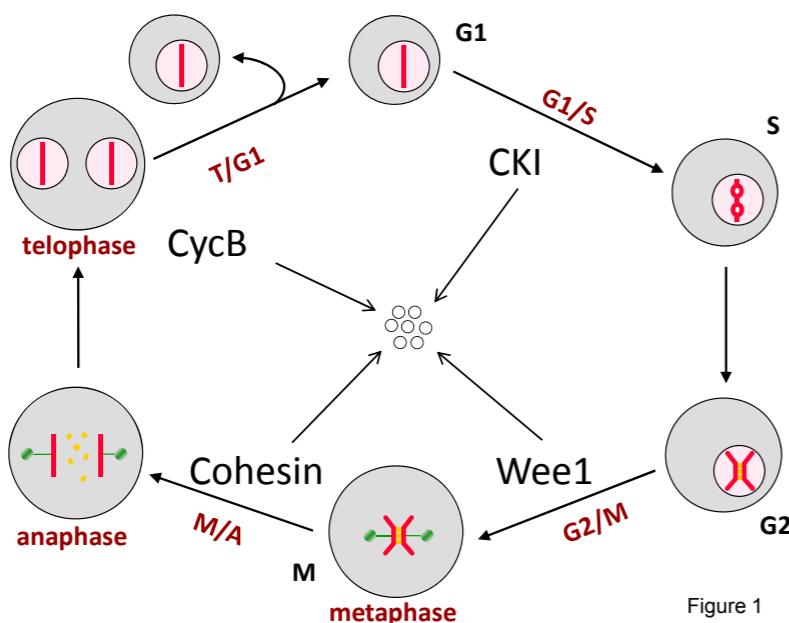
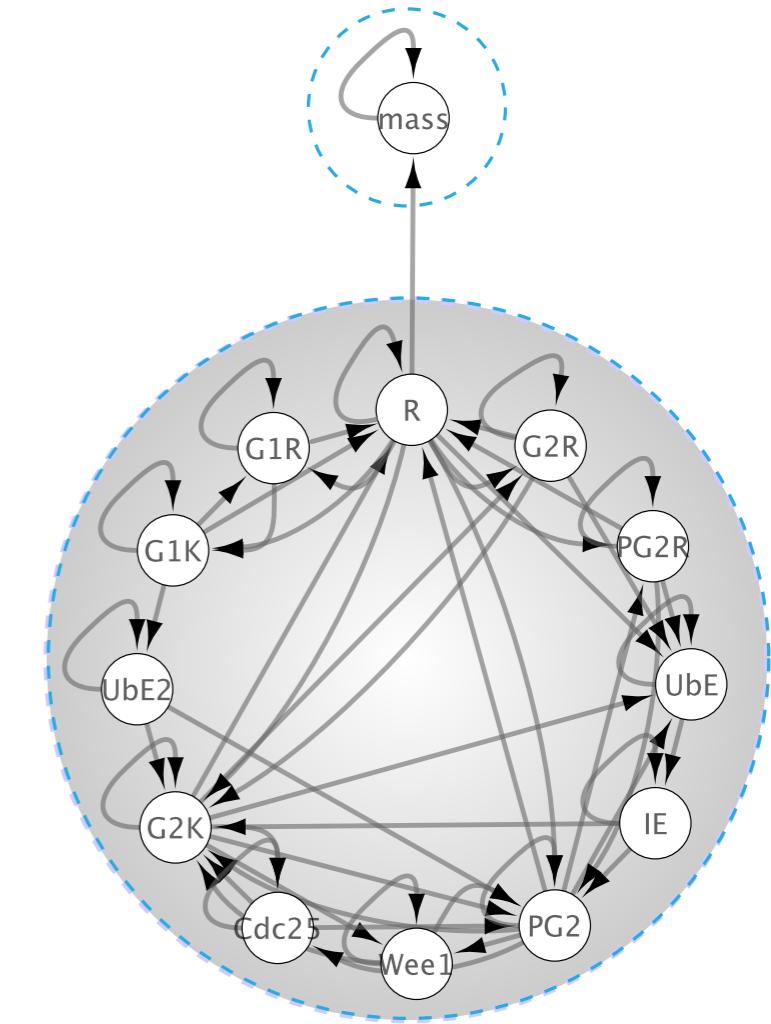


Figure 1

$$\begin{aligned}
 \frac{d[Cdc25]}{dt} &= -\frac{k_{cr}[Cdc25]}{K_{mc} + [Cdc25]} + \frac{k_c[Cdc25]([G2K] + \beta[PG2])}{K_{mc} + 1 - [Cdc25]}, \\
 \frac{d[G1K]}{dt} &= k_5 + (k_4 + k_{8r})[G1R] - k_8[G1K]R - [G1K](V_{6p}(1 - [UbE2]) + V_6[UbE2]), \\
 \frac{d[G1R]}{dt} &= -k_4[G1R] - k_{6p}[G1R] - k_{8r}[G1R] + k_8[G1K]R, \\
 \frac{d[G2K]}{dt} &= k_1 + (k_4 + k_{7r})[G2R] + (V_{2sp}(1 - [Cdc25]) \\
 &\quad + V_{25}[Cdc25])[PG2] - k_k[G2K][R] \\
 &\quad - [G2K](V_{2p}(1 - [UbE]) + V_2[UbE]) \\
 &\quad - [G2K](V_{wp}(1 - [Wee1]) + V_w[Wee1]), \\
 \frac{d[G2R]}{dt} &= -k_4[G2R] - k_{7r}[G2R] + k_7[G2K][R] \\
 &\quad - [G2R](k_{2p} + V_{2p}(1 - [UbE]) + V_2[UbE]), \\
 \frac{d[IE]}{dt} &= -\frac{k_{ir}[IE]}{K_{mir} + [IE]} + \frac{k_i[IE]([G2K] + \beta[PG2])}{K_{mi} + [IE]}, \\
 \frac{d[mass]}{dt} &= \mu [mass], \\
 \frac{d[PG2]}{dt} &= -(V_{2sp}(1 - [Cdc25]) + V_{25}[Cdc25])[PG2] \\
 &\quad + k_4[PG2R] + k_{7r}[PG2R] - k_7[PG2][R] \\
 &\quad - [PG2](V_{2p}(1 - [UbE]) + V_2[UbE]) \\
 &\quad + [G2K](V_{wp}(1 - [Wee1]) + V_w[Wee1]), \\
 \frac{d[PG2R]}{dt} &= -k_4[PG2R] - k_{7r}[PG2R] + k_7[PG2][R] \\
 &\quad - [PG2R](k_{2p} + V_{2p}(1 - [UbE]) + V_2[UbE]), \\
 \frac{d[R]}{dt} &= k_3 + k_{6p}[G1R] + k_{8r}[G1R] + k_{7r}[G2R] + k_{7r}[PG2R] \\
 &\quad - k_4[R] - k_8[G1K][R] - k_7[G2K][R] \\
 &\quad - k_7[PG2][R] \\
 &\quad - \frac{k_p[mass](Cig1) + \alpha[G1K] + [G2K] + \beta[PG2])[R]}{K_{mp} + [R]} \\
 &\quad + [G2R](k_{2p} + V_{2p}(1 - [UbE]) + V_2[UbE]) \\
 &\quad + [PG2R](k_{2p} + V_{2p}(1 - [UbE]) + V_2[UbE]), \\
 \frac{d[UbE]}{dt} &= -\frac{k_{ur}[UbE]}{K_{mur} + [UbE]} + \frac{k_u[IE](1 - [UbE])}{K_{mu} + 1 - [UbE]}, \\
 \frac{d[UbE2]}{dt} &= -\frac{k_{ur2}[UbE2]}{K_{mur2} + [UbE2]} \\
 &\quad + \frac{k_{u2}([G2K] + \beta[PG2])(1 - [UbE2])}{K_{mu2} + 1 - [UbE2]}, \\
 \frac{d[Wee1]}{dt} &= -\frac{k_w([G2K] + \beta[PG2])[Wee1]}{K_{mw} + [Wee1]} \\
 &\quad + \frac{k_{wr}(1 - [Wee1])}{K_{mwr} + 1 - [Wee1]}
 \end{aligned}$$



Genome-scale metabolic networks

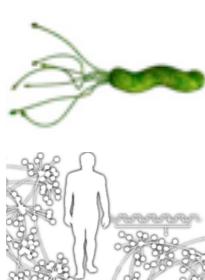
BiGG: a biochemical genetic and genomic knowledgebase of large scale metabolic reconstructions.
Schellenberger, Park, Conrad, Palsson. BMC bioinformatics (2010)



Table 1. Genome-scale Metabolic Networks

Name*	<i>N</i>	<i>L</i>	<i>R</i>	<i>N_{scc}</i>	Ω_s	<i>N_{pr}</i>	<i>N_{pp}</i>	<i>N_s</i>
<i>E. coli</i> (textbook)	72	643	153	1	72	0	0	1
<i>E. coli</i> (iJR904)	753	6,757	1,441	43	2	1	35	41
<i>E. coli</i> (iAF1260)	1,668	12,719	3,231	120	7.08×10^6	14	60	96
<i>S. cerevisiae</i> (iND750)	1,060	9,080	1,793	112	6,912	11	88	99
<i>S. aureus</i> (iSB619)	629	5,375	997	82	2	8	70	76
<i>M. barkeri</i> (iAF692)	628	5,235	955	71	256	2	60	68
<i>M. tuberculosis</i> (iNJ661)	825	7,590	1,361	73	16	5	56	65
<i>P. putida</i> (iJN746)	911	7,414	1,470	71	4	1	66	69
<i>H. pylori</i> (iT341)	485	4,046	788	32	1	0	29	30
<i>H. sapiens</i> (Recon1)	2,763	21,026	5,283	335	8.22×10^{33}	17	143	293

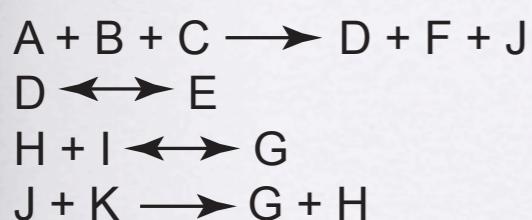
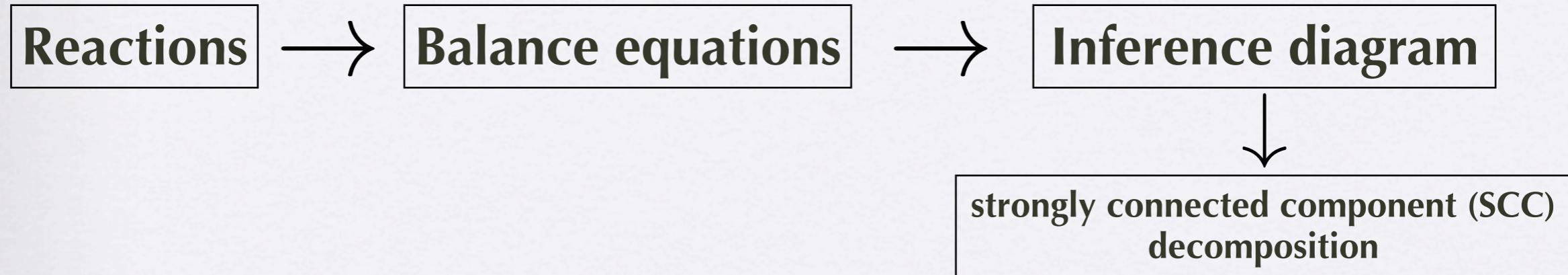
*For each metabolic network, we show the number of nodes (metabolites) *N*, edges (*L*), number of elementary reactions (*R*), the number of strongly connected components (*N_{scc}*), and the number of different sets of sensor nodes (Ω_s) in the inference diagram. The table also lists the number of pure reactants (*N_{pr}*), pure products (*N_{pp}*, that are always sensor nodes), and the minimum number of sensor nodes (*N_s*) predicted by the graphical approach.



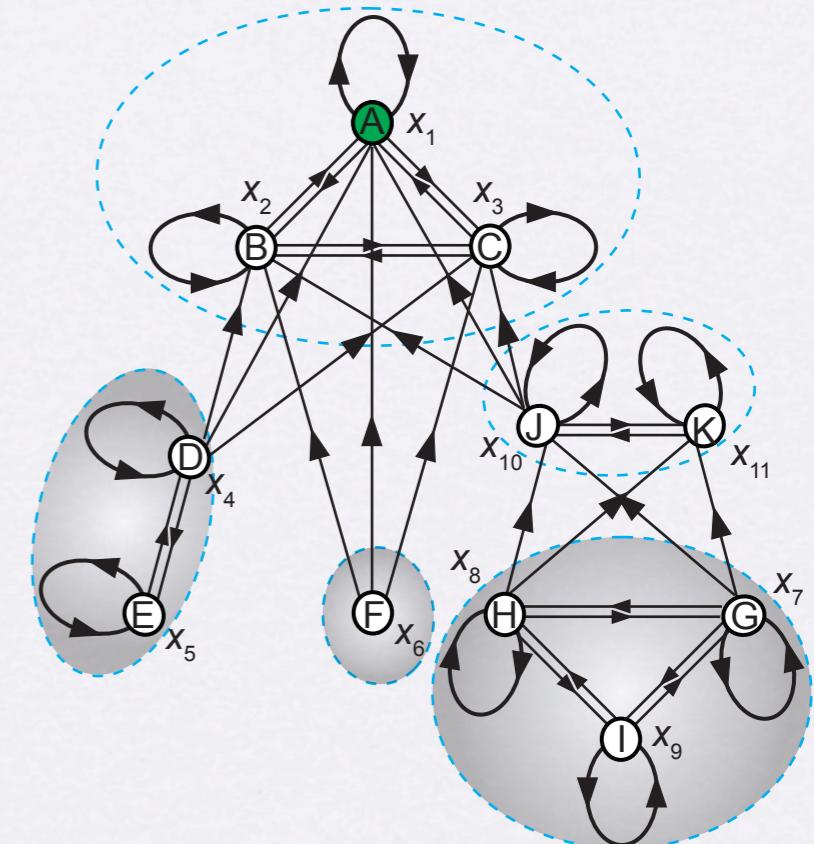
In principle one can reconstruct the state of the whole metabolism from the concentration of a relatively small fraction (5~10%) of metabolites.

Target Observability

Which sensor should we monitor if we just care about a particular metabolite's concentration?



$$\left\{ \begin{array}{l} \dot{x}_1 = -k_1 x_1 x_2 x_3 \\ \dot{x}_2 = -k_1 x_1 x_2 x_3 \\ \dot{x}_3 = -k_1 x_1 x_2 x_3 \\ \dot{x}_4 = +k_1 x_1 x_2 x_3 - k_2 x_4 + k_3 x_5 \\ \dot{x}_5 = +k_2 x_4 - k_3 x_5 \\ \dot{x}_6 = +k_1 x_1 x_2 x_3 \\ \dot{x}_7 = +k_4 x_8 x_9 - k_5 x_7 + k_6 x_{10} x_{11} \\ \dot{x}_8 = -k_4 x_8 x_9 + k_5 x_7 + k_6 x_{10} x_{11} \\ \dot{x}_9 = -k_4 x_8 x_9 + k_5 x_7 \\ \dot{x}_{10} = +k_1 x_1 x_2 x_3 - k_6 x_{10} x_{11} \\ \dot{x}_{11} = -k_6 x_{10} x_{11} \end{array} \right.$$



Target Observability

Which sensor should we monitor if we just care about a particular metabolite's concentration?

The optimal sensor for a target variable can be found by solving an optimization problem on the inference diagram.

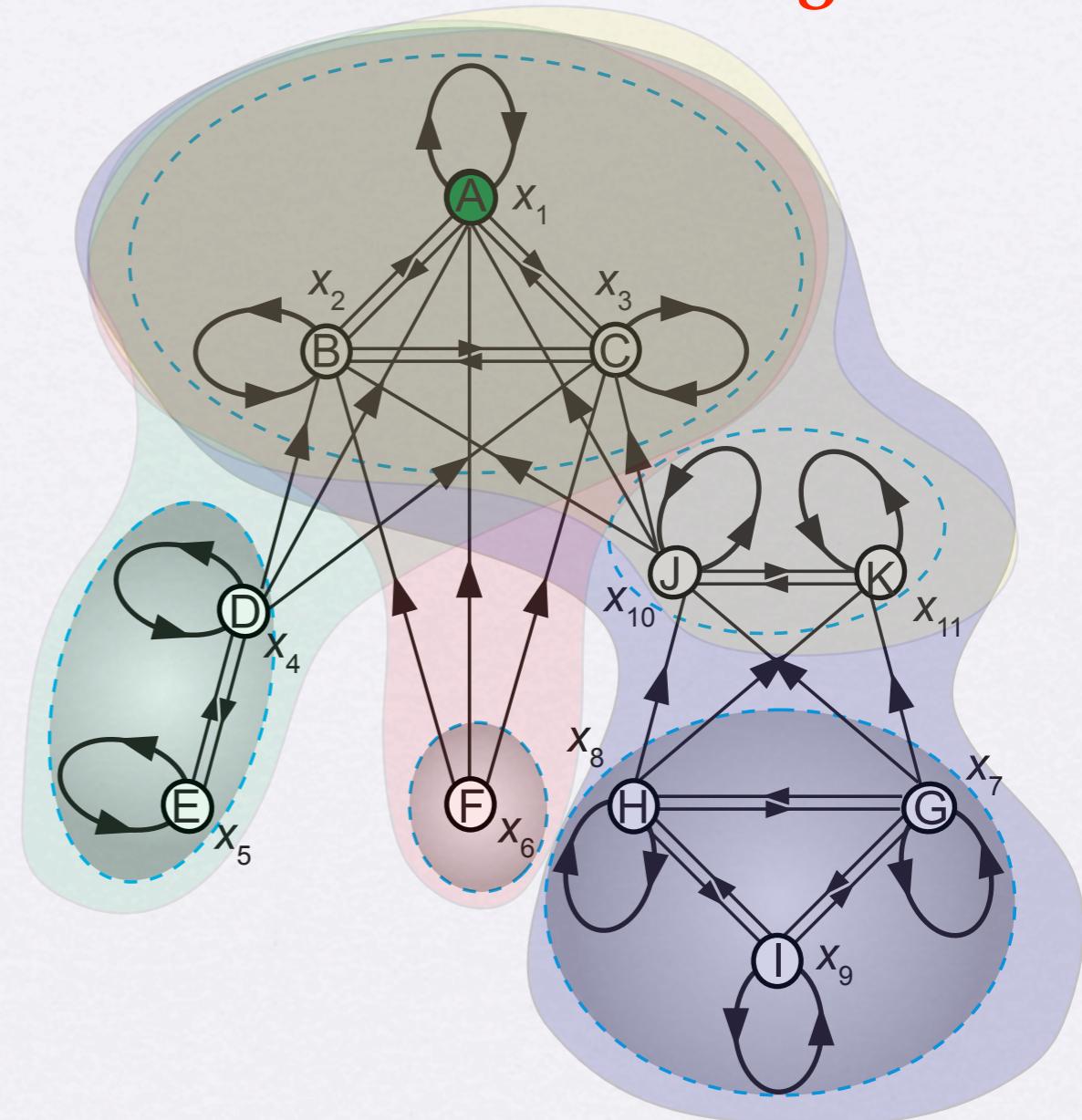
Minimize the size of the subsystem we need to observe.

$$\text{minimize} \sum_{\text{SCC} \subset \mathcal{A}_s; x_s \rightsquigarrow x_t} |\text{SCC}|$$

$x_s \rightsquigarrow x_t$: x_s can reach x_t

\mathcal{A}_s : the set of all SCCs that are reachable from x_s

$|\text{SCC}|$: size of the SCC



Summary

1. The number of sensors is lower bounded by the number of root-SCCs in the inference diagram. (True for any nonlinear dynamics.)

$$N_S \geq N_{r\text{SCC}}$$

2. Monitoring one sensor node for each root-SCC is often sufficient for observability.

- Chemical reaction systems
(Mass action; Michaelis-Menten kinetics)
- Ecological systems (Lotka-Volterra dynamics)
- Neuronal systems (Hindmarsh-Rose model)

$$N_S = N_{r\text{SCC}} \text{ in most cases.}$$

3. Target observability can be efficiently solved by using inference diagram.

Collaborators

Albert-László Barabási
NEU/DFCI/HMS

Jean-Jacques Slotine
MIT

Thank you!



JAMES S.
MCDONNELL
FOUNDATION