# Judging model reduction of complex systems

Jie Sun,[1,*] Erik M. Bollt,[2,†] and Takashi Nishikawa[2,‡]
[1]*Department of Physics & Astronomy, Northwestern University, Evanston, IL 60208*
[2]*Department of Mathematics & Computer Science, Clarkson University, Potsdam, New York 13699-5815, USA*
(Received 5 September 2010; published 27 April 2011)

Model reduction is a common goal in the study of complex systems, consisting of many components with a complex interaction structure. The quality of such reduction, however, may not be reflected correctly in the stepwise prediction error in the model since it ignores the global geometry of the dynamics. Here we introduce a general two-step framework, consisting of dimensionality reduction of the time series followed by modeling of the resulting time series, and propose the use of the shadowing distance to measure the quality of the second step. Using coupled oscillator networks as a prototypical example, we demonstrate that our approach can outperform those based on stepwise error and suggest that it sheds light on the problem of identifying and modeling low-dimensional dynamics in large-scale complex systems.

DOI: 10.1103/PhysRevE.83.046125 PACS number(s): 89.75.−k, 05.45.Tp, 05.45.Xt

## I. INTRODUCTION

Model reduction is an important concept found across various fields of science and engineering. Finding a low-dimensional model that captures the gross features of a high-dimensional system is a fundamental problem in the physics of complex systems, which touches upon such disparate fields as partial differential equations [1], dynamical systems [2], and network theory [3]. Consider, for example, a complex system modeled by a network of coupled oscillators [4,5], which may support processes such as disease spreading [6], the evolution of a food web [7], or the dynamics of a power grid [8]. It is often useful and desirable in such a system to average across parts so that a system with only a few dynamical units is sufficient to model the large-scale dynamical behavior of the original system. In much the same way that community detection methods partition the nodes of a complex network into groups based on topological similarity [9,10], dynamical units in a complex system may be grouped based on dynamical similarity.

For a given high-dimensional system, there are often many different ways to obtain a low-dimensional reduced model, which leads to the natural and fundamental question of how to choose the best one. For example, is it best to simply average the equations for individual units to obtain a reduced model? Would it be better to use a weighted average of the units reflecting their various roles within the system or to introduce an extra component into the model to compensate for the loss of information due to dimensionality reduction of the time series? Addressing these questions requires quantifying the quality of model reduction.

For a given time series, the best model of the same dimension as the time series is commonly selected based on the least-squares (LS) criterion, which minimizes the sum of squared stepwise errors. For chaotic systems, however, this criterion can be inappropriate or even misleading [11–13], due to the nature of the cost function, which is based on local, rather than global and geometric, features of the system. If reduced models of lower dimension are considered, the LS criterion can still be formulated, but its appropriateness is even more questionable.

## II. METHOD

We are unaware of previous work addressing this problem of quantifying the quality of model reduction beyond the LS criterion. In this paper we introduce a framework (illustrated in Fig. 1) that divides the problem into two steps: (i) finding a low-dimensional representation of the time series generated by the original system and (ii) finding a model that best describes the reduced-order time series. This model will then be considered a reduced-order model for the original system. Our framework allows for a separate treatment of dimensionality reduction and time-domain modeling, two aspects of the problem that can involve model deficiency of a very different nature. Using $\eta$ and $\epsilon$ to represent the error caused in the first and second steps, respectively (to be discussed in detail below), the quality of a model $f_a$ can be quantified as

$$J(f_a) = (1-\mu)\eta + \mu\epsilon, \tag{1}$$

where $\mu \in [0,1]$ can be chosen (by the modeler) to emphasize either part of the model reduction process.

The first step is essentially a problem of finding a low-dimensional manifold that best fits the original time series as a set of points in high-dimensional state space. Once a best-fit manifold is found, the loss of information $\eta$ can be quantified by measuring the amount of residual, i.e., deviation from the manifold. There are several relatively well-developed methods for this problem, such as the principal component analysis [2] for finding linear manifolds and ISOMAP [14] for finding nonlinear manifolds.

In contrast, the second step is a subtle problem that appears largely unexplored in the literature. To simplify our discussion, consider a scalar time series produced by an unknown one-dimensional map, where the goal is to measure the quality of a given map as a model for the unknown map. On the one hand, a small stepwise error (or, equivalently, a small difference in the two maps as functions) may not imply that the model is good since the error can accumulate over time, leading to a poor long-term prediction. This is particularly problematic for

*sunj@northwestern.edu
†bolltem@clarkson.edu
‡tnishika@clarkson.edu

 

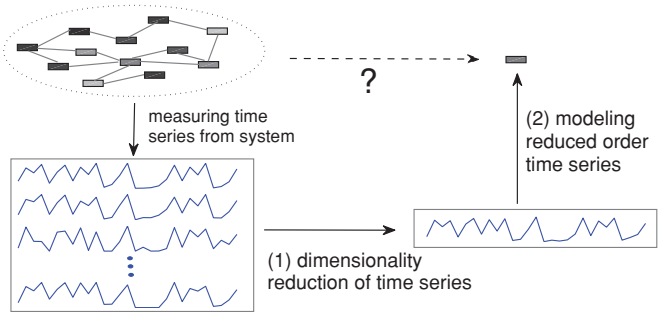


FIG. 1. (Color online) Two-step approach for judging model reduction of large complex systems. Since direct comparison is impossible due to the difference in dimensionality (top horizontal arrow), we first find a low-dimensional representation of the measured time series (bottom horizontal arrow) and then model the resulting time series (up arrow). Here we propose the use of the shadowing distance (defined in the text) to measure the quality of the second step.

a chaotic system due to the sensitivity to initial conditions. On the other hand, for a chaotic system large accumulated error does not necessarily mean that the model is poor since this can occur even when the two maps match perfectly, due to sensitivity to perturbations.

To avoid this fundamental difficulty, we propose a measure of model quality based on the concept of shadowing [15–17], which was originally designed to test whether a simulated time series suffering from numerical inaccuracy could have come from a given model system. The shadowing distance $\epsilon_{\mathrm{SD}}$ for a given finite-length time series is defined as the smallest distance within which a model trajectory can follow the time series. This can be regarded as a measure of the level of confidence in saying that the noisy time series came from the model and thus as a measure of how good the model is for the time series. (Another possibility is to use the distribution of shadowing times. See Ref. [13] for details.) Symbolically, given a time series $\{x_t\}_{t=1}^{T}$, we define the shadowing distance for a given model $f$ as

$$\epsilon_{\mathrm{SD}}(f) \equiv \inf_{y_1 \in D} \max_{1 \leqslant t \leqslant T} ||x_t - y_t||, \tag{2}$$

where $D$ is the state space, $y_t$ for $t \geqslant 2$ is given recursively by $y_{t+1} = f(y_t)$, and $||\cdot||$ denotes the Euclidean distance. This definition can be naturally extended to continuous-time systems.

To estimate $\epsilon_{\mathrm{SD}}$ we find a numerical shadowing trajectory using a procedure developed for systems of arbitrarily high dimension, in both discrete time [18] and continuous time [19]. Starting with the given time series, we iteratively generate an incrementally less noisy trajectory that still stays close to the time series. This leads to a trajectory that has a stepwise error within machine precision and whose distance to the time series is likely to be near its minimum possible value $\epsilon_{\mathrm{SD}}$. Since we do not expect a long shadowing trajectory for a very noisy time series, such as a reduced time series resulting from (nonlinear) dimensionality reduction, we estimate $\epsilon_{\mathrm{SD}}$ for all possible time-series segments of length $T_s \leqslant T$. To avoid the adverse effect of rare, nonshadowable segments (i.e., $\epsilon_{\mathrm{SD}} = \infty$), we discard the largest 10% of the estimated $\epsilon_{\mathrm{SD}}$ values and compute the average over the remaining values. This choice

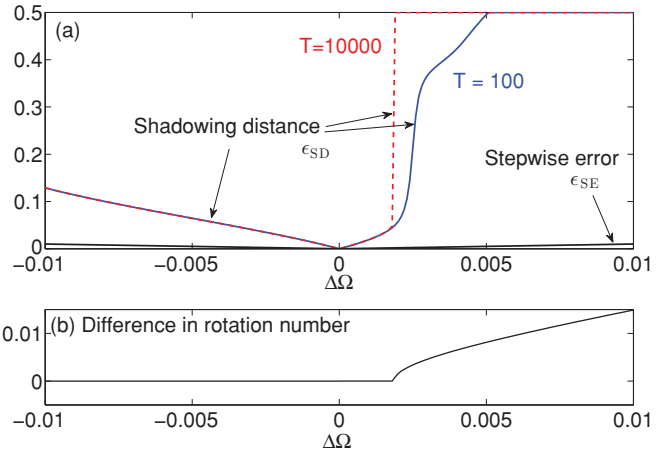


FIG. 2. (Color online) Shadowing distance $\epsilon_{\mathrm{SD}}$ vs stepwise error $\epsilon_{\mathrm{SE}}$ for judging model quality. A periodic time series $\{x_t\}_{t=1}^{T}$ was generated by the standard circle map [20], $x_{t+1} = x_t + \Omega - 0.12\sin(2\pi x_t)$, with $\Omega = 0.35$ and a random choice of $x_1$. Using the same map with $\Omega = 0.35 + \Delta\Omega$ as a model for this time series, we have $\epsilon_{\mathrm{SE}} = |\Delta\Omega|$ [lower solid curve in (a)]. Despite the small $\epsilon_{\mathrm{SE}}$, error can accumulate over time and reach 0.5 (the maximum possible error on the unit circle) for $\Delta\Omega \gtrsim 0.002$ due to (b) the difference in the rotation number for the two maps. In contrast, $\epsilon_{\mathrm{SD}}$ [solid and dashed curves in (a), computed with $T_s = T$] correctly reflects this effect, as well as the dependence on $T$ and asymmetry with respect to $\Delta\Omega$, none of which is captured at all by $\epsilon_{\mathrm{SE}}$.

of 10% is a conservative one to guarantee that the artificially large values of $\epsilon_{\mathrm{SD}}$ at glitch points do not affect the measure of model quality of typical segments of the time series, which is consistent with the quantile statistics suggested in Ref. [13], where shadowing time is considered. An appropriate choice of $T_s$ is $T_s \approx \log(\delta^{-1})/\lambda$, where $\delta$ is the machine precision and $\lambda$ is the maximum Lyapunov exponent of the system, since a longer trajectory is likely to suffer from the accumulation of computational error.

A critical feature of $\epsilon_{\mathrm{SD}}$ is that it correctly reflects the model's ability to produce a trajectory that closely follows the entire time series. In contrast, the stepwise error, which can be measured by $\epsilon_{\mathrm{SE}}(f) \equiv \sqrt{\sum_{t=1}^{T} ||f(x_t) - x_{t+1}||^2/T}$ for a given time series, may be small even when large error accumulation along the time series is unavoidable. This is clearly illustrated in Fig. 2 for a nonchaotic time series from the standard circle map. The contrast would be even more dramatic for chaotic systems. Based on this observation, we propose to use the optimal shadowing (OS) criterion, in which we select a model with the smallest shadowing distance, instead of one with smallest $\epsilon_{\mathrm{SE}}$ (the LS criterion).

## III. EXAMPLES OF APPLICATION

To demonstrate that our approach is well suited to addressing fundamental questions on model reduction, consider a general network of coupled discrete-time dynamical systems described by

$$x_{t+1}^{(i)} = g\left(x_t^{(i)}, a^{(i)}\right) - (\sigma/c_i)\sum_{j=1}^{n} \ell_{ij}\, g\left(x_t^{(j)}, a^{(j)}\right), \tag{3}$$

where $\{x^{(i)}\}_{i=1}^{n}$, with $x_t^{(i)} \in \mathbb{R}^d$, represents the states of $n$ oscillators at time $t$, the function $g(x,a)$ describes the dynamics driving individual oscillators (which can be different and are parametrized here by $a$), and $\sigma$ is the global coupling strength. The effective diffusive coupling among the oscillators is represented by the discrete Laplacian matrix, obtained from the adjacency matrix $B = [b_{ij}]$ of the network by setting $\ell_{ij} = -(1-\delta_{ij})b_{ij} + \delta_{ij}\sum_k b_{ik}$, where $\delta_{ij}$ is the Kronecker delta function. The coupling function has been chosen to have the same form as the individual dynamics $[g(x,a)]$, which corresponds to the situation where each oscillator receives a direct signal from the output of its neighbors. The normalization by $c_i = \sum_j b_{ij}$ scales the effective coupling strength at node $i$ to keep the trajectory of individual dynamics within its (bounded) domain.

As a concrete example consider networks of logistic maps: $g(x,a) = ax(1-x)$. We generate random directed networks as follows. Starting with two nodes and one directed link from node 2 to node 1, at each step, a new node $k \geqslant 3$ is added to the network with directed links pointing to $d$ ($\geqslant 2$) existing nodes, chosen at random. Each new directed link is assigned weight 1 and the associated backward link is created with weight 0.05. For each network, the coupling strength is chosen to be $\sigma = 1$ and the node parameters are set as $a^{(i)} = 3.96 + (i/n)0.04$, filling the interval [3.96,4] uniformly. With this set of parameters, the oscillators are nearly synchronized [21], i.e., the trajectory of any one oscillator stays close to that of any other oscillator.

For a given network generated as above, we ask if there is a simplified system that can model this $n$-dimensional coupled chaotic system. Since the oscillators are nearly synchronized, a reasonable guess might be a one-dimensional model $f(x) = \bar{a}x(1-x)$, where $\bar{a} = \sum_i a^{(i)}/n$ is the average parameter. Such uniform averaging, however, does not respect the nonhomogeneity of the network, resulting from the fact that the oscillators introduced at an early stage of network generation have a much larger influence than the others. Since the parameters for these nodes are closer to 3.96, we expect that an optimal value for the parameter $a$ in the family of candidate models $f_a(x) = ax(1-x)$ would be smaller than the average $\bar{a}$.

To find this optimal value we first consider a randomly generated network with $n = 500$ and $d = 8$. In Fig. 3(a) we compare the candidate models using both stepwise error ($\epsilon_{\mathrm{SE}}$) and shadowing distance ($\epsilon_{\mathrm{SD}}$). For the reduced time series, we take the average trajectory, $x_t = \sum_i x_t^{(i)}/n$ (we will discuss other possibilities below). The two criteria give clearly distinct values of $a$ as the best choice. In Fig. 3(b) we show the difference between the estimate of the largest Lyapunov exponent of the original system and that from $f_a(x)$. In Fig. 3(c) we plot the prediction error as a function of $a$ for the prediction of five steps away from present. The curves in both Figs. 3(b) and 3(c) follow closely the shadowing distance shown in Fig. 3(a) instead of the stepwise error, providing clear evidence that the model quality in this case is better and naturally judged by the proposed OS criterion. By varying the network parameter $d$ to test the generality of the proposed approach we see in the inset of Figs. 3(b) and 3(c) that the models chosen by the OS criterion clearly outperform those chosen by the LS criterion in the range of $d$ considered.

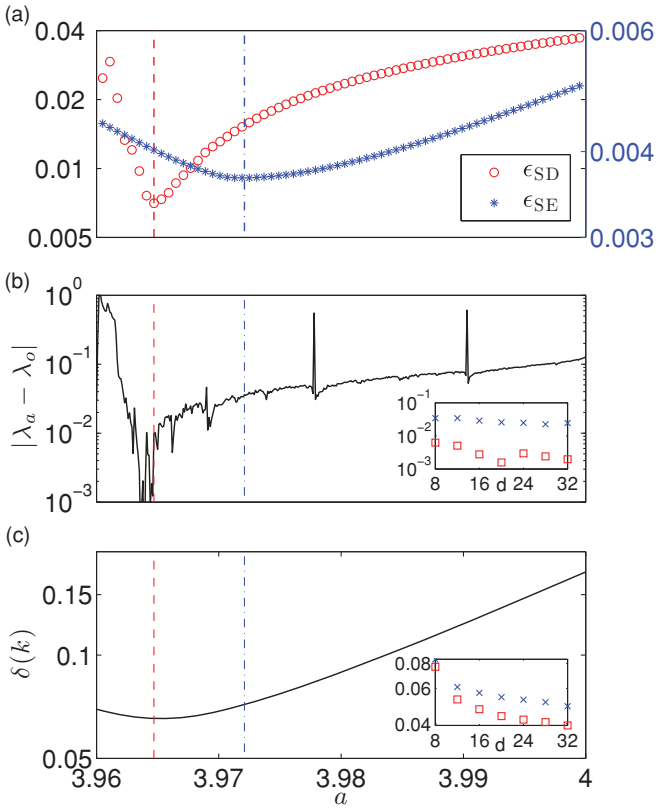


FIG. 3. (Color online) Judging the quality of $f_a(x) = ax(1-x)$ as the reduced model for a large network of coupled nonidentical logistic maps (details in the main text). (a) Shadowing distance ($\epsilon_{\mathrm{SD}}$) and stepwise error ($\epsilon_{\mathrm{SE}}$). We used $T_s = 50$ and $T = 5000$. (b) Difference between the maximum Lyapunov exponent of the original system ($\lambda_o$) and of $f_a$ ($\lambda_a$). (c) The $k$-step prediction error, defined as $\delta(k) = |f^k(x_0) - x_k|$, where $f$ is the model and $\{x_t\}$ is the time series. Here we plot the curve for $k = 5$. The insets of (b) and (c) show the estimation error of the maximum Lyapunov exponent and the prediction error, respectively, with $k = 5$ for the OS-based models (□ on the bottom) and LS-based models (× on the top). We used random directed networks of size $n = 500$ and $d = 8,12,\ldots,32$. Each data point is an average over 20 independent realizations. In all panels the vertical dashed line and the vertical dash-dotted line correspond to the values of $a$ that minimize the shadowing distance $\epsilon_{\mathrm{SD}}$ and the stepwise error $\epsilon_{\mathrm{SE}}$, respectively.

We now address the question of whether the average trajectory is the best choice as a reduced time series. It is indeed the best if we minimize only the dimensionality reduction error [$\mu = 0$ in Eq. (1)], but allowing a slightly larger reduction error might reduce the shadowing error $\epsilon_{\mathrm{SD}}$ significantly. To explore this interplay between the two components of the model reduction process and how their relative emphasis influences the choice of the best model, consider the same networked logistic map system in for Fig. 3. We now consider weighted average of the individual trajectories: $x_t = \sum_i w^{(i)} x_t^{(i)}$, where $w^{(i)} = k_i^{\beta}/\sum_j k_j^{\beta}$ for given $\beta$. Here $k_i$ is the out-degree of

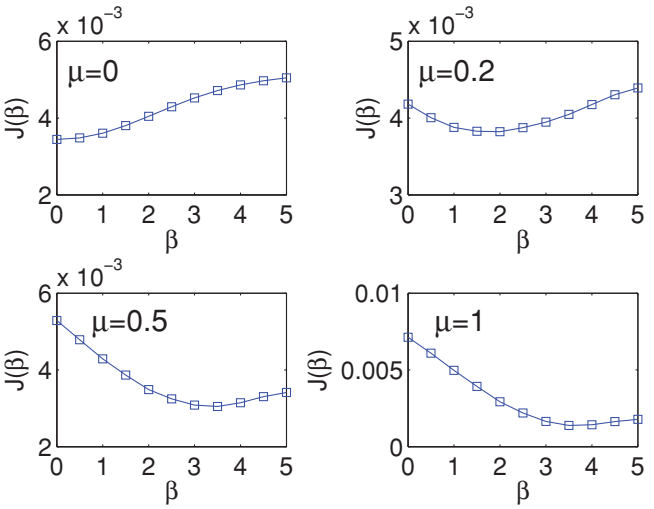


FIG. 4. (Color online) Interplay between the dimensionality reduction error $\eta(\beta)$ and the optimal shadowing distance $\epsilon_{\rm SD}(\beta)$ for the system used in Fig. 3(a). For each fixed value of $\mu$, the model reduction error $J(\beta)$ is plotted as a function of the parameter $\beta$, which determines the weights for averaging the time series (see the main text for details).

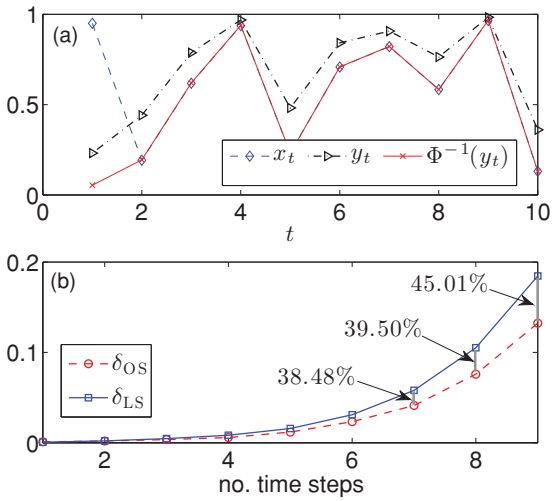


FIG. 5. (Color online) (a) Segments of the time series $x_t$ and $y_t$. Also plotted is the transformed time series $\Phi^{-1}(y_t)$, which stays close to $x_t$. (b) Prediction error (defined in the caption of Fig. 3) for the LS-based optimal model ($\delta_{\rm LS}$) and for the OS-based optimal model ($\delta_{\rm OS}$) with $T_s = 50$ and $T = 5000$.

node $i$. The dimensionality reduction error is then computed as $\eta(\beta) = \sqrt{\sum_i \sum_t |x_t - x_t^{(i)}|^2/nT}$. We define the shadowing distance $\epsilon_{\rm SD}(\beta)$ for a given $\beta$ to be the minimum value of $\epsilon_{\rm SD}$ over the parameter $a$ of the reduced model $f(x,a) = ax(1-x)$. In Fig. 4 we show the $\beta$ dependence for the model reduction error $\zeta(\beta) = (1-\mu)\eta(\beta) + \mu\epsilon_{\rm SD}(\beta)$, for different values of $\mu$, which controls the relative emphasis on the two parts. If $\mu = 0$ (upper left panel of Fig. 4), the best choice is $\beta = 0$ (uniform weights), as noted above. When we start to focus on the quality of the model for the reduced time series ($\mu > 0$, the three other panels in Fig. 4), giving more weight to nodes with a larger out-degree yields a smaller model reduction error.

To show that our OS criterion can be effective even when individual trajectories are not necessarily close to each other, but related nonlinearly, consider the simplest case of a two-unit system:

$$\begin{aligned} x_{t+1} &= f(x_t) + \sigma\{\Phi^{-1}[g(y_t)] - f(x_t)\}, \\ y_{t+1} &= g(y_t) + \sigma\{\Phi[f(x_t)] - g(y_t)\}. \end{aligned} \tag{4}$$

Here we choose $f(x) = 3.96x(1-x)$, $g(y) = \sqrt{4y(1-y)}$, $\sigma = 0.3$, and $\Phi(x) = \sqrt{x}$. Although a naive comparison of $x_t$ and $y_t$ might lead to the conclusion that no model reduction is possible, nonlinear reduction techniques can be used to find the intrinsic coherence between $x_t$ and $y_t$ [Fig. 5(a)]. A one-dimensional model for the single trajectory $\{x_t\}$ is thus sufficient to model the system. Considering the family of models of the form $f_a(x) = ax(1-x)$, Fig. 5(b) compares the predictive power of the optimal models selected by the LS and OS criteria. It clearly shows that the error in making seven- to ten-step predictions is as much as 45% more for the LS-selected model than for the OS-selected model.

## IV. CONCLUSION AND OPEN PROBLEMS

To summarize, we have proposed a general approach for judging the quality of low-dimensional reduced models for high-dimensional complex systems. A key to our approach is the decomposition of the problem into spatial and temporal domains where the modeling error can be assessed separately. Another prominent feature is the use of a shadowing-based criterion in the temporal domain, which can be combined with any nonlinear dimensionality reduction technique in the spatial domain. We have shown that the models selected based on our criterion are better than those selected by the traditional LS criterion, in terms of both the Lyapunov exponent estimates and short-term prediction error.

Using examples of random directed networks, we have provided evidence that this approach is useful in addressing fundamental questions about model reduction for large networks of nearly synchronized oscillators, especially when there is dynamical inhomogeneity among the nodes and directionality in the link structure. We emphasize that our framework applies readily to systems in which individual oscillators have a multidimensional state space and connected through arbitrary network structure.

Open problems that can now be addressed using our shadowing-based measure of model quality include the construction of an optimal model from time series when explicit equations for the dynamics are not known. One possible procedure is to select generic basis functions and search for an optimal model among all linear combinations using the OS criterion. We expect to see many cases in which the proposed criterion resolves the differences within the model that cannot be resolved by previous criteria.

## ACKNOWLEDGMENT

E.M.B. and J.S. were supported by the Army Research Office under Grant No. 51950-MA.

[1] P. Holmes, J. L. Lumley, and G. Berkooz, *Turbulence, Coherent Structures, Dynamical Systems and Symmetry* (Cambridge University Press, Cambridge, 1998).
[2] A. C. Antoulas, *Approximation of Large-Scale Dynamical Systems* (Society for Industrial and Applied Mathematics, Philadelphia, 2005).
[3] S. Itzkovitz, R. Levitt, N. Kashtan, R. Milo, M. Itzkovitz, and U. Alon, Phys. Rev. E **71**, 016127 (2005).
[4] A. Pikovsky, M. Rosenblum, and J. Kurths, *Synchronization: A Universal Concept in Nonlinear Sciences* (Cambridge University Press, Cambridge, 2001).
[5] A. Arenas, A. Diaz-Guilera, J. Kurths, Y. Moreno, and C. Zhou, Phys. Rep. **469**, 93 (2008).
[6] S. H. Strogatz and I. Stewart, Sci. Am. **269**(6), 102 (1993).
[7] M. D. Holland and A. Hastings, Nature (London) **456**, 792 (2008).
[8] P. M. Anderson and A. A. Fouad, *Power System Control and Stability* (IEEE, Piscataway, 1994).
[9] M. E. J. Newman, Proc. Natl. Acad. Sci. **103**, 8577 (2006).
[10] M. A. Porter, J.-P. Onnela, and P. J. Mucha, Notices Amer. Math. Soc. **56**, 1082 (2009).
[11] P. E. McSharry and L. A. Smith, Phys. Rev. Lett. **83**, 4285 (1999).
[12] K. Judd and T. Stemler, Phys. Rev. E **79**, 066206 (2009).
[13] L. A. Smith, M. C. Cuellar, H. Du, and K. Judd, Phys. Lett. A **374**, 2618 (2010).
[14] J. B. Tenenbaum, V. Silva, and J. C. Langford, Science **290**, 2319 (2000).
[15] D. V. Anosov, Proc. Steklov Inst. Math. **90**, 1 (1967).
[16] R. Bowen, J. Diff. Eq. **18**, 333 (1975).
[17] C. Grebogi, S. M. Hammel, J. A. Yorke, and T. Sauer, Phys. Rev. Lett. **65**, 1527 (1990).
[18] G. D. Quinlan and S. Tremaine, Mon. Not. R. Astron. Soc. **259**, 505 (1992).
[19] W. Hayes and K. R. Jackson, Appl. Numer. Math. **53**, 299 (2005).
[20] J. Hale and H. Kocak, *Dynamics and Bifurcations*, 3rd ed. (Springer-Verlag New York, 1991).
[21] J. Sun, E. M. Bollt, and T. Nishikawa, Europhys. Lett. **85**, 60011 (2009).