

Early Predictor for the Onset of Critical Transitions in Networked Dynamical Systems

Zijia Liu^{1,2}, Xiaozhu Zhang^{1,2,3}, Xiaolei Ru^{1,2}, Ting-Ting Gao^{1,2}, Jack Murdoch Moore^{1,2,*} and Gang Yan^{1,2,†}

¹MOE Key Laboratory of Advanced Micro-Structured Materials, and School of Physical Science and Engineering, Tongji University, Shanghai, People's Republic of China

²National Key Laboratory of Autonomous Intelligent Unmanned Systems, and MOE Frontiers Science Center for Intelligent Autonomous Systems, Tongji University, Shanghai, People's Republic of China

³Chair for Network Dynamics, Center for Advancing Electronics Dresden (cfaed) and Institute for Theoretical Physics, Technical University of Dresden, 01062 Dresden, Germany

 (Received 7 December 2023; revised 12 April 2024; accepted 23 May 2024; published 15 July 2024)

Numerous natural and human-made systems exhibit critical transitions whereby slow changes in environmental conditions spark abrupt shifts to a qualitatively distinct state. These shifts very often entail severe consequences; therefore, it is imperative to devise robust and informative approaches for anticipating the onset of critical transitions. Real-world complex systems can comprise hundreds or thousands of interacting entities, and implementing prevention or management strategies for critical transitions requires knowledge of the exact condition in which they will manifest. However, most research so far has focused on low-dimensional systems and small networks containing fewer than ten nodes or has not provided an estimate of the location where the transition will occur. We address these weaknesses by developing a deep-learning framework which can predict the specific location where critical transitions happen in networked systems with size up to hundreds of nodes. These predictions do not rely on the network topology, the edge weights, or the knowledge of system dynamics. We validate the effectiveness of our machine-learning-based framework by considering a diverse selection of systems representing both smooth (second-order) and explosive (first-order) transitions: the synchronization transition in coupled Kuramoto oscillators; the sharp decline in the resource biomass present in an ecosystem; and the abrupt collapse of a Wilson-Cowan neuronal system. We show that our method provides accurate predictions for the onset of critical transitions well in advance of their occurrences, is robust to noise and transient data, and relies only on observations of a small fraction of nodes. Finally, we demonstrate the applicability of our approach to real-world systems by considering empirical vegetated ecosystems in Africa.

DOI: [10.1103/PhysRevX.14.031009](https://doi.org/10.1103/PhysRevX.14.031009)

Subject Areas: Complex Systems, Nonlinear Dynamics

I. INTRODUCTION

In complex systems throughout nature, technology, and society, smooth variation of parameters can result in sudden transitions between system states exhibiting significantly different behavior [1,2]. Examples of these critical transitions include rapid global warming at the end of glacial periods [3], desertification [4], extinction events of animal or bacteria species [5–7], development of psychiatric disorders [8], allograft rejection [9], power outages [10], laser thresholds [11], the onset of convection [11], transitions in human movement [12], and many other examples [13–16]. Critical

transitions have the potential to shift a system into a new state with undesirable properties [17,18], leading to environmental damage, economic loss, and public health problems if timely corrective measures are not implemented. Thus, in numerous applications, forecasting critical transitions is important for preventing or avoiding damage [19–21].

In response to this need, various early warning signals (EWSs) have been proposed to anticipate impending critical transitions [22–27]. A popular approach exploits a phenomenon known as critical slowing down [28], whereby a dynamical system approaching a tipping point exhibits slower recovery from local perturbations [29,30]. Critical slowing down, and, thus, impending critical transitions, can be detected by increases in generic typical EWSs [29], such as rising lag-1 autocorrelation [5] and variance [31,32]. A different approach leverages deep learning to identify, in addition, whether the approaching transition is fold, Hopf, or transcritical bifurcation [33]. These methods, like most EWSs, can warn of critical transitions but cannot predict the specific location where

*jackmoore@tongji.edu.cn

†gyan@tongji.edu.cn

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

they will occur. One recent work [34] acknowledged this gap by introducing a Bayesian linear segment fit to estimate the transition time. However, this procedure can be applied only to low-dimensional systems, which limits its applicability for many complex systems of real-world interest [35,36]. Another machine-learning method proposes handling networked systems using reservoir computing [37] but does not evaluate performance on systems with more than four nodes, restricts attention to synchronization phenomena, and needs information other than nodal time series.

Established methods do not address the challenge of predicting tipping points quantitatively in diverse complex systems using only observed time series of node states. These data frequently entail additional issues: A substantial fraction of nodes may be hidden [38]; and strong observation noise may be present [39]. Therefore, we are motivated to ask the following question: how to effectively discover and integrate the information of the whole network and then find a versatile and robust way to leverage critical slowing down and other subtle features that emerge in time series prior to transitions.

In this paper, we answer this question via a machine-learning framework sandwiching graph isomorphism network (GIN) and gated recurrent units (GRU) neural network layers into a GIN-GRU architecture [40,41]. The GIN layers read in the time series of each node and identify and extract the features of individual nodes, subsequently integrating and outputting them into the collective information of the entire graph. The GRU layers then read in the output of the GIN and interpret those features, looping back on themselves to generate memory and identify recurring features in a long time series. The combination of these layers provides excellence in both graph pattern recognition and sequence prediction.

We demonstrate that the proposed GIN-GRU framework is effective in anticipating the exact location of critical transition of a wide range of complex systems. Moreover, we establish our method's robustness to the challenges

associated with most real-world complex system data by showing that it can provide early and accurate predictions even for systems measured amid significant noise, when using transient data, or when a substantial fraction of system components cannot be observed. Finally, we show the transferability of our method and its relevance to real data by considering empirical vegetated ecosystems in Africa.

II. METHOD

In this section, we present our method in detail, including the generation and processing of data and our deep-learning-based anticipation algorithm.

A. Generating and processing data

In order to train a neural machine to predict tipping points far in advance, we first need to generate relevant training data. Given specific system dynamics equations that would have transitions with certain parameters, after altering the value of control parameter ε and simulating, the important thing we have to attach attention to is how to identify a system's precise critical transition point ε_c [see Fig. 1(b)], which in the following would be regarded as the label in our supervised machine-learning training process and the basis of evaluating prediction accuracy. Many widely studied approaches can be used to detect transitions. For instance, different order parameters r are used to measure synchronization of coupled oscillators [42], and, in some continuous systems, the real part of the dominant eigenvalue λ (the eigenvalue whose real part has the smallest magnitude) of the system's Jacobian matrix increases to 0 [$\text{Re}(\lambda) \rightarrow 0$] [43] when the critical transition occurs.

Motivated by the significance of early prediction, we opt to use only a small segment of the time series, X_s , instead of the whole. Therefore, we employ a sliding window approach as illustrated in Fig. 1(b). The window, with a width of $w = 20$ time steps, slides from just before the transition at ε_c (where $s = 1$) to the earliest available data

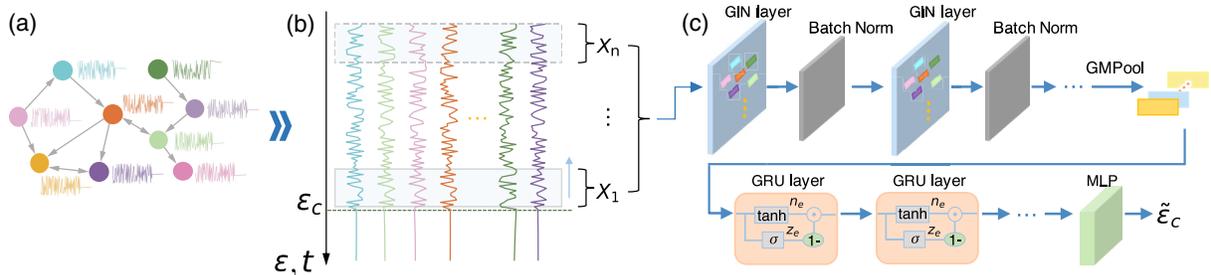


FIG. 1. Overview of our approach to anticipate transitions. (a) Complex network topology underlying observed time series. (b) Identification of the precise tipping point ε_c and preservation of the data X_1, \dots, X_n for prediction with a sliding window (numbers $1, \dots, n$ correspond to different lead distances). Control parameter ε changes over time t (not necessarily linearly). (c) Overall architecture of our deep-learning anticipation algorithm (i.e., GIN-GRU). The input, X_1, \dots, X_n , first passes six GIN layers, each followed by batch normalization [44]. Then, we let the graph-level embeddings obtained by the global max pooling (GMPool) operation go through a four-layer GRU. Finally, after a multilayer perceptron (MLP), we acquire the predicted $\tilde{\varepsilon}_c$.

(the initial w time steps, where $s = n$). In real systems, the control parameter ε may not change linearly over time. Thus, we define the lead distance as $|\varepsilon_s - \varepsilon_c|$. Our approach aims to forecast the exact location ε_c of the critical transition in advance, from some point ε_s by using only successive data X_s . Note that knowledge of the value of ε_s is not required in the training or prediction procedure of our method, and it is just used to assess the method performance.

B. Deep-learning method

After a large number G of simulations, a massive and diverse dataset has been built, with tens of thousands of data $X_1^1, \dots, X_{n(1)}^1, \dots, X_1^G, \dots, X_{n(G)}^G$ and their corresponding labels. We then split the dataset (after randomly shuffling) into train/validation/test data by 7:1:2, avoiding splitting individual simulations to avoid data leakage. In the training phase, as shown in Fig. 1(c), we first let each training datum pass a six-layer GIN. In every layer, the time series of each node \mathbf{x}_i is calculated following the formula

$$\mathbf{x}'_i = h_{\Theta} \left((1 + \varepsilon) \cdot \mathbf{x}_i + \sum_{j \in \mathcal{N}(i)} \mathbf{x}_j \right), \quad (1)$$

where ε denotes a learnable parameter or a fixed scalar, $\mathcal{N}(i)$ denotes the set of i 's neighboring nodes, and h_{Θ} denotes a multilayer perceptron (MLP) with trainable parameters Θ . Note that, since the topology information is unseen, $\sum_{j \in \mathcal{N}(i)} \mathbf{x}_j$ is actually not calculated. After several layers' iterations and computing, the representation of each node is obtained. We then employ a MaxAggregation operation $\mathbf{r} = \max_{i=1}^N \mathbf{x}_i$ to garner the entire network's representation. Thus far, we have successfully transformed a multidimensional complex network time series into single-dimensional graph-level outputs. Subsequently, we utilize four GRU layers to further explore the latent transition characteristics within the outputs. For each element r_e in the input sequence \mathbf{r} , each layer computes

$$h_e = (1 - z_e) \odot n_e + z_e \odot h_{(e-1)}, \quad (2)$$

where h_e is the hidden state at time e , z_e and n_e denote the update and new gates, respectively, and \odot is the Hadamard product defined by $(A_{i,j}) \odot (B_{i,j}) = (A_{i,j} B_{i,j})$. To avoid overfitting, between every two GRU layers there is a dropout layer with a dropout probability $D_p = 0.1$ to avoid overfitting. Finally, the outputs are passed to an MLP to generate a predicted $\tilde{\varepsilon}_c$. The objective function is the mean square error \mathcal{L} given by

$$\mathcal{L} = \frac{1}{B} \sum_{b=1}^B (\tilde{\varepsilon}_{c_b} - \varepsilon_{c_b})^2, \quad (3)$$

where B denotes the batch size of the training set. In the prediction phase, the dropout operation is removed, and the test result is recorded when the best result is achieved on the validation set.

III. RESULTS

In this section, we demonstrate the ability of our method to detect both first- and second-order phase transitions in systems drawn from ecology, neuroscience, and collective motion.

A. Predicting synchronization of nonidentical phase oscillators

First, we explore the ability of our framework to predict the onset of second-order transitions by considering synchronization of symmetrically coupled Kuramoto oscillators [45], the dynamics of which are governed by the equation

$$\frac{d\theta_i}{dt} = \omega_i + \varepsilon \sum_{j=1}^N A_{i,j} \sin(\theta_j - \theta_i), \quad (4)$$

where ω_i denotes natural frequency of each oscillator, ε here denotes the coupling strength of the interactions, and $A = (A_{i,j})$ is the adjacency matrix of the network, which is defined such that $A_{i,j} = 1$ ($A_{i,j} = 0$) when there is an edge (when there is no edge) between node i and node j . The dynamics of coupled Kuramoto oscillators are highly dependent on the strength ε of the interactions. When coupling is weak (i.e., small ε), oscillators rotate almost independently with their respective frequencies ω_j . However, as the coupling becomes sufficiently strong (i.e., $\varepsilon \rightarrow \varepsilon_c$), a fully phase-locked state emerges in a critical transition. In this state, all oscillators synchronize to a common frequency such that constant phase differences are established between each pair of nodes: $\theta_i - \theta_j = \text{const}$.

When generating data, to emulate the diversity of real-world complex networks, for each simulation we use a distinct combination of parameter values for a scale-free (SF) network [46] with scale exponent 2.5 and different initial values for each oscillator's phase θ_i . Specifically, the mean degree is chosen randomly between 3.5 and 6.5, the number of oscillators is randomly chosen between 65 to 135, and the initial phase of each oscillator is randomly chosen between 0 and 2π . Because reciprocity in interactions is common in many applications concerning oscillators, we consider undirected networks, i.e., a symmetric adjacency matrix $A_{i,j} = A_{j,i}$. After initializing, we increase ε gradually, from 0 to 5, in steps of 0.05. For each value of ε , the system is evolved for 100 units of time to allow it to equilibrate, at which point the final value of each oscillator's phase θ_i is recorded. We choose the natural frequency ω_i of each oscillator randomly between 0 and

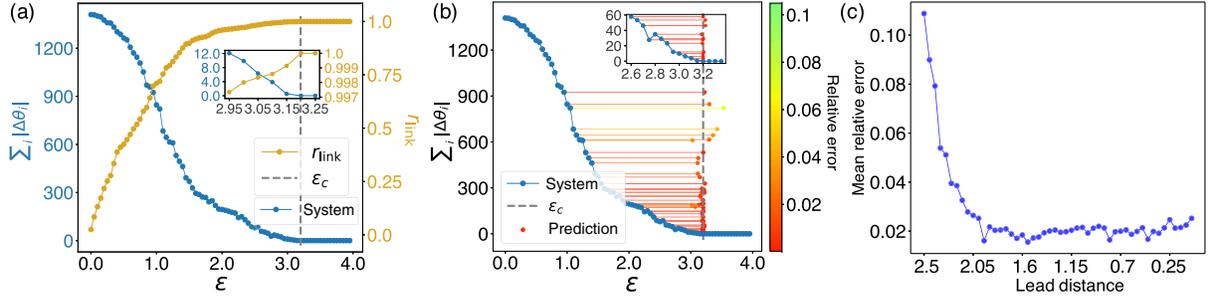


FIG. 2. Predicting the onset of synchronization transition of the Kuramoto model. (a),(b) Presentation of an example randomly chosen from the test data. (a) Indicators measuring the synchronization degree, underlying system quantity $\sum_i |\Delta\theta_i|$ and order parameter r_{link} , respectively, versus control parameter ϵ (i.e., coupling strength). The tipping point ϵ_c is denoted by a vertical gray dashed line, and the inset shows the magnification of $\epsilon \in [2.95, 3.25]$. (b) Independent predictions $\tilde{\epsilon}_c$ from different lead distances. The horizontal lines show the interval between the start of anticipation ϵ_s (i.e., the blue dots) and the predicted critical transition location $\tilde{\epsilon}_c$. The relative error, i.e., the anticipation inaccuracy I , in each prediction is represented by the color bar on the right, and the inset focuses on the $\epsilon \in [2.6, 3.35]$. (c) Mean anticipation inaccuracy \bar{I} versus lead distances derived from the entire test dataset.

2π before translating each natural frequency by the same amount to consider a corotating frame with $\sum_{i=1}^N \omega_i = 0$. As a result, when the fully phase-locked state occurs, all oscillators synchronize to the common frequency ($d\theta_i/dt = 0$). Therefore, the sum $\sum_i |\Delta\theta_i|$ over oscillators of the magnitude $|\Delta\theta_i|$ of the change in phase throughout the final $\Delta t = 10$ interval can be used as a parameter to evaluate the degree of system synchrony. This is illustrated in Fig. 2(a) for an example system: When the critical transition occurs, $\sum_i |\Delta\theta_i| = 0$ (blue curve), the point from which the order parameter r_{link} [47] attains unity, $r_{\text{link}} = 1$ (brown curve). The statistics r_{link} and $\sum_i |\Delta\theta_i|$ change almost monotonically as the transition point approaches but do not provide an estimate of the exact onset of complete synchronization, and, furthermore, computation of r_{link} requires knowledge of network topology. In contrast, as Fig. 2(b) illustrates, deep learning can accurately estimate the critical transition long before its occurrence, based on time series alone.

On detecting ϵ_c , we start recording data with the sliding window (see Sec. II A) containing time series segment X_s (each comprising 20 time steps) of different lead distances prior to the critical transition, where $s \in [1, n]$ and $n \leq 50$. Here, n can vary among simulations depending on the duration between the start of the simulation and the simulation's critical transition, and the maximum considered n is set to 50. We run $G = 1000$ simulations, each leading to $n \in [1, 50]$ data, providing a total of 32 632 data X_s . We assess prediction performance using the relative anticipation inaccuracy

$$I = \frac{|\tilde{\epsilon}_c - \epsilon_c|}{\epsilon_{\text{max}} - \epsilon_{\text{min}}},$$

where $\tilde{\epsilon}_c$ is our prediction of the critical value ϵ_c and ϵ_{min} and ϵ_{max} are the minimum and maximum value of the control parameter, here, 0 and 5, respectively. Despite the

wide range of critical coupling strengths ϵ_c in the unseen test data (i.e., $\epsilon_c \in [1.35, 4.85]$), which poses a significant challenge for forecasting, the mean over all test data of the anticipation inaccuracy I is only 0.0215. In Fig. 2(c), we investigate how mean relative anticipation inaccuracy \bar{I} of test data changes with lead distances. For lead distances about 2.0 or less, our neural model starts to exhibit high predictive accuracy, with $\bar{I} \leq 0.0252$. This illustrates how, far before the critical transition, our deep-learning algorithm can already effectively detect both critical slowing down and other subtle features emerging from the time series. Thus, we can accurately predict the tipping points at an early stage, not having to wait until they are imminent, which greatly facilitates planning necessary to accommodate or avert the transition.

B. Predicting sharp decline of resource biomass

The synchronization transition of the Kuramoto model we have discussed above belongs to the category of second-order phase transitions [48,49], in which a physical quantity characterizing the macroscopic state changes continuously through the transition point. Now, we explore the capacity of our machine-learning scheme in anticipating another kind of transition, the first-order phase transition [50,51], characterized by an abrupt onset in which the underlying quantity varies discontinuously at the transition point.

We consider an ecosystem model widely used to simulate harvested crops [52,53], animal population systems [52,53], insect pest systems [54], and human host-parasite systems [55]. The dynamics of the model are governed by the equation

$$\frac{dx_i}{dt} = rx_i \left(1 - \frac{x_i}{k} \right) - \epsilon \frac{x_i^2}{x_i^2 + 1} + d \sum_{j=1}^N A_{i,j} (x_j - x_i) + \eta W_i, \quad (5)$$

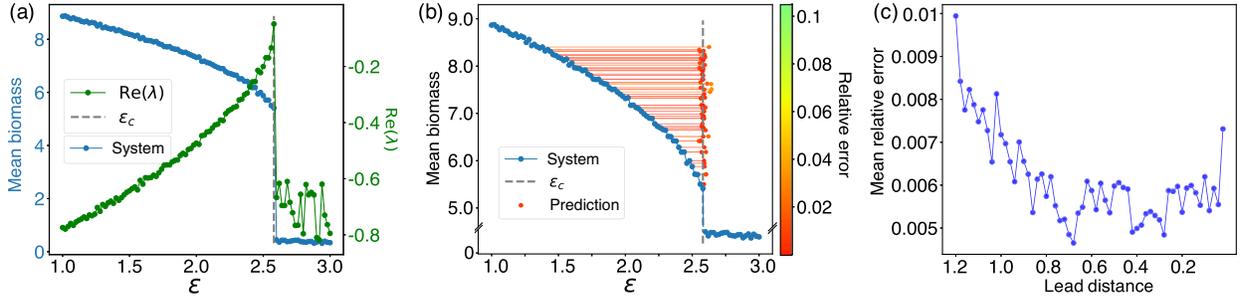


FIG. 3. Predicting occurrence moment of sharp decline of resource biomass system. (a),(b) An example system randomly chosen from the unseen test dataset. (a) Underlying system quantity \bar{x} representing the average resource biomass and another transition indicator $\text{Re}(\lambda)$ versus the grazing rate control parameter ε . The tipping point ε_c is denoted by a vertical gray dashed line. (b) Independent predictions $\tilde{\varepsilon}_c$ for different lead distances. The horizontal lines show the interval between the start of anticipation ε_s (i.e., the blue dots) and the predicted critical transition location $\tilde{\varepsilon}_c$. The relative error, i.e., the anticipation inaccuracy I , in each prediction is represented by the color bar on the right. (c) Mean anticipation inaccuracy \bar{I} versus lead distance derived from the entire test dataset.

where x_i denotes the state of one small area. The interpretation of the parameters of this model depends on which of the ecosystems listed above is considered. For concreteness, we refer to a grazing system with control parameter ε representing the grazing rate, for which state variables x_i represent biomass and the parameters k , r , d , and η denote the carrying capacity, the self-growth rate, the degree of influence by others, and the intensity of noise, respectively, while W_i represents measurement noise with a standard Gaussian distribution [56,57].

When generating data, for each simulation, we use a distinct combination of SF network parameters and different initial values for each node's biomass x_i . Specifically, the network mean degree is randomly chosen from 2 to 6, the number of nodes is randomly selected from 5 to 400, the growth rate r in the interval $[0.7, 1)$, and the influence rate d in $[0.5, 1)$, while the carrying capacity k and the intensity of noise η are fixed as $k = 10$ and $\eta = 0.5$. After initializing, we gradually increase the control parameter ε , from 1 to 3, in steps of 0.02. Each time, the final value of each node would be recorded after the system reaches equilibrium. In contrast to the second-order case, in this first-order phase transition, the mean resource biomass $\bar{x} = \sum_i x_i$ changes abruptly at the tipping point ε_c as shown in Fig. 3(a) (blue curve). This corresponds to a critical transition from underexploitation to overexploitation when grazing pressure surpasses a threshold. At the same time, the real part $\text{Re}(\lambda)$ of the dominant eigenvalue (see Sec. II A) approaches zero from below before suddenly plummeting (green curve). Figure 3(b) illustrates that our method, using the time series alone, accurately anticipates $\tilde{\varepsilon}_c$ of the example system long before the moment of transition.

Upon detecting ε_c , we begin recording data of time series segment X_s for different lead distances, where $s \in [1, n]$ and $n \leq 60$. We consider $G = 600$ different simulations to generate 26 191 data in total. Despite the wide range of ε_c in the massive unseen test data (i.e., $\varepsilon_c \in [1.9, 2.66]$), the mean over all test data of the anticipation inaccuracy I is only 6.77×10^{-3} . In addition, the average over test data

depending on lead distances of the anticipation inaccuracy, shown in Fig. 3(c), is very low even for the largest lead distance considered, with $\bar{I} \leq 9.94 \times 10^{-3}$. First of all, this elucidates that even the earliest sequence of 20 consecutive observations is sufficient for our algorithm to predict accurately. More importantly, this further demonstrates the effectiveness of our GIN-GRU algorithm, which can capture the overall structural information of complex networks and study and excavate from the time series the bifurcations in a system's future, even in the case of abrupt first-order transitions.

C. Predicting abrupt collapse of neuronal system

To further demonstrate the versatility of the GIN-GRU approach, we consider again the challenging case of first-order phase transitions, this time while examining a distinct class of networks and a dynamical model drawn from neuroscience. Specifically, we utilize Erdős-Rényi (ER) networks and the Wilson-Cowan model [36,58], which describes the firing-rate activity of a population of neurons according to the differential equations

$$\frac{dx_i}{dt} = -x_i + \sum_{j=1}^N w_{i,j}(1 - \varepsilon) \frac{1}{1 + e^{-\tau(x_j - \mu)}}, \quad (6)$$

where τ and μ control the steepness of the activation function and the firing-rate threshold, respectively, the element $w_{i,j} \in \mathbb{R}$ encodes for the strength of the directed interaction from node j to node i , and the control parameter ε here denotes the fraction by which link strengths have decreased.

In the brain networks of a variety of organisms, topological perturbations, including decreasing link strengths, or node or link removal, are ubiquitous [35,58]. As shown in Fig. 4(a), when the decreasing fraction ε is small, the system exhibits a minor response, and the mean activity of all neurons \bar{x} is still positive. However, when ε is large

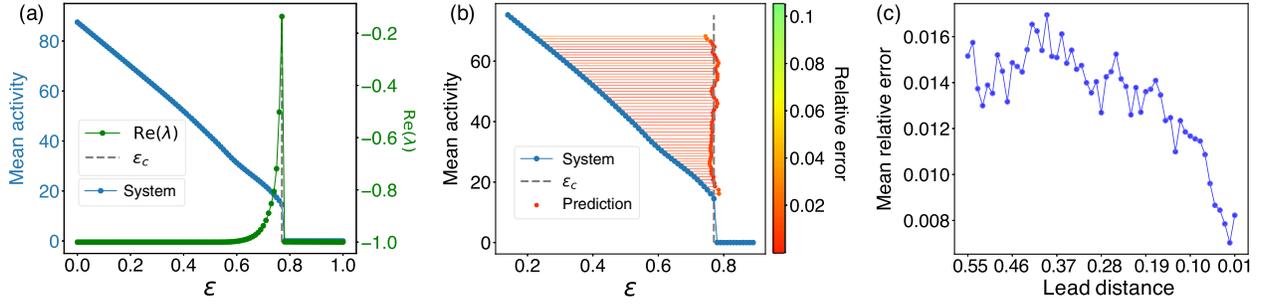


FIG. 4. Predicting occurrence moment of abrupt collapse of Wilson-Cowan neuronal system. (a),(b) Presentation of an example randomly chosen from the test dataset. (a) Underlying system quantity \bar{x} representing the average activity of all neurons and another transition indicator $\text{Re}(\lambda)$ versus link degradation control parameter ε . The tipping point ε_c is denoted by a vertical gray dashed line. (b) Independent predictions $\tilde{\varepsilon}_c$ for different lead distances. The horizontal lines show the interval between the start of anticipation ε_s (i.e., the blue dots) and the predicted critical transition location $\tilde{\varepsilon}_c$. The relative error, i.e., the anticipation inaccuracy I , in each prediction is represented by the color bar on the right. (c) Mean anticipation inaccuracy \bar{I} versus lead distance derived from the entire test dataset.

enough, a major collapse abruptly occurs, the system transitions from a functional to a dysfunctional dynamic state, and \bar{x} is suddenly suppressed in a first-order phase transition. When generating data, for each simulation, we use a distinct combination of ER network parameter values and different initial values for each node's activity x_i . Specifically, the mean degree of the ER network is randomly set between 3 and 6, the number of nodes is randomly selected from between 300 and 700, and the weight of each edge is randomly set between 10 and 20. After initializing, we then increase the weight attenuation factor ε gradually from 0 to 1 in steps of 0.01. Each time, the final value of each node would be recorded after the system reaches equilibrium. Similarly to the situation of resource biomass system, in the Wilson-Cowan model, \bar{x} has a sudden shift around ε_c with a small change of ε . Simultaneously, $\text{Re}(\lambda)$ approaches close to 0 [see Fig. 4(a) and Sec. II A]. Figure 4(b) shows, for an example system, the prediction $\tilde{\varepsilon}_c$ of the critical value of the attenuation factor which our method provides based only on time series.

On detecting ε_c , we start recording X_s for different lead distances, where $s \in [1, n]$ and $n \leq 55$. $G = 1000$ different simulations have been done and generated 28 233 data in total. Despite the wide range of ε_c in the unseen test data, i.e., $\varepsilon_c \in [0.38, 0.84]$, the mean over all test data of anticipation inaccuracy is only 0.0132. Moreover, as shown in Fig. 4(c), not only does our neural machine exhibit high predictive accuracy in the vicinity of the critical transition, where the lead distance is small, but the anticipation inaccuracy tends to plateau to a reasonable level, less than 0.0169, as the lead distance gradually increases.

IV. ROBUSTNESS ANALYSIS

Accurate network structure and precise knowledge of relevant external conditions are difficult to ensure [38,39], and, hence, we have restricted ourselves to using only time series without providing to our machine-learning prediction algorithm any topological information or values

of the control parameter. In practice, measurement noise is ubiquitous, it can be difficult to observe all relevant variables, and the system is not allowed to reach equilibrium. Therefore, in this section, we validate the robustness of our GIN-GRU approach against incomplete data, observational noise, and transient data.

A. Incomplete data

In practice, we commonly encounter the challenging scenario where only a partial observation of the full system is available [38]. For example, measurement limitations or unrecognized salient variables observations may restrict the detection range to only a subset of nodes [59], while the other nodes are regarded as “hidden” because direct observation or information about them is unattainable. Concretely, in our case, the time series of only a randomly chosen subset of all nodes is observed. The problem is then expressed as anticipating tipping point $\tilde{\varepsilon}_c$ given only the time series of the nonhidden nodes.

To uncover the impact of missing data, we consider Wilson-Cowan neuronal systems with the same distribution of topological and dynamical properties as previously. The result of anticipation inaccuracy versus different ratios of hidden nodes is shown in Fig. 5(a). Overall results (thick purple line and filled circles) and results for specific lead distances are plotted separately. Even when 81% of the nodes are unobserved, our algorithm exhibits prediction accuracy comparable to the scenario where complete data are available. Performance tends to improve as s becomes smaller, but trends are similar across different s . Even when anticipating from an early stage (i.e., for high s) and with a high fraction (i.e., 99%) of nodes unobserved, the mean anticipation inaccuracy \bar{I} remains low, at less than 0.039.

B. Observational noise

Another common challenge in realistic networked systems is noise [39,60], either dynamical or observational.

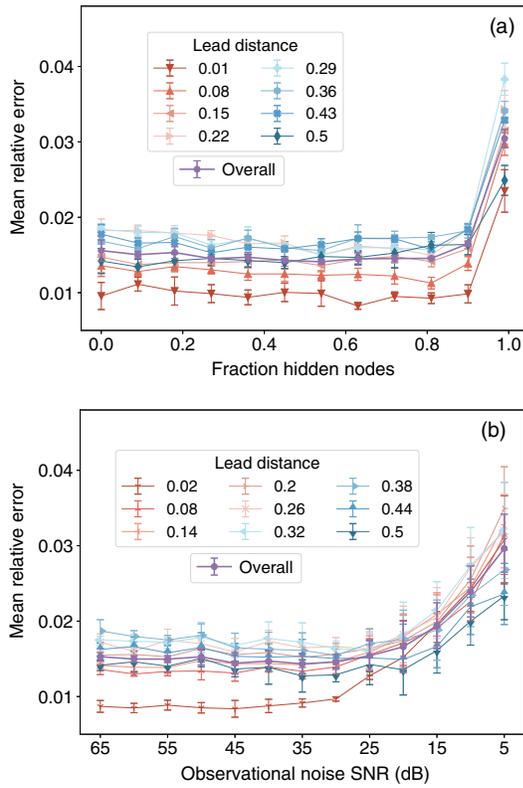


FIG. 5. Anticipation robustness against incompleteness and noise. (a) Mean anticipation inaccuracy versus fraction of nodes which are hidden (unobserved). (b) Mean anticipation inaccuracy versus observational noise intensity quantified by signal-to-noise ratio (SNR). Different lead distances and overall inaccuracy are plotted separately. The error bar represents standard deviation over five independent trials.

Dynamical noise refers to the intrinsic stochasticity in dynamics, which we have already considered in the resource biomass system (see Sec. III B), while observational noise is induced by the measuring process. We consider observational noise $a\beta_X$ added to original time series X such that the final observed time series used in our prediction model is $X^{\text{obs}} = X + a\beta_X$, where β_X are drawn independently for each time step from a standard normal distribution and the intensity a determines the signal-to-noise ratio (SNR). Our problem becomes predicting tipping point $\tilde{\epsilon}_c$ based on noisy time series data X^{obs} .

Now we test the impact of observational noise on the performance of our deep-learning approach. Wilson-Cowan neuronal system abrupt collapse anticipation inaccuracy (MRE) versus different SNR (dB) of observational noise is displayed in Fig. 5(b), showing overall performance (thick purple line and filled circles) and specific lead distances separately. Our algorithm exhibits remarkable resilience, tolerating 35 dB SNR without obvious deterioration. Similarly to the case of incomplete data, overall trends and those for different lead distances are similar. Even in the most challenging circumstance—predicting far in advance (i.e., large lead distance) and the highest noise intensity (5 dB)—the mean anticipation inaccuracy remains low, at less than 0.035.

C. Transient data

Real dynamical systems often do not maintain an equilibrium state. Therefore, in this subsection, we demonstrate that our approach can also be applied to predict critical transitions from transient data. Specifically, we reduce the time interval between a change in the control

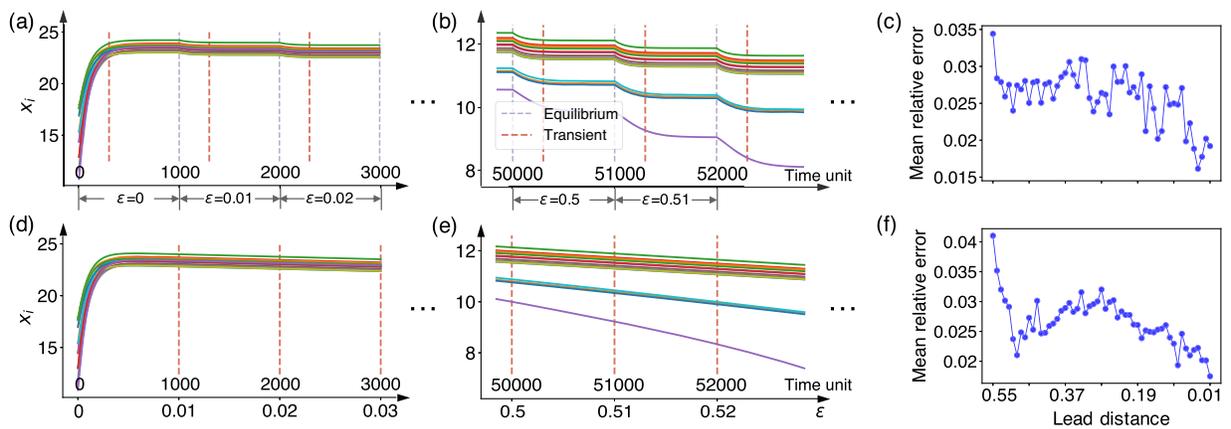


FIG. 6. Anticipation robustness against transient data. (a),(b),(d),(e) Example of evolution of a subset of nodes in a Wilson-Cowan system. The gray dashed line indicates the equilibrium data we utilize in Sec. III, while the vertical orange dashed line shows the transient data we consider in Sec. IV C. (a),(b) Shortening the interval between changing the control parameter and recording system state. (d),(e) Continuously changing the control parameter as the system evolves. (c),(f) Mean anticipation inaccuracy \bar{I} versus lead distances derived from the entire test dataset for (c) a shorter interval before recording system state and (f) a continuously changing control parameter.

parameter and the recording of the system state. While in Sec. III C each state of the Wilson-Cowan neuronal system is recorded after 1000 time units, by which time it had equilibrated, here the state is recorded after only 300 time steps, at which point the system clearly has not reached equilibrium [Figs. 6(a) and 6(b)]. In Fig. 6(c), we depict the mean anticipation inaccuracy versus lead distance when relying on such transient data, showing that the average anticipation inaccuracy over all test data is 0.026, only slightly higher than for the equilibrium data previously considered.

In Figs. 6(d) and 6(e), we present the challenging circumstance in which the control parameter continuously increases as the system evolves. In this case, when recording the data point, we expect the system to be significantly out of equilibrium. However, our method still performs robustly, with no substantial decrease in prediction performance [see Fig. 6(f)].

V. TRANSFERABILITY AND APPLICATION

In scenarios where critical transitions are pertinent and limited training data are available from the system under study, directly training a neural machine may not be feasible. To address this challenge, we propose a strategy: leveraging the neural machine already well trained on synthetic data and fine-tuning it on a small dataset from the system of ultimate interest. This approach enables the neural machine to adapt to the new system and capture its latent mechanism effectively. In this section, we demonstrate the transferability of our algorithm to empirical vegetated ecosystems in Africa.

In regions with semiarid climates, vegetated ecosystems can undergo sudden and unforeseen changes, including transitioning from forest to savannah due to fluctuating or time-varying factors like rainfall [61]. These transitions can have severe consequences for both the ecosystem and its dependent communities. To study this type of system, we utilize satellite imagery from the Moderate Resolution Imaging Spectroradiometer (MODIS) at a resolution of 250 meters per pixel (MOD44B) [62] to obtain continuous measurements of tree coverage.

We integrate vegetation data with mean annual precipitation data obtained from the Tropical Rainfall Measuring Mission (TRMM) 3B43 dataset, which has a resolution of 0.25° [63]. To assign each vegetation pixel an annual precipitation value, we linearly interpolate the precipitation data at the vegetation grid level, averaging over a span of 20 yr. Within the dataset, regions showcasing bistability between tropical forests and savannahs within an intermediate range of mean annual precipitation have long been recognized [64,65]. We identify three such districts in Central Africa that exhibit bistability and an abrupt transition between the two states (see Fig. 7).

Each data point in Fig. 7 represents a $500\text{-pixel} \times 500\text{-pixel}$ area and is recorded sequentially along the variation of latitude with a step of 50 pixels. We build a network comprising 100 nodes, each spanning five pixels in the longitudinal direction and without overlap between distinct nodes. To apply our established framework, we linearly interpolate the original data to obtain more data points to meet the time series segment X_s length requirement (each comprising 20 time steps) mentioned in Sec. II A. In addition, owing to the different widths of three districts, i.e., from 500

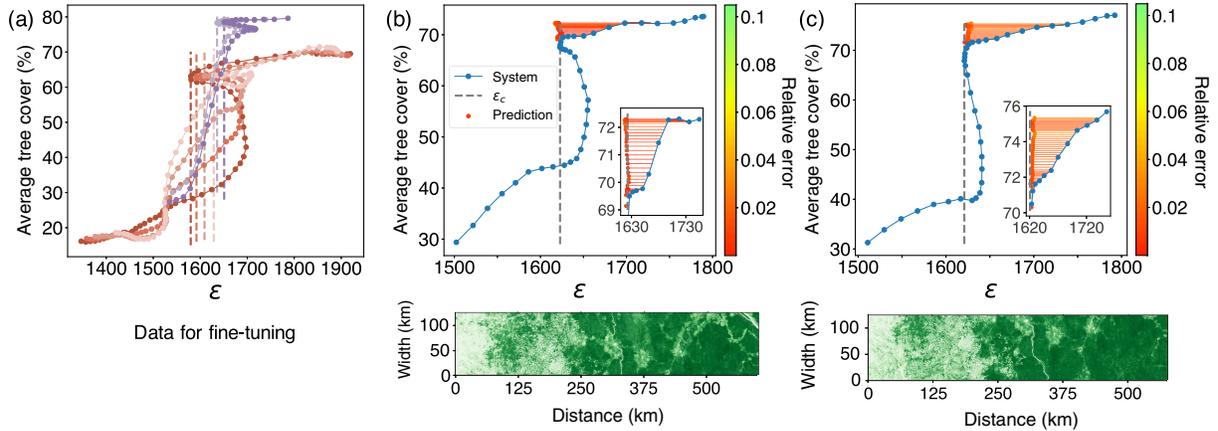


FIG. 7. Predicting occurrence moment of abrupt transition in empirical vegetated ecosystems using transfer learning. (a) Presentation of real data used for fine-tuning, which shows the underlying system quantity \bar{x} representing the average tree cover versus the annual rainfall control parameter ε . Tipping points ε_c are shown by vertical dashed lines. (b),(c) For different lead distances, mean anticipation result $\tilde{\varepsilon}_c$ of two areas from a new district. To reduce statistical fluctuations due to data scarcity, for each lead distance we show the mean over ten independent fine-tuning trials. The tipping point ε_c is denoted by a vertical gray dashed line, and the inset shows a magnification of all predictions. The horizontal lines show the interval between the start of anticipation ε_s (i.e., the blue dots) and the predicted critical transition location $\tilde{\varepsilon}_c$. The relative error, i.e., the anticipation inaccuracy I , in each prediction is represented by the color bar on the right. Each plotted blue dot corresponds to a window of the empirical tree coverage shown at the bottom in (b) and (c).

to 800 pixels, we can obtain different numbers of transition data with the same sliding width of 100 pixels. To be specific, we obtain two and four subsets of data from two districts [mean locations are (26.119439°E, 2.498051°N) and (20.892508°E, 3.957983°N), respectively] and use them to fine-tune our neural machine. Finally, we anticipate the critical transition of rainfall for a brand new district [mean location (23.957871°E, 3.957249°N)].

Given the limited availability of real data for vegetated ecosystems, we employ a transfer learning strategy [66] to predict real critical transitions. Initially, we train a neural machine on synthetic data to obtain a pretrained deep-learning (DL) model. Subsequently, this pretrained model is fine-tuned using the limited data available from real vegetated systems. This fine-tuning process allows the DL model to adapt to the new and underlying mechanisms of the real system.

To enhance the generalizability of our DL model, we generate synthetic training data from previously discussed first-order phase transition systems, including the resource biomass and neuronal systems, along with a new mathematical model of ecosystem focusing on the interactions between macrophyte coverage and turbidity in shallow lakes [57,67]. In this mathematical ecosystem, we vary the number of nodes randomly between 10 and 200, the scale exponent for the scale-free network between 2.1 and 2.9, and the network mean degree between 2 and 6. Following pretraining on synthetic data, we fine-tune the DL model using limited real data from two districts of the empirical vegetated ecosystem [as depicted in Fig. 7(a)]. This fine-tuning process involves further training the DL model for a relatively small number of epochs (300 epochs compared to 1000 epochs in the pretraining stage).

This final DL model is used to predict the critical mean annual rainfall for a completely new district. It is worth noting that the fine-tuning data are sourced from two districts that are different from the target district, ensuring there is no data leakage when making predictions. As shown in Figs. 7(b) and 7(c), despite the distinct patterns and critical rainfall (ϵ_c) values in the new, unseen district compared to those used for fine-tuning, our deep-learning approach maintains a low anticipation inaccuracy. The mean relative errors are merely 0.008 and 0.016, respectively. Furthermore, our algorithm demonstrates the capability to quantitatively predict critical transitions far from the transition onset, underscoring its transferability and relevance to real-world systems.

VI. DISCUSSION

Anticipating critical transitions is a crucial task in dealing with nonlinear systems. In this work, we propose a deep-learning-based framework for quantitative prediction of tipping points solely using time series data. Through three synthetic examples and one empirical system, we demonstrate the effectiveness and robustness of our methods. Even

when predicting critical transitions well in advance, our approach maintains precision. The ubiquity of critical threshold phenomenon in real-world systems and potential catastrophe risks means that the method has broad application prospects. For instance, provided the difference between the system's current control parameter value and the value at the transition is known, natural disasters would have more possibilities to be managed effectively.

There are some methods in the literature aiming for early warning. A novel deep-learning approach was proposed in Ref. [33], but the work is dedicated to early bifurcation classifications rather than estimates of transition onset. Traditional EWSs such as rising variance [31] and lag-1 autocorrelation [5], as well as indicators such as r_{link} (see Sec. III A), also show changes well before a critical transition but do not provide a quantitative prediction. In addition, many indicators utilize information, such as the full topology of the network, which is challenging to access in more realistic situations. In contrast, our approach does not remain at the qualitative level but, instead, leverages only time series data to sense and capture latent transition laws and transform them into quantitative anticipation. Observed trends in accuracy are consistent with the different properties of first- and second-order phase transitions. Specifically, accuracy depends on lead distance more strongly for the biomass and Wilson-Cowan systems, which are first-order transitions characterized by abrupt changes, than for the second-order Kuramoto system.

Many works that focus on EWSs only consider a small number of variables (generally no more than 3), which renders uncertain their applicability to more realistic scenarios. For example, a recent approach based on bifurcation theory also works well in anticipating low-dimensional cases [68], but, since it is derived from a one-dimensional time series, may be challenged by high-dimensional systems. Another method has been proposed to anticipate synchronization transitions of networked systems [37], by employing a “parameter-aware” reservoir computing neural machine which can predict the locations of the transition points for both smooth and explosive transitions. Unfortunately, because the computational load of reservoir computing is quite large, the method's applicability is restricted to sufficiently small networks. In addition, the method relies on knowledge of control parameter values, but this information may not be available in practical settings. In our GIN-GRU combined deep-learning scheme, these restrictions are all lifted.

We also conducted experiments with graph convolutional networks [69], graph attention networks [70], and other graph neural networks, as well as many recurrent neural networks, but found that the GIN-GRU architecture achieves the highest anticipation accuracy. This can be attributed to GIN's incorporation of graph isomorphism within its network structure, enabling GIN to more effectively leverage global information from local features through a specialized

feature mapping function and a learnable layerwise aggregation function. We found that the long short-term memory network [71] yielded similar effects to GRU, but GRU is much more efficient.

The computational complexity of our proposed algorithm mainly depends on two parts. The GIN has complexity about $O(LNN_n)$, where N , L , and N_n denote the number of nodes, number of layers of GIN, and number of neurons of the multilayer perceptron, respectively. The GRU incurs complexity approximately $O(L_s d_i d_h)$, where L_s , d_i , and d_h represent the length of sequence, the dimension of the input, and number of hidden neurons, respectively. Our machine-learning implementation is based on PYTORCH, and experiments are conducted on a local machine equipped with two NVIDIA V100 32 GB GPUs. All our codes are available in Ref. [72].

Our work raises several questions worthy of future pursuit. First, the size of sliding window (i.e., w) is presently set to 20. Although this is already relatively small, it would be interesting to investigate how much smaller windows can be while still providing accurate predictions. Second, some networked dynamical systems may undergo critical transitions more than once [73,74]. Predicting the locations of all transitions solely based on pretransition data is a challenging but valuable goal. Third, the nodes in a complex system can have higher-order interactions that may impact the transition of networked systems [75,76]. Hence, it is an interesting direction to extend the approach to anticipating higher-order network tipping points. Finally, regarding our framework for predicting the critical transitions of empirical systems based on a deep-learning model well trained on synthetic data, it would be worthwhile to integrate more synthetic systems to increase the versatility of the pretrained model.

ACKNOWLEDGMENTS

We thank Xinjie Zhang for helpful discussion and providing code, Renjie Ni for downloading some real data, and the anonymous reviewers for suggestions and corrections which significantly improved the paper. This work is supported by the National Natural Science Foundation of China (Grants No. T2225022, No. 12161141016, No. 12375036, and No. 62088101), Shuguang Program of Shanghai Education Development Foundation and Shanghai Municipal Education Commission (Grant No. 22SG21), Shanghai Municipal Science and Technology Major Project (Grant No. 2021SHZDZX0100), Science and Technology Commission of Shanghai Municipality (Grant No. 19511132101), and the Fundamental Research Funds for the Central Universities.

[1] M. Scheffer, S. Carpenter, J. A. Foley, C. Folke, and B. Walker, *Catastrophic shifts in ecosystems*, *Nature (London)* **413**, 591 (2001).

[2] M. Scheffer, *Critical Transitions in Nature and Society* (Princeton University Press, Princeton, NJ, 2020), Vol. 16.

[3] P. U. Clark, N. G. Pisias, T. F. Stocker, and A. J. Weaver, *The role of the thermohaline circulation in abrupt climate change*, *Nature (London)* **415**, 863 (2002).

[4] R. Corrado, A. M. Cherubini, and C. Pennetta, *Early warning signals of desertification transitions in semiarid ecosystems*, *Phys. Rev. E* **90**, 062705 (2014).

[5] V. Dakos, M. Scheffer, E. H. Van Nes, V. Brovkin, V. Petoukhov, and H. Held, *Slowing down as an early warning signal for abrupt climate change*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 14308 (2008).

[6] A. J. Veraart, E. J. Faassen, V. Dakos, E. H. Van Nes, M. Lürling, and M. Scheffer, *Recovery rates reflect distance to a tipping point in a living system*, *Nature (London)* **481**, 357 (2012).

[7] V. Dakos and J. Bascompte, *Critical slowing down as early warning for the onset of collapse in mutualistic communities*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 17546 (2014).

[8] I. A. van de Leemput, M. Wichers, A. O. Cramer, D. Borsboom, F. Tuerlinckx, P. Kuppens, E. H. Van Nes, W. Viechtbauer, E. J. Giltay, S. H. Aggen *et al.*, *Critical slowing down as early warning for the onset and termination of depression*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 87 (2014).

[9] I. Izrailtyan, J. Y. Kresh, R. J. Morris, S. C. Brozena, S. P. Kutalek, and A. S. Wechsler, *Early detection of acute allograft rejection by linear and nonlinear analysis of heart rate variability*, *J. Thorac. Cardiovasc. Surg.* **120**, 737 (2000).

[10] E. Cotilla-Sanchez, P. D. Hines, and C. M. Danforth, *Predicting critical transitions from time series synchrophasor data*, *IEEE Trans. Smart Grid* **3**, 1832 (2012).

[11] H. Haken, *Synergetics: An Introduction: Nonequilibrium Phase Transitions and Self-Organization in Physics, Chemistry, and Biology* (Springer-Verlag, New York, 1978).

[12] H. Haken, J. S. Kelso, and H. Bunz, *A theoretical model of phase transitions in human hand movements*, *Biol. Cybern.* **51**, 347 (1985).

[13] J. Xie, F. Meng, J. Sun, X. Ma, G. Yan, and Y. Hu, *Detecting and modelling real percolation and phase transitions of information on social media*, *Nat. Hum. Behav.* **5**, 1161 (2021).

[14] S. Strogatz, *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*, 2nd ed. (CRC Press, Boca Raton, 2015).

[15] J. Lim and B. I. Epureanu, *Forecasting a class of bifurcations: Theory and experiment*, *Phys. Rev. E* **83**, 016203 (2011).

[16] E. Gopalakrishnan, Y. Sharma, T. John, P. S. Dutta, and R. Sujith, *Early warning signals for critical transitions in a thermoacoustic system*, *Sci. Rep.* **6**, 35310 (2016).

[17] P. D. Ritchie, J. J. Clarke, P. M. Cox, and C. Huntingford, *Overshooting tipping point thresholds in a changing climate*, *Nature (London)* **592**, 517 (2021).

[18] T. M. Lenton, J. Rockström, O. Gaffney, S. Rahmstorf, K. Richardson, W. Steffen, and H. J. Schellnhuber, *Climate tipping points—Too risky to bet against*, *Nature (London)* **575**, 592 (2019).

[19] D. Eroglu, M. Tanzi, S. van Strien, and T. Pereira, *Revealing dynamics, communities, and criticality from data*, *Phys. Rev. X* **10**, 021047 (2020).

- [20] G. I. Hagstrom and S. A. Levin, *Marine ecosystems as complex adaptive systems: Emergent patterns, critical transitions, and public goods*, *Ecosystems* **20**, 458 (2017).
- [21] K. C. Ewel, C. Cressa, R. T. Kneib, P. S. Lake, L. A. Levin, M. A. Palmer, P. Snelgrove, and D. H. Wall, *Managing critical transition zones*, *Ecosystems* **4**, 452 (2001).
- [22] V. Dakos, S. R. Carpenter, W. A. Brock, A. M. Ellison, V. Guttal, A. R. Ives, S. Kéfi, V. Livina, D. A. Seekell, E. H. van Nes *et al.*, *Methods for detecting early warnings of critical transitions in time series illustrated using simulated ecological data*, *PLoS One* **7**, e41010 (2012).
- [23] X. Peng, M. Small, Y. Zhao, and J. M. Moore, *Detecting and predicting tipping points*, *Int. J. Bifurcation Chaos Appl. Sci. Eng.* **29**, 1930022 (2019).
- [24] L. Xu, D. Patterson, S. A. Levin, and J. Wang, *Non-equilibrium early-warning signals for critical transitions in ecological systems*, *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2218663120 (2023).
- [25] F. Zhang and J. Wang, *Nonequilibrium indicator for the onset of epileptic seizure*, *Phys. Rev. E* **108**, 044111 (2023).
- [26] J. Liang, Y. Hu, G. Chen, and T. Zhou, *A universal indicator of critical state transitions in noisy complex networked systems*, *Sci. Rep.* **7**, 42857 (2017).
- [27] S. Carpenter and W. Brock, *Early warnings of unknown nonlinear shifts: A nonparametric approach*, *Ecology* **92**, 2196 (2011).
- [28] M. Scheffer, J. Bascompte, W. A. Brock, V. Brovkin, S. R. Carpenter, V. Dakos, H. Held, E. H. Van Nes, M. Rietkerk, and G. Sugihara, *Early-warning signals for critical transitions*, *Nature (London)* **461**, 53 (2009).
- [29] E. H. Van Nes and M. Scheffer, *Slow recovery from perturbations as a generic indicator of a nearby catastrophic shift*, *Am. Nat.* **169**, 738 (2007).
- [30] C. Wissel, *A universal law of the characteristic return time near thresholds*, *Oecologia* **65**, 101 (1984).
- [31] V. Dakos, E. H. Van Nes, P. d'Odorico, and M. Scheffer, *Robustness of variance and autocorrelation as indicators of critical slowing down*, *Ecology* **93**, 264 (2012).
- [32] X. Zhang, C. Kuehn, and S. Hallerberg, *Predictability of critical transitions*, *Phys. Rev. E* **92**, 052905 (2015).
- [33] T. M. Bury, R. Sujith, I. Pavithran, M. Scheffer, T. M. Lenton, M. Anand, and C. T. Bauch, *Deep learning for early warning signals of tipping points*, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2106140118 (2021).
- [34] M. Heßler and O. Kamps, *Bayesian on-line anticipation of critical transitions*, *New J. Phys.* **24**, 063021 (2022).
- [35] J. Gao, B. Barzel, and A.-L. Barabási, *Universal resilience patterns in complex networks*, *Nature (London)* **530**, 307 (2016).
- [36] E. Laurence, N. Doyon, L. J. Dubé, and P. Desrosiers, *Spectral dimension reduction of complex dynamical networks*, *Phys. Rev. X* **9**, 011042 (2019).
- [37] H. Fan, L.-W. Kong, Y.-C. Lai, and X. Wang, *Anticipating synchronization with machine learning*, *Phys. Rev. Res.* **3**, 023237 (2021).
- [38] W. Deng, C. Yang, K. Huang, and W. Wu, *A two-stage reconstruction method for complex networked system with hidden nodes*, *Chaos* **32**, 053105 (2022).
- [39] T.-T. Gao and G. Yan, *Autonomous inference of complex network dynamics from incomplete and noisy data*, *Nat. Computat. Sci.* **2**, 160 (2022).
- [40] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, *How powerful are graph neural networks?*, [arXiv:1810.00826](https://arxiv.org/abs/1810.00826).
- [41] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, *Learning phrase representations using rnn encoder-decoder for statistical machine translation*, [arXiv:1406.1078](https://arxiv.org/abs/1406.1078).
- [42] M. Schröder, M. Timme, and D. Witthaut, *A universal order parameter for synchrony in networks of limit cycle oscillators*, *Chaos* **27**, 073119 (2017).
- [43] T. M. Bury, C. T. Bauch, and M. Anand, *Detecting and distinguishing tipping points using spectral early warning signals*, *J. R. Soc. Interface* **17**, 20200482 (2020).
- [44] S. Ioffe and C. Szegedy, *Batch normalization: Accelerating deep network training by reducing internal covariate shift*, in *Proceedings of the 32nd International Conference on Machine Learning, Lille, France (PMLR, 2015)*, pp. 448–456.
- [45] Y. Kuramoto, *Self-entrainment of a population of coupled non-linear oscillators*, in *Proceedings of the International Symposium on Mathematical Problems in Theoretical Physics: 1975, Kyoto University, Kyoto/Japan* (Springer, New York, 1975), pp. 420–422.
- [46] M. Pósfai, Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabási, *Effect of correlations on network controllability*, *Sci. Rep.* **3**, 1067 (2013).
- [47] J. Gómez-Gardenes, Y. Moreno, and A. Arenas, *Paths to synchronization on complex networks*, *Phys. Rev. Lett.* **98**, 034101 (2007).
- [48] Y. Kuramoto, *Chemical Oscillations, Waves, and Turbulence* (Springer, New York, 1984).
- [49] S. J. Blundell and K. M. Blundell, *Concepts in Thermal Physics* (Oxford University Press, New York, 2010).
- [50] K. Binder, *Theory of first-order phase transitions*, *Rep. Prog. Phys.* **50**, 783 (1987).
- [51] R. Lipowsky, *Surface critical phenomena at first-order phase transitions*, *Ferroelectrics* **73**, 69 (1987).
- [52] I. Noy-Meir, *Stability of grazing systems: An application of predator-prey graphs*, *J. Ecol.* **63**, 459 (1975).
- [53] F. Brauer and D. A. Sanchez, *Constant rate population harvesting: Equilibrium and stability*, *Theor. Popul. Biol.* **8**, 12 (1975).
- [54] M. Hassell, J. Lawton, and J. Beddington, *Sigmoid functional responses by invertebrate predators and parasitoids*, *J. Anim. Ecol.* **46**, 249 (1977).
- [55] G. Macdonald, *The Epidemiology and Control of Malaria* (Oxford University Press, New York, 1957).
- [56] R. M. May, *Thresholds and breakpoints in ecosystems with a multiplicity of stable states*, *Nature (London)* **269**, 471 (1977).
- [57] V. Dakos, E. H. van Nes, R. Donangelo, H. Fort, and M. Scheffer, *Spatial correlation as leading indicator of catastrophic shifts*, *Theor. Ecol.* **3**, 163 (2010).
- [58] H. Sanhedrai, J. Gao, A. Bashan, M. Schwartz, S. Havlin, and B. Barzel, *Reviving a failed network through microscopic interventions*, *Nat. Phys.* **18**, 338 (2022).
- [59] H. Huang, *Effects of hidden nodes on network structure inference*, *J. Phys. A* **48**, 355002 (2015).

- [60] M. E. Newman, *Network structure from rich but noisy data*, *Nat. Phys.* **14**, 542 (2018).
- [61] G. Tirabassi and C. Masoller, *Entropy-based early detection of critical transitions in spatial vegetation fields*, *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2215667120 (2023).
- [62] C. DiMiceli, R. Sohlberg, and J. Townshend, *MODIS/Terra vegetation continuous fields yearly L3 global 250m SIN grid V061 [dataset]*, NASA EOSDIS land processes DAAC (2022).
- [63] Tropical Rainfall Measuring Mission (TRMM), TRMM (TMPA/3B43) rainfall estimate L3 1 month 0.25 degree x 0.25 degree V7, Greenbelt, MD, Goddard Earth Sciences Data and Information Services Center (GES DISC) (2011).
- [64] C. Ciemer, N. Boers, M. Hirota, J. Kurths, F. Müller-Hansen, R. S. Oliveira, and R. Winkelmann, *Higher resilience to climatic disturbances in tropical vegetation exposed to more variable rainfall*, *Nat. Geosci.* **12**, 174 (2019).
- [65] B. Wuyts, A. R. Champneys, and J. I. House, *Amazonian forest-savanna bistability and human impact*, *Nat. Commun.* **8**, 15519 (2017).
- [66] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, *A comprehensive survey on transfer learning*, *Proc. IEEE* **109**, 43 (2020).
- [67] M. Scheffer *et al.*, *Ecology of Shallow Lakes* (Springer, New York, 1998), Vol. 1.
- [68] F. Grziwotz, C.-W. Chang, V. Dakos, E. H. van Nes, M. Schwarzländer, O. Kamps, M. Heßler, I. T. Tokuda, A. Telschow, and C.-h. Hsieh, *Anticipating the occurrence and type of critical transitions*, *Sci. Adv.* **9**, eabq4558 (2023).
- [69] T. N. Kipf and M. Welling, *Semi-supervised classification with graph convolutional networks*, arXiv:1609.02907.
- [70] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, *Graph attention networks*, arXiv:1710.10903.
- [71] S. Hochreiter and J. Schmidhuber, *Long short-term memory*, *Neural Comput.* **9**, 1735 (1997).
- [72] <https://github.com/m-serious/tipping-predictor.git>.
- [73] J. J. Lever, I. A. van de Leemput, E. Weinans, R. Quax, V. Dakos, E. H. van Nes, J. Bascompte, and M. Scheffer, *Foreseeing the future of mutualistic communities beyond collapse*, *Ecol. Lett.* **23**, 2 (2020).
- [74] C. D. Brummitt, G. Barnett, and R. M. D'Souza, *Coupled catastrophes: Sudden shifts cascade and hop among interdependent systems*, *J. R. Soc. Interface* **12**, 20150712 (2015).
- [75] F. Battiston, E. Amico, A. Barrat, G. Bianconi, G. Ferraz de Arruda, B. Franceschiello, I. Iacopini, S. Kéfi, V. Latora, Y. Moreno *et al.*, *The physics of higher-order interactions in complex systems*, *Nat. Phys.* **17**, 1093 (2021).
- [76] R. Lambiotte, M. Rosvall, and I. Scholtes, *From networks to optimal higher-order models of complex systems*, *Nat. Phys.* **15**, 313 (2019).