



Predicting network dynamics without requiring the knowledge of the interaction graph

Bastian Prasse^{a,1} and Piet Van Mieghem^a

Edited by David Weitz, Harvard University, Cambridge, MA; received March 29, 2022; accepted September 18, 2022

A network consists of two interdependent parts: the network topology or graph, consisting of the links between nodes and the network dynamics, specified by some governing equations. A crucial challenge is the prediction of dynamics on networks, such as forecasting the spread of an infectious disease on a human contact network. Unfortunately, an accurate prediction of the dynamics seems hardly feasible, because the network is often complicated and unknown. In this work, given past observations of the dynamics on a fixed graph, we show the contrary: Even without knowing the network topology, we can predict the dynamics. Specifically, for a general class of deterministic governing equations, we propose a two-step prediction algorithm. First, we obtain a surrogate network by fitting past observations of every nodal state to the dynamical model. Second, we iterate the governing equations on the surrogate network to predict the dynamics. Surprisingly, even though there is no similarity between the surrogate topology and the true topology, the predictions are accurate, for a considerable prediction time horizon, for a broad range of observation times, and in the presence of a reasonable noise level. The true topology is not needed for predicting dynamics on networks, since the dynamics evolve in a subspace of astonishingly low dimension compared to the size and heterogeneity of the graph. Our results constitute a fresh perspective on the broad field of nonlinear dynamics on complex networks.

dynamics on networks | predicting dynamics | network reconstruction

The interplay of dynamics and structure lies at the heart of myriad processes on networks, ranging from predator–prey interactions on ecological networks (1) and epidemic outbreaks on physical contact networks (2) to brain activity on neural networks (3). To relate the network structure, or graph, and the process dynamics, there are two approaches of opposing directions. On the one hand, a large body of research (4–6) focuses on the question, What is the impact of the network structure on the dynamics of a process? For instance, what is the impact of the network of online social media friendships on the spread of fake news. On the other hand, network reconstruction methods (7–11) consider the inverse problem: Given some observations of the dynamics, to what extent can the network structure be inferred? As an example, one may ask to determine the path of an infectious virus from one individual to another, given observations of the epidemic outbreak.

The prediction of the dynamics on an unknown network seems to require the combination of both directions (see, e.g., the discussion in ref. 12): first, the reconstruction of the network structure based on past observations of the dynamics and, second, the estimation of the future dynamics based on the inferred network. Intuitively, one may expect that an accurate prediction of the dynamics is possible only if the network topology is available. In this work, paradoxically, we show the contrary: It is possible to accurately predict a general class of dynamics without the network structure!

1. Modeling Dynamics on Networks

The network's graph is represented by the $N \times N$ weighted adjacency matrix A whose elements are denoted by a_{ij} . If there is a directed link from node j to node i , then $a_{ij} > 0$; otherwise $a_{ij} = 0$. Hence, we focus on nonnegative entries $a_{ij} \geq 0$ in this work, which is in agreement with the considered empirical networks detailed in *SI Appendix, section A*. However, our prediction approach can be adjusted to entries $a_{ij} \in \mathbb{R}$ in a straightforward manner, as we argue in Section 3. Furthermore, we consider a fixed weighted adjacency matrix A and that the underlying graph does not change with time.

We denote the nodal state of node i at time t by $x_i(t)$ and the nodal state vector by $x(t) = (x_1(t), \dots, x_N(t))^T$. We consider a general class of dynamical models on networks (7, 12–14) that describe the evolution of the nodal state $x_i(t)$ of any node i as

$$\frac{dx_i(t)}{dt} = f_i(x_i(t)) + \sum_{j=1}^N a_{ij}g(x_i(t), x_j(t)). \quad [1]$$

Significance

Dynamics on networks describe a plethora of physical phenomena, including the viral spread on contact networks, the competition between species on predator–prey networks, and magnetoencephalography activity on the human connectome. Of particular interest is the prediction of dynamics on networks. While the network is decisive for the dynamics, the precise network structure is unknown in most applications. Thus, it seems necessary to reconstruct the underlying network, which constitutes a tremendous, if not infeasible, obstacle to predicting dynamics. Here, we show the opposite: The prediction of a general class of dynamics may be possible, even if the underlying graph cannot be reconstructed. Our work is an important step toward reducing the obstacle of an accurate network reconstruction to predict dynamics.

Author affiliations: ^aFaculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, 2600 GA Delft, The Netherlands

Author contributions: B.P. and P.V.M. designed research; B.P. performed research; B.P. and P.V.M. analyzed data; and B.P. and P.V.M. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2022 the Author(s). Published by PNAS. This article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

¹To whom correspondence may be addressed. Email: bastian.prasse@rwth-aachen.de.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2205517119/-DCSupplemental>.

Published October 24, 2022.

Table 1. Models of dynamics on networks

Model	$f_i(x_i(t))$	$g(x_i(t), x_j(t))$
LV	$x_i(t)(\alpha_i - \theta_i x_i(t))$	$-x_i(t)x_j(t)$
MP	$x_i(t)(\alpha_i - \theta_i x_i(t))$	$x_i(t)x_j^2(t)(1 + x_j^2(t))^{-1}$
MM	$-x_i(t)$	$x_j^h(t)(1 + x_j^h(t))^{-1}$
SIS	$-\delta_i x_i(t)$	$(1 - x_i(t))x_j(t)$
KUR	ω_j	$\sin(x_i(t) - x_j(t))$
WC	$-x_i(t)$	$(1 + \exp(-\tau(x_j(t) - \mu)))^{-1}$

The function $f_i(x_i(t))$ describes the self-dynamics of node i . The sum in [1] represents the interactions of node i with its neighbors. The interaction between two nodes i and j depends on the adjacency matrix A and the interaction function $g(x_i(t), x_j(t))$. A broad spectrum of models follows from [1] by specifying the self-dynamics function f_i and the interaction function g . For instance, by specifying $f_i(x_i(t)) = 0$ and $g(x_i(t), x_j(t)) = -Ax_j(t)$, the model [1] reduces to the linear dynamics $dx(t)/dt = -Ax(t)$. In this work, we study six particular models of dynamics on networks, which are summarized in Table 1.

1.1. Lotka–Volterra Population Dynamics. The Lotka–Volterra model (LV) (15) describes the population dynamics of competing species. The nodal state $x_i(t)$ denotes the population size of species i , the growth parameters of species i equal $\alpha_i > 0$ and $\theta_i > 0$, and the link weight a_{ij} quantifies the competition rate, or predation rate, of species j on species i .

1.2. Mutualistic Population Dynamics. We adopt the model of Harush and Barzel (16) to describe mutualistic population dynamics (MP). The nodal state $x_i(t)$ denotes the population size of species i , the growth parameters of species i are denoted by $\alpha_i > 0$ and $\theta_i > 0$, and the link weight $a_{ij} > 0$ quantifies the strength of mutualism between species i and species j .

1.3. Michaelis–Menten Regulatory Dynamics. The dynamics of gene regulatory networks can be described by the Michaelis–Menten equation (16–18), Michaelis–Menten regulatory dynamics (MM). The nodal state $x_i(t)$ is the expression level of gene i , the Hill coefficient is denoted by h , and the link weights $a_{ij} > 0$ are the reaction rate constants.

1.4. Susceptible–Infected–Susceptible Epidemics. Spreading phenomena, such as the epidemic of an infectious disease, can be described by the susceptible–infected–susceptible model (SIS) (2, 19–21). The nodal state $x_i(t)$ equals the infection probability of node i . The parameter $\delta_i > 0$ denotes the curing rate, and the link weight a_{ij} is the infection rate from node j to node i .

1.5. Kuramoto Oscillators. The Kuramoto model (KUR) (22) has been applied to various synchronization phenomena of phase oscillators (23), such as magnetoencephalography (MEG) activity of brain regions (3). Here, the nodal state $x_i(t)$ corresponds to the phase of oscillator i , the parameter ω_i denotes the natural frequency of node i , and the coupling strength from node j to node i is given by the link weight a_{ij} .

1.6. Wilson–Cowan Neural Firing. The firing rates of neurons can be described by the Wilson–Cowan model (WC) (14, 24). Here, the nodal state $x_i(t)$ is the activity of neuron i , and the parameters τ and μ are the slope and the threshold of the neural activation function. The link weight a_{ij} specifies the number and strength of synapses from neuron j to neuron i .

As stated in ref. 25, there are three possibilities for the qualitative long-term behavior of the dynamical system [1]. First, the

nodal state $x(t)$ might approach a steady state $x_\infty = \lim_{t \rightarrow \infty} x(t)$.

At the steady state x_∞ , the nodal state does not change any longer; thus $dx(t)/dt = 0$. Second, the nodal state $x(t)$ might converge to a limit cycle, a curve on which the nodal state $x(t)$ circulates forever. Third, the nodal state $x(t)$ might never come to rest, nor enter a repeating cycle. Then, the state $x(t)$ perpetually continues to move in an irregular pattern.

Additionally to the functions $f_i(x_i(t))$ and $g(x_i(t), x_j(t))$, given by Table 1, the differential equations [1] require the specification of the adjacency matrix A . Each dynamical process above evolves on a respective real-world network: LV, food web of Little Rock Lake (26); MP, mutualistic insect interactions (27, 28); MM, gene regulatory network of the yeast *Saccharomyces cerevisiae* (29); SIS, face-to-face contacts between visitors of the “Infectious: Stay Away” exhibition (30); KUR, structural connectivity between brain regions (31, 32); and WC, *Caenorhabditis elegans* neuronal connectivity (33, 34). The real-world networks specify an adjacency matrix A , based on which we simulate the dynamics [1]. Some of the real-world networks are disconnected. Throughout the work, we consider the whole network, which may consist of multiple connected components, without confining to the largest connected component. *SI Appendix, section A* states the real-world networks and model parameters in detail.

The focus of this work is to predict the nodal state $x(t)$ at times $t > t_{\text{obs}}$, where t_{obs} denotes the observation time. Furthermore, we focus on predicting transient dynamics of the nodal state $x(t)$. The input to the prediction is given by $n + 1$ nodal state observations $x(0), x(\Delta t), \dots, x(n\Delta t)$ in the transient regime of the dynamics.* Here, $\Delta t > 0$ denotes the sampling time with $n\Delta t = t_{\text{obs}}$. We emphasize that we do not assume any knowledge of the matrix A for our prediction algorithm.

In refs. 35–37, sophisticated methods were derived to compute the steady state $x_\infty = \lim_{t \rightarrow \infty} x(t)$ from only a few observations of the nodal state vector $x(t)$ for consensus dynamics on graphs. There are two substantial differences between the studies (35–37) and our work. First, beyond predicting only the steady state x_∞ , we predict the complete dynamics of the nodal state $x(t)$ at all times $t \geq t_{\text{obs}}$. We stress that the equilibrium $x_{\infty, i}$ of the consensus dynamics in refs. 35–37 is, by definition, the same for all nodes i . In contrast, for the dynamics [1], the nodal states $x_i(t)$ of different nodes i are typically strongly heterogeneous. Second, consensus dynamics are linear, but we study nonlinear dynamics [1] on networks. In contrast to linear dynamics, there is no general closed-form solution for any of the nonlinear dynamics specified above. Thus, predicting nonlinear dynamics [1] requires a fundamentally different approach than refs. 35–37, which build upon closed-form expressions of linear dynamics. In summary, the methods in refs. 35–37 do not address the central problem of our work: Nonlinearity, rather than linearity, is the norm for dynamical processes on networks (5, 25, 38), and the complete future dynamics provide a considerably richer picture than the long-term equilibria.

In our work, we assume that the self-dynamics function $f_i(x_i(t))$ and the interaction function $g(x_i(t), x_j(t))$ are known, for two reasons. First, when describing the nodal state $x(t)$ of a real-world process, the underlying kind of physical process is usually known, which specifies the functions f_i and g at least approximately. Second, the focus of our work is on the

*We denote the number of observations by $n + 1$ for simplicity: Given $n + 1$ observations, there are n nodal state transitions from $x(k\Delta t)$ to $x((k + 1)\Delta t)$ with the discrete time $k = 0, \dots, n$. The n nodal state transitions form the basis for our prediction algorithm in Section 3.

fundamental interdependence of the network topology and the general dynamics [1]. While assuming the functions $f_i(x_i(t))$ and $g(x_i(t), x_j(t))$ as unknowns would make predictions more challenging, our fundamental result would remain unaffected: An accurate network reconstruction is neither possible nor necessary to predict the trajectory of the nodal state $x(t)$. We emphasize that the observations and predictions of the nodal state are on the same trajectory, which starts at some initial state $x(0)$. Lastly, we refer to the recent proposal of Gao and Yan (12) for a setting reverse to our work, where the adjacency matrix A is known but the functions f_i and g are unknown.

2. Dynamics on Networks Are Low Dimensional

The number N of nodes or the size N of real-world networks can be very large. What is the impact of the network size N on the predictability of dynamics? Intuitively, it seems that more nodes N lead to a more complicated dynamics. Is there a maximum network size N , above which an accurate prediction is impossible? In this section, we report two observations. First, the dynamics on networks are in a subspace $\mathcal{X} \subset \mathbb{R}^N$ of small dimension $m \ll N$. The dimension m equals the degrees of freedom of the network dynamics. The smaller the dimension m , the simpler the dynamics that is to be predicted. Second, we infer the low-dimensional subspace \mathcal{X} from the dynamics in a short observation time interval. Thus, predicting dynamics on a network with N nodes simplifies to predicting $m \ll N$ degrees of freedom.

Our analysis relies on the proper orthogonal decomposition (POD) (39–43), which is a powerful tool for discovering low-dimensional structures in dynamics. At any time t , the POD approximates the $N \times 1$ nodal state vector $x(t)$ by

$$x(t) \approx \sum_{p=1}^m c_p(t) y_p. \quad [2]$$

Here, the agitation modes y_1, \dots, y_m are orthonormal vectors, which do not change over time t . The scalar functions $c_p(t)$ are obtained by projecting the nodal state vector $x(t)$ on the respective agitation mode,

$$c_p(t) = y_p^T x(t). \quad [3]$$

The number of agitation modes m is a parameter which determines the accuracy of the POD [2]. The more agitation modes m , the more degrees of freedom of the POD [2]. Hence, the more modes m , the more accurate the approximation [2]. More precisely, numerical results in *SI Appendix, section B.1* suggest that the approximation error (with respect to the Euclidean norm over a finite time interval) of the POD [2] decreases exponentially quickly as the number of agitation modes m increases. If $m = N$, then the approximation [2] is exact, because any $N \times 1$ vector $x(t)$ can be written as the linear combination of N orthogonal vectors. Intuitively speaking, if the POD [2] is accurate for $m \ll N$ modes or degrees of freedom, then the nodal state vector $x(t)$ is barely agitated.

If the POD [2] is accurate, then the nodal state $x(t)$ is practically an element of the m -dimensional subspace $\mathcal{X} = \text{span}\{y_1, \dots, y_m\}$, where the span of the vectors y_1, \dots, y_m equals the set of all linear combinations[†]

$$\text{span}\{y_1, \dots, y_m\} = \left\{ \sum_{p=1}^m \alpha_p y_p \mid \alpha_p \in \mathbb{R} \right\}.$$

What makes the POD [2] so interesting for predicting dynamics on networks? Suppose we know the agitation modes y_1, \dots, y_m for which the POD [2] is accurate at future times $t > t_{\text{obs}}$. Then, we must predict only m functions $c_1(t), \dots, c_m(t)$ to predict the $N \times 1$ nodal state vector $x(t)$. Thus, not the network size N but the number of agitation modes m is decisive for the difficulty of predicting dynamics on networks.

However, at the observation time t_{obs} , we do not know the agitation modes y_1, \dots, y_m for which the POD [2] is accurate at future times $t > t_{\text{obs}}$. In the following, we show that the agitation modes y_p can be estimated from observing the dynamics from time $t = 0$ until $t = t_{\text{obs}}$. More precisely, we estimate the agitation modes y_p from the nodal state observations $x(0), \dots, x(n\Delta t)$ in two steps; see refs. 41, 42, and 44. First, we define the $N \times (n+1)$ nodal state matrix as $X = (x(0), x(\Delta t), \dots, x(n\Delta t))$. Second, we obtain the agitation modes y_1, \dots, y_m as the first m left-singular vectors of the nodal state matrix X .

Fig. 1 demonstrates the accuracy of the POD [2], with $m = 15 \ll N$ agitation modes. Here, the agitation modes y_p follow from the nodal state $x(t)$ until the observation time t_{obs} as stated above, and the scalar functions $c_p(t)$ are computed by [3]. Surprisingly, the POD [2] is accurate at times $t > t_{\text{obs}}$, even though the nodal state $x(t)$ at times $t > t_{\text{obs}}$ was not used for computing the agitation modes y_p . Hence, during the time interval $[0, t_{\text{obs}}]$, the nodal state $x(t)$ quickly locks into only a few agitation modes y_p , which govern the dynamics also at future times $t > t_{\text{obs}}$. The POD approximation error $\bar{\epsilon}$ in Fig. 1 varies for the different models (also after normalizing by the maximum magnitude of the model-specific nodal states $x_i(t)$), since the number of agitation modes m depends on the specific model (*SI Appendix, section B.2*), but Fig. 1 considers, for simplicity, the same number of agitation modes $m = 15$ for all dynamic models. Furthermore, in *SI Appendix, section B.2*, we argue that the number m of agitation modes is constant as the number of nodes N grows, for fixed time intervals. Hence, the dynamics remain at a given level of simplicity (a constant number m of agitation modes) despite an increasingly complex network (due to a larger number N of nodes and a more heterogeneous degree distribution as N grows).

We stress that the POD [2] cannot be used (directly) to predict the nodal state $x(t)$. Whereas the agitation modes y_p could be inferred within a small observation time interval $[0, t_{\text{obs}}]$, we do not know the functions $c_p(t)$ at future times $t \geq t_{\text{obs}}$. For Fig. 1, we computed the functions as $c_p(t) = y_p^T x(t)$, requiring the nodal state $x(t)$ at times $t \geq t_{\text{obs}}$. However, since the agitation modes y_p can be estimated from observing past dynamics, the crucial implication of Fig. 1 is that predicting dynamics on networks with N nodes reduces to predicting $m \ll N$ scalar functions $c_p(t)$. Hence, predicting future dynamics on unknown networks seems possible, even for large networks. In the remaining part of this work, we design and evaluate a concrete prediction algorithm.

3. Prediction Algorithm Based on a Surrogate Network

If we knew the adjacency matrix A , then we could predict the evolution of the nodal state $x(t)$ by numerically solving the differential Eq. 1. However, the true adjacency matrix A is unknown. Thus, we resort to predicting the dynamics by using a surrogate

[†]Strictly speaking, unless the POD [2] is exact, the vector $x(t)$ is not an element of the subspace \mathcal{X} . However, if the POD [2] is sufficiently accurate, then the difference $x(t) - \text{proj}_{\mathcal{X}}(x(t))$ is negligible. Here, we denote the projection of the nodal state $x(t)$ onto the subspace \mathcal{X} by $\text{proj}_{\mathcal{X}}(x(t))$, which is element of the subspace \mathcal{X} .

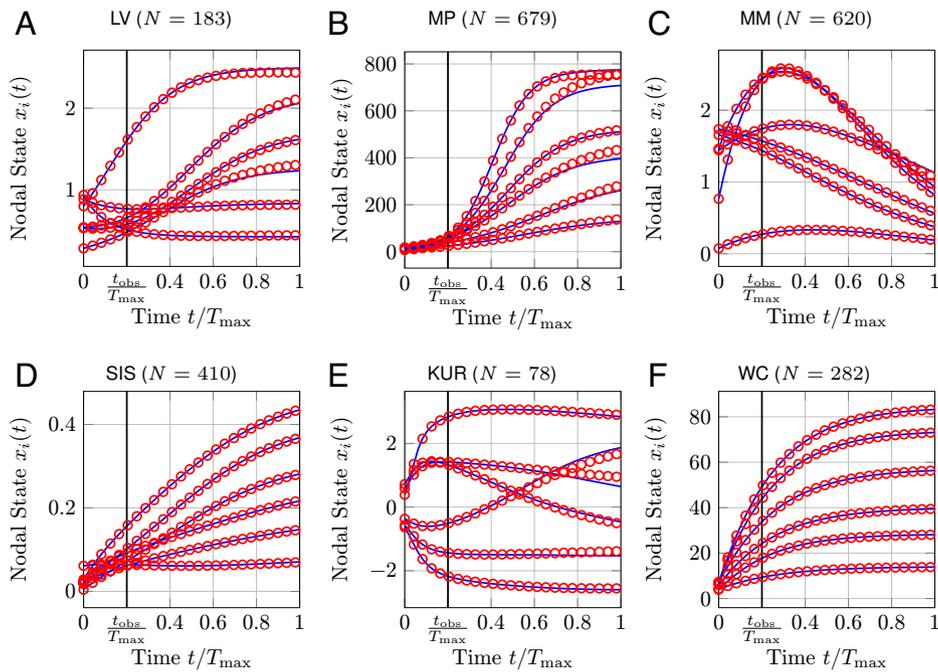


Fig. 1. POD of dynamics on networks. The exact nodal state $x(t)$ is shown in blue, and the approximation by the POD [2] is shown in red. The maximum prediction time T_{\max} is different for each dynamic model, and the observation time equals $t_{\text{obs}} = T_{\max}/5$. The number of observations is $n = 100$. For readability, only six nodal states $x_i(t)$ are depicted for each network. In this and the following figures, the six nodes i are chosen such that the six nodal states $x_i(T_{\max})$ at the maximum time T_{\max} are as evenly spaced as possible. The approximation equals the linear combination [2] of $m = 15$ agitation modes y_1, \dots, y_m , which are computed by observing the nodal state $x(t)$ from time $t = 0$ to $t = t_{\text{obs}}$. The average of the POD approximation error $\epsilon_i(t) = |x_i(t) - \sum_{p=1}^m c_p(t) (y_p)_i|$, with respect to all nodes i and future times $t \in (t_{\text{obs}}, T_{\max})$, is denoted by $\bar{\epsilon}$ and equals (A) $\bar{\epsilon} = 1.01 \cdot 10^{-2}$, (B) $\bar{\epsilon} = 2.17$, (C) $\bar{\epsilon} = 1.10 \cdot 10^{-3}$, (D) $\bar{\epsilon} = 7.70 \cdot 10^{-5}$, (E) $\bar{\epsilon} = 6.26 \cdot 10^{-2}$, and (F) $\bar{\epsilon} = 1.79 \cdot 10^{-3}$.

adjacency matrix \hat{A} with elements \hat{a}_{ij} .[‡] More specifically, we compute the nodal state prediction $\hat{x}(t)$ at times $t > t_{\text{obs}}$ with the prediction model

$$\frac{d\hat{x}_i(t)}{dt} = f_i(\hat{x}_i(t)) + \sum_{j=1}^N \hat{a}_{ij} g(\hat{x}_i(t), \hat{x}_j(t)) \quad [4]$$

with the initial condition $\hat{x}(t_{\text{obs}}) = x(t_{\text{obs}})$. We obtain the surrogate matrix \hat{A} by fitting the prediction model [4] to the past observations of the nodal state $x(t)$, as explained in more detail below.

Before presenting the details of how we extract the surrogate \hat{A} from observing the dynamics, we consider the fundamental question: When does the surrogate \hat{A} predict the dynamics accurately? The predicted nodal state $\hat{x}(t)$ is initialized at the observation time t_{obs} as $\hat{x}(t_{\text{obs}}) = x(t_{\text{obs}})$. Hence, the prediction $\hat{x}(t)$ is exact if $d\hat{x}(t)/dt = dx(t)/dt$ at all future times $t \geq t_{\text{obs}}$. By comparing the true dynamics [1] and the surrogate model [4], we find that the predictions are exact if

$$\sum_{j=1}^N \hat{a}_{ij} g(x_i(t), x_j(t)) = \sum_{j=1}^N a_{ij} g(x_i(t), x_j(t)) \quad [5]$$

for all nodes i at future times $t \geq t_{\text{obs}}$, where we replaced $g(\hat{x}_i(t), \hat{x}_j(t))$ by $g(x_i(t), x_j(t))$ on the left side of [5], because $\hat{x}(t) = x(t)$. The coupling function g is usually nonlinear. Thus, Eq. 5 is generally not linear with respect to the nodal state $x(t)$. However, we emphasize the crucial observation that, for all nodes

[‡]The sole purpose of the surrogate matrix \hat{A} lies in the prediction of the nodal state $x(t)$. Particularly, as we argue in Section 4, the matrix \hat{A} should not be interpreted as an estimate of the adjacency matrix A .

i and times $t \geq t_{\text{obs}}$, [5] represents a linear system with respect to the surrogate matrix \hat{A} . The linear system [5] determines which, and how many, surrogate matrices \hat{A} yield accurate predictions of the nodal state $x(t)$. The true adjacency matrix $\hat{A} = A$ is always a solution of [5]. But a linear system may have more than one solution if the set of linear equations is not of full rank. Hence, there may be surrogate matrices $\hat{A} \neq A$ that obey [5]. Then, the nodal state $x(t)$ can be predicted accurately based on surrogates \hat{A} which are different than the true adjacency matrix A .

We argue that, due to the low-dimensional dynamics of the nodal state $x(t)$, there are numerous surrogates \hat{A} that solve [5] and yield accurate predictions. For ease of exposition, we focus on SIS dynamics in the following, and we refer to *SI Appendix, section C* for other dynamics.

Example. For SIS dynamics, the coupling function equals $g(x_i(t), x_j(t)) = (1 - x_i(t))x_j(t)$, and the linear system [5] becomes

$$(1 - x_i(t)) \left(\hat{A}x(t) \right)_i = (1 - x_i(t)) (Ax(t))_i \quad [6]$$

for all nodes i . Hence, the surrogate \hat{A} predicts perfectly if $\hat{A}x(t) = Ax(t)$ at all times t . As shown in Section 2, the nodal state dynamics are low-dimensional. Suppose the POD [2] is exact. (When the POD [2] does not hold exactly, we can argue similarly; see *SI Appendix, section D*.) Then, [6] becomes

$$(1 - x_i(t)) \left(\hat{A} \sum_{p=1}^m c_p(t) y_p \right)_i = (1 - x_i(t)) \left(A \sum_{p=1}^m c_p(t) y_p \right)_i$$

for all nodes i , which is satisfied if

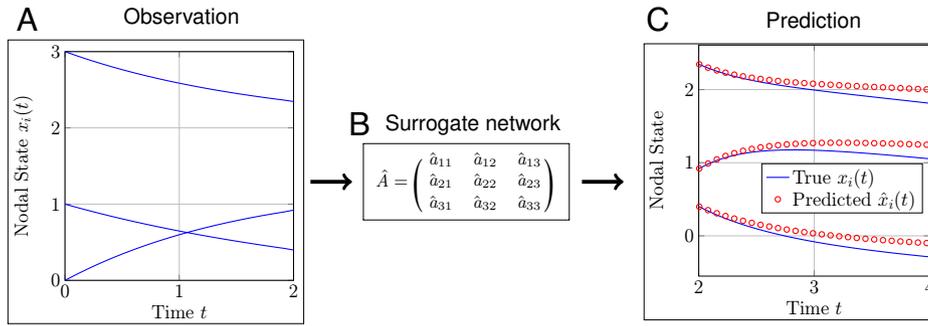


Fig. 2. Framework for predicting dynamics on unknown networks. This example shows a small network of $N = 3$ nodes. (A) The nodal state $x_i(t)$ is observed for all nodes i until the observation time $t_{\text{obs}} = 2$. The evolution of the nodal state $x_i(t)$ obeys the differential Eq. 1 with the known functions f_i and g and the unknown adjacency matrix A . (B) Based on the nodal state observations $x(0), x(\Delta t), \dots, x(n\Delta t)$ until the observation time $n\Delta t = t_{\text{obs}}$, we obtain the surrogate matrix \hat{A} from the optimization problem [9]. (C) For any time $t \geq t_{\text{obs}}$, we predict the dynamics by the model [4], which follows from the dynamics [1] by replacing the unknown matrix A with the surrogate matrix \hat{A} . The nodal state prediction is initialized as $\hat{x}_i(t_{\text{obs}}) = x_i(t_{\text{obs}})$ for all nodes i .

$$\hat{A} \sum_{p=1}^m c_p(t) y_p = A \sum_{p=1}^m c_p(t) y_p.$$

Thus, the surrogate \hat{A} yields exact predictions of the nodal state $x(t)$ if

$$\hat{A} y_p = A y_p \quad [7]$$

for every agitation mode $p = 1, \dots, m$. The linear system [7] has N^2 unknowns, namely, the entries \hat{a}_{ij} of the surrogate \hat{A} . But there are only mN equations, namely, the vectors $A y_1, \dots, A y_m$, each with N entries. As shown in Section 2, the number of agitation modes m is much smaller than the number of nodes N . Thus, there is a dramatic difference between the number of equations mN and the number N^2 of unknowns \hat{a}_{ij} . Since $mN \ll N^2$, there are countless surrogate matrices $\hat{A} \neq A$ that solve [7] and, thus, perfectly predict the nodal state $x(t)$.[§]

Our prediction framework for dynamics on unknown networks is illustrated by Fig. 2. We rely on two steps to obtain a unique surrogate matrix \hat{A} from the nodal state observations $x(0), x(\Delta t), \dots, x(n\Delta t)$. For simplicity, we introduce the notation $x_i[k] = x_i(k\Delta t)$ for all observation times $k = 0, 1, \dots, n$. First, we approximate the derivative $dx_i(t)/dt$ in [4] by a difference quotient,

$$\frac{x_i[k+1] - x_i[k]}{\Delta t} \approx f_i(x_i[k]) + \sum_{j=1}^N \hat{a}_{ij} g(x_i[k], x_j[k]). \quad [8]$$

Second, we obtain the surrogate \hat{A} that minimizes the difference of the left and right side of [8]. Specifically, we obtain the surrogate \hat{A} by solving

$$\begin{aligned} \underset{\hat{a}_{i1}, \dots, \hat{a}_{iN}}{\text{argmin}} \quad & \sum_{k=0}^{n-1} \left(\frac{x_i[k+1] - x_i[k]}{\Delta t} - f_i(x_i[k]) \right. \\ & \left. - \sum_{j=1}^N \hat{a}_{ij} g(x_i[k], x_j[k]) \right)^2 + \rho_i \sum_{j=1}^N \hat{a}_{ij} \quad [9] \\ \text{s.t.} \quad & \hat{a}_{ij} \geq 0 \quad j = 1, \dots, N \end{aligned}$$

for every node i . In [9], the scalar $\rho_i > 0$ denotes the regularization parameter, which is set by hold-out cross-validation (46). The

optimization problem [9] is known as the least absolute shrinkage and selection operator (LASSO) (47–50), which is an established and powerful approach to infer the network structure from its dynamics (7, 8, 51–53). The first, sum-of-squares, term in [9] fits the surrogate model [4] to the observations $x[0], \dots, x[n]$. As illustrated by the *Example*, numerous surrogate matrices \hat{A} result in the same value of $\sum_{j=1}^N \hat{a}_{ij} g(x_i[k], x_j[k])$, due to the low-dimensional dynamics of the nodal state $x(t)$. Thus, there are many surrogates \hat{A} that minimize the first term in [9]. By including the second, ℓ_1 -regularization, term $\rho_i \sum_{j=1}^N \hat{a}_{ij}$, we obtain a well-defined optimization problem. More precisely, the second term results in a sparser solution \hat{A} of the LASSO [4]; see refs. 47 and 50. The larger the regularization parameter $\rho_i > 0$, the sparser the solution $\hat{a}_{i1}, \dots, \hat{a}_{iN}$ to the LASSO [4]. Hence, the surrogate in-degree \hat{d}_i of node i , which equals the number of entries $\hat{a}_{i1}, \dots, \hat{a}_{iN}$ that are positive, decreases as the regularization parameter ρ_i increases. The LASSO formulation [4] does not constrain the surrogate network \hat{A} to be connected. Indeed, as we show in *SI Appendix, section H*, the resulting surrogate network \hat{A} may have multiple disconnected components, even if the true matrix A is connected.

For all considered dynamics, the corresponding empirical network A detailed in *SI Appendix, section A* has nonnegative entries $a_{ij} \geq 0$. Hence, the solution \hat{A} to the LASSO optimization problem [9] is more accurate when the nonnegativity constraint $\hat{a}_{ij} \geq 0$ is included. Additionally, the optimization problem [9] could be adjusted to estimate adjacency matrices A with entries $a_{ij} \in \mathbb{R}$ by omitting the constraint $\hat{a}_{ij} \geq 0$. In *SI Appendix, section E*, we state the details of our method to extract the surrogate matrix \hat{A} .

Fig. 3 shows the performance of our prediction method. The predictions are surprisingly accurate. In particular, we can accurately predict the nodal state $x(t)$ until 5 times the observation time interval, $t = 5t_{\text{obs}}$, except for the Kuramoto model. The predictions for the Kuramoto are accurate until $t \approx 2t_{\text{obs}}$. The Kuramoto oscillators are the only dynamics in Fig. 3 that do not converge to a steady state x_{∞} . Hence, the Kuramoto dynamics are significantly more complex, which explains the worse prediction accuracy.

We chose the maximum time T_{max} in Fig. 3 such that the observed dynamics look sufficiently rich. We refer to *SI Appendix, section G* for an extensive sensitivity analysis of the prediction method, with respect to the maximum time T_{max} , the observation time t_{obs} , the network size N , model errors, and heterogeneous

[§]The rank of the linear system [7] equals mN . Hence, if $mN < N^2$, then the linear system is underdetermined (45), and there are infinitely many solutions for the surrogate \hat{A} .

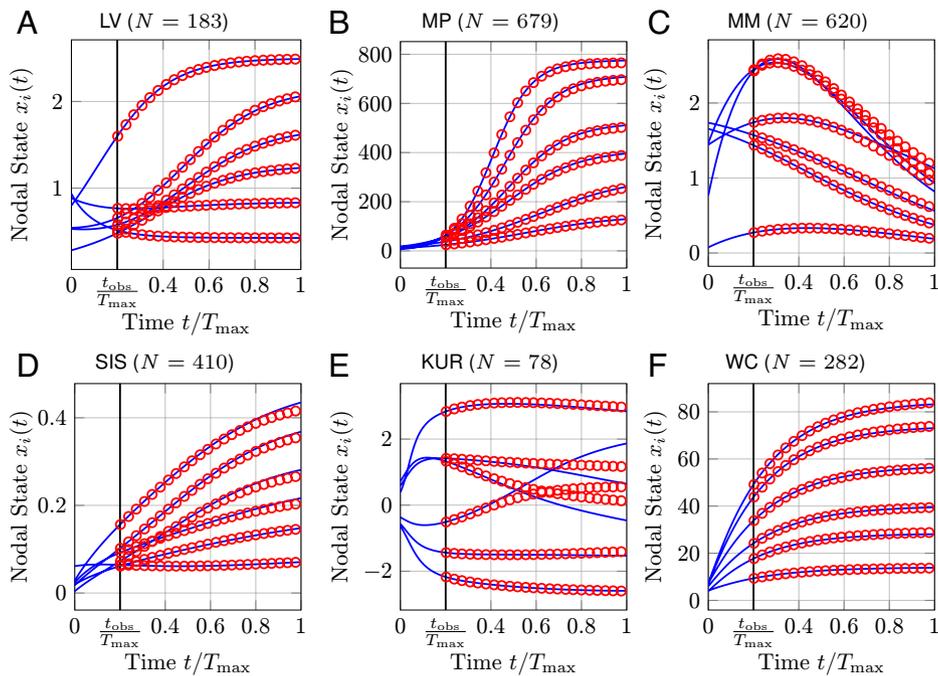


Fig. 3. Accuracy of the prediction method. Based on $n = 100$ nodal state observations $x(0), x(\Delta t), \dots, x(n\Delta t)$ until time $t_{\text{obs}} = n\Delta t$, the nodal state $x(t)$ is predicted at times $t > t_{\text{obs}}$. The blue curves are the true nodal states $x_i(t)$. The red marks are the nodal state predictions $\hat{x}_i(t)$ based on the surrogate matrix \hat{A} , initialized as $\hat{x}(t_{\text{obs}}) = x(t_{\text{obs}})$. For clarity, only six nodal states $x_i(t)$ are depicted for each network. The average of the prediction error $\epsilon_i(t) = |x_i(t) - \hat{x}_i(t)|$, with respect to all nodes i and future times $t \in [t_{\text{obs}}, T_{\text{max}}]$, is denoted by $\bar{\epsilon}$ equals (A) $\bar{\epsilon} = 1.83 \cdot 10^{-3}$, (B) $\bar{\epsilon} = 3.87 \cdot 10^{-1}$, (C) $\bar{\epsilon} = 1.37 \cdot 10^{-2}$, (D) $\bar{\epsilon} = 1.26 \cdot 10^{-3}$, (E) $\bar{\epsilon} = 0.1$, and (F) $\bar{\epsilon} = 8.79 \cdot 10^{-2}$.

coupling functions g_i . Simulation results in *SI Appendix, section J* suggest that the prediction algorithm has a quasi-polynomial runtime.

In *SI Appendix, section K*, we propose a predictability parameter that assesses the fundamental limitations of predicting the dynamics [1] from nodal state observations. The predictability parameter is closely related to the Lyapunov exponent and connects the prediction accuracy of our approach with potentially chaotic dynamics, which are subject to fundamental prediction limits (54).

4. The Surrogate Network Topology

As shown in Fig. 3, the surrogate matrix \hat{A} yields accurate nodal state predictions $\hat{x}(t)$. Does the high prediction accuracy imply a similarity of surrogate network topology with the true network topology? Here, we make a clear distinction between the network topology and the interaction strengths (55). The network topology, graph or network structure, is the set of all links: all node pairs (i, j) for which $a_{ij} > 0$. If there is a link from node j to node i , then the interaction strength is specified by the link weight a_{ij} . For instance, consider the two 3×3 adjacency matrices

$$A = \begin{pmatrix} 0 & 0.1 & 0 \\ 2.5 & 0 & 0 \\ 0 & 3 & 1 \end{pmatrix}, \quad \hat{A} = \begin{pmatrix} 0 & 9 & 0 \\ 0.7 & 0 & 0 \\ 0 & 0.5 & 3 \end{pmatrix}.$$

For all nodes i, j , it holds that $a_{ij} > 0$ if and only if $\hat{a}_{ij} > 0$. Hence, the two matrices A and \hat{A} have the same network topology. However, the interaction strengths, for instance, from node 2 to node 1, is different, because $a_{12} = 0.1$ but $\hat{a}_{12} = 9$.

We quantify the similarity of the networks A and \hat{A} by two topological metrics. First, we consider the area under the receiver operating characteristic (ROC) curve (AUC) (56), which we compute with the Matlab command `perfcurve`. To compute the

AUC, we consider a given rounding threshold $\epsilon \geq 0$. Then, the true positive rate $\text{TPR}(\epsilon) \in [0, 1]$ equals the number of node pairs i, j for which both $\hat{a}_{ij} \geq \epsilon$ and $a_{ij} > 0$, divided by the number of entries $a_{ij} > 0$. Similarly, the false positive rate $\text{FPR}(\epsilon)$ equals the number of node pairs i, j for which $\hat{a}_{ij} \geq \epsilon$ but $a_{ij} = 0$, divided by the number of entries $a_{ij} = 0$. The ROC curve is obtained by plotting the rate $\text{TPR}(\epsilon)$ versus the rate $\text{FPR}(\epsilon)$ for $\epsilon \geq 0$, and the AUC equals the area under the ROC curve. If the surrogate \hat{A} were obtained by tossing a coin for every possible link, then the corresponding AUC would be 0.5. The closer the AUC is to one, the greater the similarity of the surrogate topology to the true topology.

Second, we consider the in-degree distribution of the matrices A and \hat{A} . The (unweighted) in-degree d_i of node i equals the number of links that end at node i . The surrogate network algorithm detailed in *SI Appendix, section E* does generate entries $\hat{a}_{ij} = 0$ that are exactly equal to zero, and we define the estimated degree \hat{d}_i of node i as the number of strictly positive surrogate entries $\hat{a}_{ij} > 0, j = 1, \dots, N$. The in-degree distribution is given by $\Pr[D \geq d]$, where D is the degree of a randomly chosen node in the network.

Fig. 4 compares the surrogate network \hat{A} to the true network A . We emphasize that Fig. 4 compares a single realization of the true matrix A and an initial nodal state $x[0]$, namely, the same matrix A and nodal state $x[0]$ that generated the nodal dynamics shown in Fig. 3. The AUC value is almost 0.5 for all models. Hence, the surrogate network topology is completely different from the true network topology! Moreover, the degree distribution $\Pr[D \geq d]$ of the surrogate network differs strongly from the degree distribution of the true network, except for Fig. 4 C and F. We remark that, even if two networks have the same degree distribution $\Pr[D \geq d]$, the network topologies can be entirely different. For instance, the AUC value equals only 0.53 in Fig. 4F.

In *SI Appendix, section F*, we show that a similar contrast of prediction accuracy and topological similarity also holds for

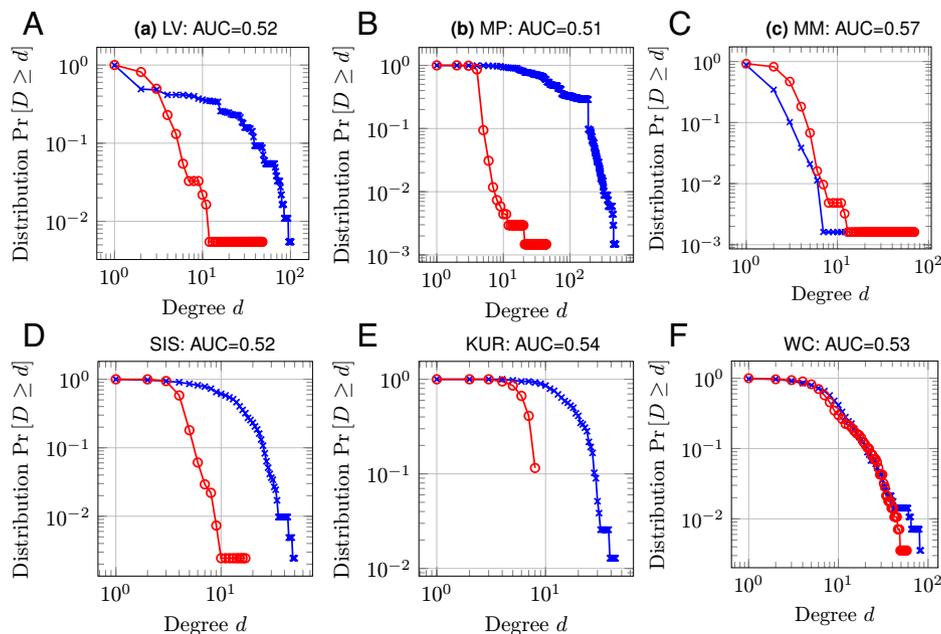


Fig. 4. The surrogate topology versus the true topology. Shown is a comparison of the topologies of the surrogate network \hat{A} with the true network A with respect to two topological metrics. The parameters are the same as in Fig. 3. First, the AUC value of the surrogate network \hat{A} , which equals (A) AUC = 0.52, (B) AUC = 0.51, (C) AUC = 0.57, (D) AUC = 0.52, (E) AUC = 0.54, and (F) AUC = 0.53. Second, the in-degree distributions $\text{Pr}[D \geq d]$ for the surrogate matrix \hat{A} are shown in red and, for the true matrix A , in blue.

random graphs. For a discussion on potential rotational symmetries of the surrogate network \hat{A} and the true network A and a comparison with respect to the eigenvalue spectra of the two matrices, we refer to [SI Appendix, section F.1](#).

5. Discussion

The prediction of general dynamics on unknown networks is studied, based on past observations of the dynamics. We propose a prediction framework which consists of two steps. First, we obtain a surrogate network by fitting the dynamical model to the past observations. Our fitting method is based on the LASSO. Second, we predict the dynamics by computing the dynamical model in [1] where the true adjacency matrix A is replaced by the surrogate's adjacency matrix \hat{A} . Counterintuitively, even though the surrogate network topology has no similarity with the true topology, the predictions are accurate, for a considerable prediction time horizon, for a broad range of observation times, and in the presence of a reasonable noise level! The true network topology is not needed for accurate predictions, because the network dynamics are barely agitated.

The observation that dynamics on networks can be predicted without the true topology has far-reaching consequences. The majority of network topologies are complicated, and a sufficiently accurate network reconstruction is a difficult, perhaps impossible, task in practice. However, in this work, we reveal a stark contrast: The network graph is complicated and large, but the nodal state dynamics follow a simple linear combination of only a few agitation modes y_1, \dots, y_m . Hence, our results suggest a promising research direction for dynamics on networks: Rather than the interplay between all the numerous nodes, dynamics on networks can be understood as the interplay of a few agitation modes. On a conceptual level, the nonlinear dynamics [1] behave surprisingly similarly to a linear system $dx(t)/dt = Ax(t)$, for which the agitation modes y_p directly follow from the eigenvectors and eigenvalues of the matrix A .

It is an open question whether it is possible to generalize our prediction method to time-varying graphs, which seems challenging, since there are more processes active on a time-varying network: 1) the dynamics on the graph, 2) the process that changes the graph, and 3) possibly a coupling process between processes 1 and 2. For example, when COVID-19 spreads in a population, we distinguish between 1) the viral infection process, 2) the human mobility process that generates the underlying time-variant contact graph, and 3) a coupling or interference process due to awareness or observation of infections in a close neighborhood that may inspire individuals to change contacts or wear protection. If the network varies over time, it is unclear whether there are time-invariant agitation modes y_p . Time-varying networks can be aggregated to obtain a static network, or a sequence of static networks that correspond to different time intervals. For different methods of constructing static networks from time-varying networks and their limitations, in the context of epidemics on networks, we refer to refs. 57–60. In ref. 61, a method was proposed to embed time-varying networks into a low-dimensional space. An interesting future research direction is to explore a possible interplay of low-dimensional representations of time-varying networks with the agitation of the nodal state $x(t)$.

We emphasize that the agitation modes depend on the initial state $x(0)$. Consequently, also the surrogate network \hat{A} depends on the initial state $x(0)$. Thus, using the same surrogate matrix \hat{A} for predicting dynamics $\tilde{x}(t)$ with a different initial state $\tilde{x}(0) \neq x(0)$ results in a lower prediction accuracy than for the trajectory $x(t)$ starting with the initial condition $x(0)$. If the agitation modes \tilde{y}_p and y_p of the trajectories $\tilde{x}(t)$ and $x(t)$ are very similar, then we can expect that the surrogate network \hat{A} yields an accurate prediction also for the trajectory $\tilde{x}(t)$. Furthermore, in [SI Appendix, section I](#), we consider that multiple trajectories are observed on the same network. We show that there is an increasing benefit, as the number of observed trajectories grows, for predicting the future of the dynamical system with another initial nodal state. In summary, it is possible to accurately predict a

trajectory $\tilde{x}(t)$ with different initial conditions $\tilde{x}(0) \neq x(0)$, even if the trajectory $\tilde{x}(t)$ has not been observed for long, provided that 1) the agitation modes \tilde{y}_p, y_p are similar or 2) sufficiently many trajectories were observed. For some applications, neither of these two conditions might be satisfied, for example, for local outbreaks of infectious diseases where the initial condition varies strongly. However, the prediction method could be applied, for instance, to the spreading of information on online social media networks, where the initial seeding of information frequently originates from the same set of nodes (which may correspond to news stations or influencers).

We confined to autonomous dynamics [1] without any control. In some applications (52, 62–64), it might be possible to control the nodal state $x(t)$ and there might be an additive control $u_i(t) \in \mathbb{R}$ to the dynamics [1] of one, or multiple, nodes i . If the control $(u_1(t), \dots, u_N(t))^T$ is high dimensional, then the dynamics of the nodal state $x(t)$ might not be low dimensional. For a sufficiently high-dimensional network dynamics, it is conceptually possible to apply the sparse identification of nonlinear dynamics (SINDy) algorithm by Brunton et al. (65), which may reconstruct the complete governing Eq. 1, that is, the adjacency matrix A and the functions f_i and g .[¶] Additionally, we refer to refs. 53 and 66 for model-free (i.e., without the knowledge of the functions f_i and g) network reconstruction methods, provided the network dynamics are sufficiently high dimensional.

We emphasize that we considered deterministic governing equations. Developing similar prediction methods that make use of a surrogate network for stochastic processes is an open research question. If there are mean-field equations for stochastic process that resemble [1], such as for the stochastic SIS process (21, 67), then the results of our work are at least applicable to stochastic processes in the parameter regimes where the mean-field equations are accurate. Furthermore, we assumed that the nodal state $x_i(t)$ is observed for every node i . While assuming that only some nodal states $x_i(t)$ can be observed is clearly an interesting generalization of our prediction method, our result that an accurate prediction is possible without requiring the underlying graph is unaffected.

The agitation modes y_p were extracted in a data-driven manner from past observations of the dynamics. Obtaining a more thorough, analytic, understanding of the connection between topology, network dynamics, and agitation modes y_p stands on the agenda of future research. Here, we would like to mention four points. First, provided that the network has an equitable partition, the POD [2] is exact, and the agitation modes y_p follow from the cells of the partition for a plethora of dynamical models

[¶]This is the case provided that the functions f_i and g are in the SINDy library of candidate functions, which must be preconstructed by the user of the SINDy algorithm.

1. R. May, *Stability and Complexity in Model Ecosystems* (Princeton Landmarks in Biology, Princeton University Press, 2001), vol. 6.
2. R. Pastor-Satorras, C. Castellano, P. Van Mieghem, A. Vespignani, Epidemic processes in complex networks. *Rev. Mod. Phys.* **87**, 925–979 (2015).
3. J. Cabral, M. L. Kringelbach, G. Deco, Exploring the network dynamics underlying brain activity during rest. *Prog. Neurobiol.* **114**, 102–131 (2014).
4. S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang, Complex networks: Structure and dynamics. *Phys. Rep.* **424**, 175–308 (2006).
5. A. Barrat, M. Barthelemy, A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, 2008).
6. M. A. Porter, J. P. Gleeson, *Dynamical Systems on Networks* (Frontiers in Applied Dynamical Systems: Reviews and Tutorials, Springer, 2016) vol. 4.
7. M. Timme, J. Casadiego, Revealing networks from dynamics: An introduction. *J. Phys. A Math. Theor.* **47**, 343001 (2014).
8. W.-X. Wang, Y.-C. Lai, C. Grebogi, Data based identification and prediction of nonlinear and complex dynamical systems. *Phys. Rep.* **644**, 1–76 (2016).
9. M. Newman, Network structure from rich but noisy data. *Nat. Phys.* **14**, 542–545 (2018).
10. T. P. Peixoto, Network reconstruction and community detection from dynamics. *Phys. Rev. Lett.* **123**, 128301 (2019).

(70–75).[#] Second, under some assumptions (76), if the network has a negligible degree correlation (18), then the dynamics can be approximated by the POD with one agitation mode y_1 . Third, if the basic reproduction number R_0 is close to one, then the SIS dynamics on any network reduce to $m = 1$ agitation mode y_1 , which enables the derivation of a closed-form solution (77). Fourth, on a complete graph, the SIS dynamics reduce to $m = 2$ agitation modes y_1 and y_2 , which, again, enables the derivation of a closed-form solution (75). We believe that the results (75, 77), which relate agitation modes and network structure for the SIS process, can be extended to obtain a deeper understanding of the general dynamics [1].

Our discovery, showing that only a few agitation modes are relevant for predicting general dynamics on static networks, might provide insight into why deep learning methods are working so wonderfully well irrespective of the network topology (78).^{||} We show that only one trajectory of the process (e.g., one epidemic outbreak) suffices to learn the essence of future dynamics, but not the graph structure that couples individual dynamical processes. As shown in *SI Appendix, section I*, increasing the number of realizations (i.e., “learning more”) further increases the prediction accuracy, while gradually revealing the underlying graph. Although deep learning dynamics is not directly described by [1], the process appears similar: A neural network architecture remains fixed, and only the neural network weights are changed recursively during learning (79).

5.1. Data, Materials, and Software Availability. Code and data (with the exception of human brain network data) have been deposited in GitHub (<https://github.com/bprasse/Predicting-network-dynamics-without-the-graph>) (80). Some study data are available (human brain network detailed in *SI Appendix, section A.5* was shared with us by Prejaas Tewarie; researchers can access this data by contacting Prejaas Tewarie, Department of Clinical Neurophysiology and MEG Center, Amsterdam UMC, VU University Amsterdam. p.tewarie@amsterdamumc.nl).

ACKNOWLEDGMENTS. We thank Prejaas Tewarie for providing data on the structural brain network. Furthermore, we are grateful to Baruch Barzel and Stojan Trajanovski for helpful discussions regarding this work. Part of this research has been funded by the European Research Council under the European Union's Horizon 2020 research and innovation program (Grant Agreement 101019718).

[#]Consider a partition \mathcal{P} of the node set $\mathcal{N} = \{1, \dots, N\}$ into m disjoint subsets, $\mathcal{P} = \{\mathcal{N}_1, \dots, \mathcal{N}_m\}$. Then, the partition \mathcal{P} is called equitable (68–70) if, for any two subsets $\mathcal{N}_p, \mathcal{N}_j$, it holds that $\sum_{k \in \mathcal{N}_i} a_{ik} = \sum_{k \in \mathcal{N}_j} a_{jk}$ for all nodes $i, j \in \mathcal{N}_p$.

^{||}Deep learning is applicable in a wide range of problems beyond network science (e.g., image classification), and we admit the widely accepted perspective that its success is due to 1) accumulated big amount of data, 2) people's tolerance in accepting theoretically inaccurate predictions as long as they are computed within a decent latency, and 3) advancement in computational power.

11. A. Das, I. R. Fiete, Systematic errors in connectivity inferred from activity in strongly recurrent networks. *Nat. Neurosci.* **23**, 1286–1296 (2020).
12. T.-T. Gao, G. Yan, Autonomous inference of complex network dynamics from incomplete and noisy data. *Nat. Comput. Sci.* **2.3**, 160–168 (2022).
13. B. Barzel, A.-L. Barabási, Universality in network dynamics. *Nat. Phys.* **9**, 673–681 (2013).
14. E. Laurence, N. Doyon, L. J. Dubé, P. Desrosiers, Spectral dimension reduction of complex dynamical networks. *Phys. Rev. X* **9**, 011042 (2019).
15. R. MacArthur, Species packing and competitive equilibrium for many species. *Theor. Popul. Biol.* **1**, 1–11 (1970).
16. U. Harush, B. Barzel, Dynamic patterns of information flow in complex networks. *Nat. Commun.* **8**, 2181 (2017).
17. U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits* (CRC, 2006).
18. J. Gao, B. Barzel, A.-L. Barabási, Universal resilience patterns in complex networks. *Nature* **530**, 307–312 (2016).
19. N. T. J. Bailey, *The Mathematical Theory of Infectious Diseases and Its Applications* (Charles Griffin, London, ed. 2, 1975).
20. A. Lajmanovich, J. A. Yorke, A deterministic model for gonorrhoea in a nonhomogeneous population. *Math. Biosci.* **28**, 221–236 (1976).
21. P. Van Mieghem, J. Omic, R. Kooij, Virus spread in networks. *IEEE/ACM Trans. Netw.* **17**, 1–14 (2009).
22. Y. Kuramoto, *Chemical Oscillations, Waves, and Turbulence* (Courier Corporation, 2003).

23. T. Stankovski, T. Pereira, P. V. McClintock, A. Stefanovska, Coupling functions: Universal insights into dynamical interaction mechanisms. *Rev. Mod. Phys.* **89**, 045001 (2017).
24. H. R. Wilson, J. D. Cowan, Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys. J.* **12**, 1–24 (1972).
25. S. H. Strogatz, Exploring complex networks. *Nature* **410**, 268–276 (2001).
26. N. D. Martinez, Artifacts or attributes? Effects of resolution on the Little Rock Lake food web. *Ecol. Monogr.* **61**, 367–392 (1991).
27. M. Kato, T. Kakutani, T. Inoue, T. Itino, Insect-flower relationship in the primary beech forest of Ashu, Kyoto: An overview of the flowering phenology and the seasonal pattern of insect visits. *Contrib. Biol. Lab. Kyoto Univ.* **27**, 309–375 (1990).
28. E. L. Rezende, J. E. Lavabre, P. R. Guimarães, P. Jordano, J. Bascompte, Non-random coextinctions in phylogenetically structured mutualistic networks. *Nature* **448**, 925–928 (2007).
29. R. Milo *et al.*, Network motifs: Simple building blocks of complex networks. *Science* **298**, 824–827 (2002).
30. L. Isella *et al.*, What's in a crowd? Analysis of face-to-face behavioral networks. *J. Theor. Biol.* **271**, 166–180 (2011).
31. D. C. Van Essen *et al.*; WU-Minn HCP Consortium, The WU-Minn Human Connectome Project: An overview. *Neuroimage* **80**, 62–79 (2013).
32. P. Tewarie *et al.*, How do spatially distinct frequency specific MEG networks emerge from one underlying structural connectome? The role of the structural eigenmodes. *Neuroimage* **186**, 211–220 (2019).
33. J. G. White, E. Southgate, J. N. Thomson, S. Brenner, The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **314**, 1–340 (1986).
34. B. L. Chen, D. H. Hall, D. B. Chklovskii, Wiring optimization can relate neuronal structure and function. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 4723–4728 (2006).
35. S. Sundaram, C. N. Hadjicostis, "Finite-time distributed consensus in graphs with time-invariant topologies," in *2007 American Control Conference* (Institute of Electrical and Electronics Engineers, 2007), pp. 711–716.
36. Y. Yuan, G.-B. Stan, L. Shi, M. Barahona, J. Goncalves, Decentralised minimum-time consensus. *Automatica* **49**, 1227–1235 (2013).
37. N. O'Clery, Y. Yuan, G.-B. Stan, M. Barahona, Global network prediction from local node dynamics. *arXiv [Preprint]* (2018). <https://doi.org/10.48550/arXiv:1809.00409> (Accessed 15 February 2022).
38. M. Newman, *Networks* (Oxford University Press, 2018).
39. A. C. Antoulas, *Approximation of Large-Scale Dynamical Systems* (Advances in Design and Control, Society for Industrial and Applied Mathematics, 2005), vol. 6.
40. G. Kerschen, J.-c. Golinval, A. F. Vakakis, L. A. Bergman, The method of proper orthogonal decomposition for dynamical characterization and order reduction of mechanical systems: An overview. *Nonlinear Dyn.* **41**, 147–169 (2005).
41. P. Benner, S. Gugercin, K. Willcox, A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM Rev.* **57**, 483–531 (2015).
42. S. L. Brunton, J. N. Kutz, *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control* (Cambridge University Press, 2019).
43. M. Morrison, J. N. Kutz, Nonlinear control of networked dynamical systems. *IEEE Trans. Netw. Sci. Eng.* **8**, 174–189 (2020).
44. L. Sirovich, Turbulence and the dynamics of coherent structures. I. Coherent structures. *Q. Appl. Math.* **45**, 561–571 (1987).
45. S. Boyd, L. Vandenberghe, *Introduction to Applied Linear Algebra: Vectors, Matrices, and Least Squares* (Cambridge University Press, 2018).
46. C. Bergmeir, J. M. Benítez, On the use of cross-validation for time series predictor evaluation. *Inf. Sci.* **191**, 192–213 (2012).
47. R. Tibshirani, Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. B* **58**, 267–288 (1996).
48. S. S. Chen, D. L. Donoho, M. A. Saunders, Atomic decomposition by basis pursuit. *SIAM Rev.* **43**, 129–159 (2001).
49. E. J. Candès, M. B. Wakin, An introduction to compressive sampling. *IEEE Signal Process. Mag.* **25**, 21–30 (2008).
50. T. Hastie, R. Tibshirani, M. Wainwright, *Statistical Learning with Sparsity: The Lasso and Generalizations* (CRC, 2015).
51. Z. Shen, W.-X. Wang, Y. Fan, Z. Di, Y.-C. Lai, Reconstructing propagation networks with natural diversity and identifying hidden sources. *Nat. Commun.* **5**, 4323 (2014).
52. B. Prasse, P. Van Mieghem, Network reconstruction and prediction of epidemic outbreaks for general group-based compartmental epidemic models. *IEEE Trans. Netw. Sci. Eng.* **7**, 2755–2764 (2020).
53. M. Nitzan, J. Casadiego, M. Timme, Revealing physical interaction networks from statistics of collective dynamics. *Sci. Adv.* **3**, e1600396 (2017).
54. S. H. Strogatz, *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering* (CRC, 2018).
55. A. Barrat, M. Barthélemy, R. Pastor-Satorras, A. Vespignani, The architecture of complex weighted networks. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 3747–3752 (2004).
56. T. Fawcett, An introduction to ROC analysis. *Pattern Recognit. Lett.* **27**, 861–874 (2006).
57. C. Presigny, P. Holme, A. Barrat, Building surrogate temporal network data from observed backbones. *Phys. Rev. E* **103**, 052304 (2021).
58. P. Holme, Epidemiologically optimal static networks from temporal network data. *PLOS Comput. Biol.* **9**, e1003142 (2013).
59. A. Machens *et al.*, An infectious disease model on empirical networks of human contact: Bridging the gap between dynamic network data and contact matrices. *BMC Infect. Dis.* **13**, 185 (2013).
60. M. Génio, C. L. Vestergaard, C. Cattuto, A. Barrat, Compensating for population sampling in simulations of epidemic spread on temporal contact networks. *Nat. Commun.* **6**, 8860 (2015).
61. M. Torricelli, M. Karsai, L. Gauvin, weg2vec: Event embedding for temporal networks. *Sci. Rep.* **10**, 7164 (2020).
62. M. Timme, Revealing network connectivity from response dynamics. *Phys. Rev. Lett.* **98**, 224101 (2007).
63. Y.-Y. Liu, J.-J. Slotine, A.-L. Barabási, Controllability of complex networks. *Nature* **473**, 167–173 (2011).
64. C. Nowzari, V. M. Preciado, G. J. Pappas, Analysis and control of epidemics: A survey of spreading processes on complex networks. *IEEE Contr. Syst. Mag.* **36**, 26–46 (2016).
65. S. L. Brunton, J. L. Proctor, J. N. Kutz, Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 3932–3937 (2016).
66. J. Casadiego, M. Nitzan, S. Hallerberg, M. Timme, Model-free inference of direct network interactions from nonlinear collective dynamics. *Nat. Commun.* **8**, 2192 (2017).
67. P. Van Mieghem, The N-Intertwined SIS epidemic network model. *Computing* **93**, 147–169 (2011).
68. A. J. Schwenk, "Computing the characteristic polynomial of a graph" in *Graphs and Combinatorics*, R. A. Bari, F. Harary, Eds. (Springer, 1974), pp. 153–172.
69. M. T. Schaub, L. Peel, Hierarchical community structure in networks. *arXiv [Preprint]* (2020). <https://doi.org/10.48550/arXiv:2009.07196> (Accessed 15 February 2022).
70. P. Van Mieghem, *Graph Spectra for Complex Networks* (Cambridge University Press, 2010).
71. N. O'Clery, Y. Yuan, G.-B. Stan, M. Barahona, Observability and coarse graining of consensus dynamics through the external equitable partition. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **88**, 042805 (2013).
72. S. Bonaccorsi, S. Ottaviano, D. Mugnolo, F. D. Pellegrini, Epidemic outbreaks in networks with equitable or almost-equitable partitions. *SIAM J. Appl. Math.* **75**, 2421–2443 (2015).
73. M. T. Schaub *et al.*, Graph partitions and cluster synchronization in networks of oscillators. *Chaos* **26**, 094821 (2016).
74. K. Devriendt, R. Lambiotte, Nonlinear network dynamics with consensus-dissensus bifurcation. *J. Nonlinear Sci.* **31**, 18 (2021).
75. B. Prasse, K. Devriendt, P. Van Mieghem, Clustering for epidemics on networks: A geometric approach. *Chaos* **31**, 063115 (2021).
76. P. Kundu, H. Kori, N. Masuda, Accuracy of a one-dimensional reduction of dynamical systems on networks. *Phys. Rev. E* **105**, 024305 (2022).
77. B. Prasse, P. Van Mieghem, Time-dependent solution of the NIMFA equations around the epidemic threshold. *J. Math. Biol.* **81**, 1299–1355 (2020).
78. W. L. Hamilton, R. Ying, J. Leskovec, "Inductive representation learning on large graphs" in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, U. von Luxburg, I. Guyon, Eds. (Association for Computing Machinery, 2017), pp. 1025–1035.
79. Y. LeCun, Y. Bengio, G. Hinton, Deep learning. *Nature* **521**, 436–444 (2015).
80. B. Prasse, P. Van Mieghem, Predicting-network-dynamics-without-the-graph. GitHub. <https://github.com/bprasse/Predicting-network-dynamics-without-the-graph>. Deposited 1 October 2022.

SI Appendix to “Predicting network dynamics without requiring the knowledge of the interaction graph” by Bastian Prasse and Piet Van Mieghem

A Details on the empirical networks and model parameters

Here, we provide details on the empirical networks and the parameters for the respective network dynamics in Section 2. For every network topology, we obtain the link weights a_{ij} as follows. If there is a link from node j to node i , then we set the element a_{ij} to a uniformly distributed random number in $[0.5, 1.5]$. If there is no link from node j to node i , then we set the respective element to $a_{ij} = 0$.

A.1 Lotka-Volterra population dynamics

For the competitive population dynamics described by the Lotka-Volterra equations, we consider the *Little Rock Lake* network [1], which we accessed via the *Konect* network collection [2]. The asymmetric and connected network consists of $N = 183$ nodes, which correspond to different species. There are $L = 2494$ directed links which specify the predation of one species upon another.

For every species i , we set the growth parameters α_i and θ_i to uniformly distributed random numbers in $[0.5, 1.5]$. Furthermore, we set the the initial nodal state $x_i(0)$ to a uniformly distributed random number in $[0, 1]$ for every species i . We set the maximum prediction time to $T_{\max} = 5$.

A.2 Mutualistic population dynamics

Kato *et al.* [3] studied the relationship between 679 insect species and 91 plants in a beech forest in Kyoto by specifying which insects pollinate or disperse which plant. We accessed the insect-plant network via the supplementary data in [4]. The insect-plant network determines a mutualistic insect-insect network [5]: If two insect species i and j pollinate or disperse the same plant, then both insect species i and j contribute to, and benefit from, the abundance of the plant. Thus, if two insect species i and j are linked to the same plant, then we set a_{ij} to a uniformly distributed random number in $[0.5, 1.5]$, and $a_{ij} = 0$ otherwise. As a result, we obtain a symmetric and disconnected network with $N = 679$ nodes and $L = 30,905$ links.

For every species i , we set the growth parameters α_i and θ_i to a uniformly distributed random number in $[0.5, 1.5]$. Furthermore, we set the the initial nodal state $x_i(0)$ to a uniformly distributed random number in $[0, 20]$ for every species i . We set the maximum prediction time to $T_{\max} = 0.025$.

A.3 Michaelis-Menten regulatory dynamics

We consider the transcription interactions between regulatory genes in the yeast *S. Cerevisiae* [6]. The asymmetric and disconnected network has $N = 620$ and $L = 869$ links. The influence from gene j to gene i is in either an activation or inhibition regulation. Since the activator interactions account for more than 80% of the links between genes, we only consider activation interactions, see also [7]. In line with Harush and Barzel [5], we consider degree avert regulatory dynamics by setting the Hill

coefficient to $h = 2$. We set the the initial nodal state $x_i(0)$ to a uniformly distributed random number in $[0, 2]$ for every node i . We set the maximum prediction time to $T_{\max} = 3$.

A.4 Susceptible-Infected-Susceptible (SIS) epidemics

The SIS contagion dynamics are evaluated on the contact network of the *Infectious: Stay Away* exhibition [8] between $N = 410$ individuals, accessed via [2]. The connected and symmetric network has $L = 5530$ links. A link between two nodes i, j indicates that the respective two individuals had a face-to-face contact that lasted for at least 20 seconds.

A crucial quantity for the SIS dynamics is the basic reproduction number R_0 , which is defined as [9]

$$R_0 = \rho \left(\text{diag}(\delta_1, \dots, \delta_N)^{-1} B \right). \quad (10)$$

Here, the spectral radius of an $N \times N$ matrix M is denoted by $\rho(M)$, and $\text{diag}(\delta_1, \dots, \delta_N)$ denotes the $N \times N$ diagonal matrix with the curing rates $\delta_1, \dots, \delta_N$ on its diagonal. If the basic reproduction number R_0 is less than or equal to 1, then the epidemic dies out [10], i.e., $x(t) \rightarrow 0$ as $t \rightarrow \infty$. We would like to study the spread of a virus that does not die out, and we aim to set the basic reproduction number to $R_0 = 1.5$: First, we set the “initial curing rate” $\delta_i^{(0)}$ to a uniformly distributed random number in $[0.5, 1.5]$ for every node i . Then, we set the curing rates to $\delta_i = c\delta_i^{(0)}$, where the multiplicity c is chosen such that the basic reproduction number in (10) equals $R_0 = 1.5$. We set the the initial nodal state $x_i(0)$ to a uniformly distributed random number in $[0, 0.1]$ for every node i . Furthermore, we set the maximum prediction time to $T_{\max} = 0.5$.

A.5 Kuramoto oscillators

We consider Kuramoto oscillator dynamics on the structural human brain network [11] of size $N = 78$. Every node corresponds to a brain region of the automated anatomical labelling (AAL) atlas [12]. The structural brain network specifies the anatomical connectivity between regions, i.e., the physical connections between regions based on white matter tracts. White matter tracts were estimated using fibre tracking from diffusion MRI data from the *Human Connectome Project* [13] as outlined in [14]. The network is symmetric and has $L = 696$ links.

For every node i , we set the natural frequency ω_i to a normally distributed random number with zero mean and standard deviation 0.1π . Furthermore, we set the the initial nodal state $x_i(0)$ to a uniformly distributed random number in $[-\pi/4, \pi/4]$ for every node i . We set the maximum prediction time to $T_{\max} = 1$.

A.6 Wilson-Cowan neural firing

We consider the modified Wilson-Cowan neural firing model of Laurence *et al.* [15] on the neuronal connectivity of the adult *Caenorhabditis elegans* hermaphrodite worm. Originally, White *et al.* [16] compiled the neuronal connectivity of *C. elegans*. In [17, 18], the neural wiring was updated, which we accessed online via the *Wormatlas* online database⁸. The somatic nervous system has $N = 282$

⁸Under the link: <http://www.wormatlas.org/neuronalwiring.html>

neurons and $L = 2994$ synapses. A link from node j to node i indicates the presence of at least one synapse from neuron j to neuron i .

The slope and the threshold of the neural activation functions are set to $\tau = 1$ and $\mu = 1$, respectively. The initial state $x_i(0)$ of every node i is set to a uniformly distributed random number in $[0, 10]$. We set the maximum prediction time to $T_{\max} = 4$.

B Details on the accuracy of the proper orthogonal decomposition

In the following, we provide more details on the accuracy on the POD (2).

B.1 Accuracy versus the number of agitation modes

Here, we evaluate the trade-off between accuracy and the number m of agitation modes y_1, \dots, y_m . The number of agitation modes m is equal to the dimension of the POD. We consider the nodal state $x(t)$ over the time period $0 \leq t \leq T_{\max}$. We consider $n = 1000$ equidistant observations, and we obtain the agitation modes y_p from the nodal state matrix $X = (x(0), x(\Delta t), \dots, x((n-1)\Delta t))$ as described in Section 3. In contrast to Section 3, we set the last observation time point to $(n-1)\Delta t = T_{\max}$, since we aim to evaluate the POD, and hence the dimensionality of the nodal state $x(t)$, over a given time interval $0 \leq t \leq T_{\max}$.

We define the contribution of agitation mode y_p as the Euclidean norm $\|c_p(t)y_p\|_2$ relative to the Euclidean norm of the nodal state $\|x(t)\|_2$, integrated over the time $t \in [0, T_{\max}]$, as

$$\Phi_{c_p} = \int_0^{T_{\max}} \|c_p(t)y_p\|_2 dt \left(\int_0^{T_{\max}} \|x(t)\|_2 dt \right)^{-1}. \quad (11)$$

Since $c_p(t) = y_p^T x(t)$ and $\|y_p\|_2 = 1$, it follows that

$$\Phi_{c_p} = \int_0^{T_{\max}} \|y_p^T x(t)\|_2 dt \left(\int_0^{T_{\max}} \|x(t)\|_2 dt \right)^{-1}.$$

Figure S1 shows that the contribution Φ_{c_p} decreases exponentially fast as the agitation mode index p increases. At some value of the index p , the contribution Φ_{c_p} plateaus at a very low value. More precisely, $\Phi_{c_p} \leq 10^{-15}$ as the index p is sufficiently large. Hence, Figure S1 demonstrates that the contributions Φ_{c_p} of the first few agitation modes y_p dominate the POD (2).

Furthermore, we are interested in the accuracy of the POD (2) versus the number of agitation modes m . Analogously to the contribution Φ_{c_p} of a single agitation mode c_p in (11), we define the approximation error of the POD with respect to the Euclidean norm over a finite time interval $t \in [0, T_{\max}]$ as

$$\Phi_{\text{POD},m} = \int_0^{T_{\max}} \left\| x(t) - \sum_{p=1}^m c_p(t)y_p \right\|_2 dt \left(\int_0^{T_{\max}} \|x(t)\|_2 dt \right)^{-1}.$$

Hence, the approximation error $\Phi_{\text{POD},m}$ measures the deviation of the nodal state $x(t)$ to the linear combination of m agitation modes y_1, \dots, y_m , whereas the contribution Φ_{c_p} above focussed only on a single agitation mode y_p .

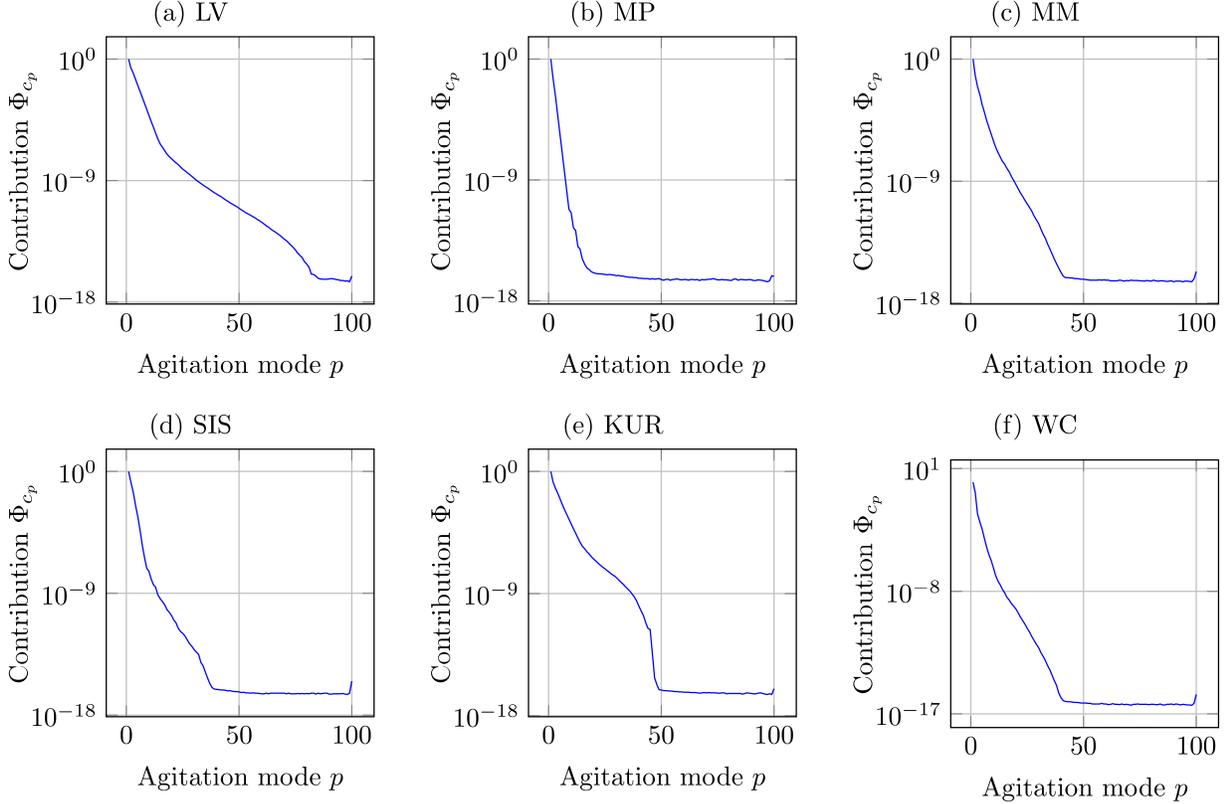


Figure S1: **The contribution of different agitation modes in the proper orthogonal decomposition.** The contribution Φ_{c_p} of the agitation mode y_p in the POD versus the agitation mode index $p = 1, \dots, N$ on a semi-logarithmic plot. The network size is $N = 100$, and the results are averaged over 100 Barabási-Albert random graphs.

Figure S2 shows that the approximation error $\Phi_{\text{POD},m}$ decreases exponentially as the number of agitation modes m increases. At some number of agitation modes m , the $\Phi_{\text{POD},m}$ plateaus at a very low value.

B.2 The number of agitation modes for large networks

We focus here on the accuracy of the POD (2) not with respect to the number of agitation modes m , as in Appendix B.1, but with respect to the number of nodes N in the underlying graph. For every network size N , we determine the number m of agitation modes, such that the POD (2) is virtually exact – up to machine precision. We determine the number of agitation modes m to represent the nodal state observations $x[1], \dots, x[n]$, where $n = \Delta t$, as the numerical rank of the $N \times n$ matrix $(x[1], \dots, x[n])$, which we compute with the Matlab command `rank`.

We set the parameters as described in the beginning of Appendix G, with exception of the maximum time T_{max} . Since we are interested in the number of agitation modes m for the nodal state $x(t)$ in a large time interval, we set the time T_{max} to ten times the value given in Appendix G. To keep the sampling time Δt at the same small value as in Appendix G, we replace (29) by $\Delta t = T_{\text{max}}/2000$. We range the number of nodes from $N = 100$ to $N = 1000$ and consider six different value of the

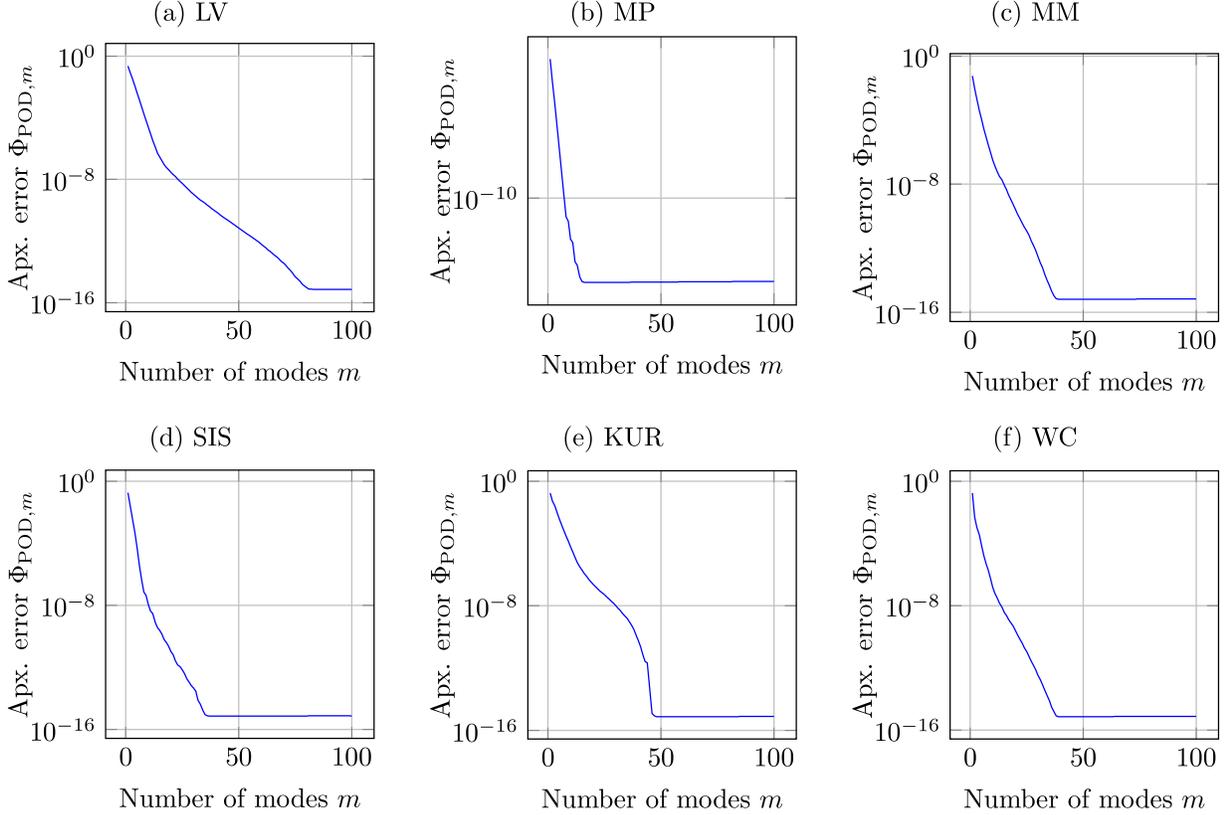


Figure S2: **The accuracy of the proper orthogonal decomposition versus the number of agitation modes.** The approximation error $\Phi_{\text{POD},m}$ of the POD with agitation modes y_1, \dots, y_m versus the number of agitation modes $m = 1, \dots, N$ on a semi-logarithmic plot. The network size is $N = 100$, and the results are averaged over 100 Barabási-Albert random graphs.

observation time $t_{\text{obs}} = 100\Delta t, 200\Delta t, 400\Delta t, 600\Delta t, 800\Delta t, 1000\Delta t$. (Hence, the smallest and largest observation times equal $t_{\text{obs}} = T_{\text{max}}/10$ and $t_{\text{obs}} = T_{\text{max}}$, respectively.)

Figure S3 suggests that, for any time interval $t \in [0, t_{\text{obs}}]$ with fixed time T_{max} , the number of agitation modes m is constant as $N \rightarrow \infty$. Hence, the dynamics remains at a given level of simplicity (a constant number of agitation modes m) despite a more and more complex network (due to a larger number of nodes N and a more heterogeneous degree distribution as N grows). The contrast between the simplicity of the dynamics and the network complexity enables accurate predictions of dynamics without precise knowledge of the network topology.

C Low-dimensional dynamics imply multiple possible surrogate matrices

From Example 1 for SIS dynamics, we found that if the low-dimensional POD (2) is exact for few agitation modes y_p , then many surrogates $\hat{A} \neq A$ predict the dynamics on the true matrix A . Here, we extend the claim to general dynamics (1): *if the POD (2) is accurate, then there many surrogates $\hat{A} \neq A$ that yield accurate predictions.*

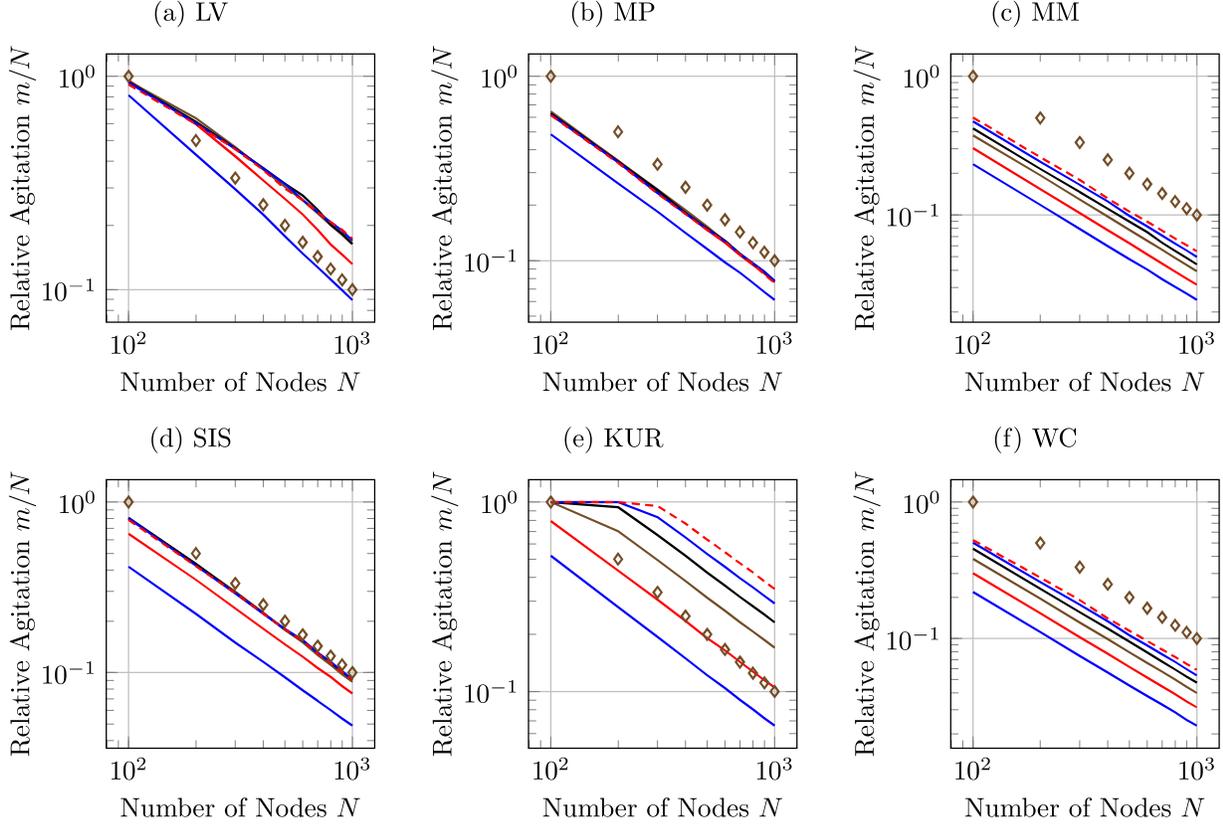


Figure S3: **The number of agitation modes versus the network size.** The number of agitation modes m , normalised by the network size N , versus the number of nodes N on a loglog-scale. The six different lines correspond to different values of the observation time $t_{\text{obs}} = 100\Delta t, 200\Delta t, 400\Delta t, 600\Delta t, 800\Delta t, 1000\Delta t$, and the diamond marks show the a constant number of agitation modes $m = 100$, i.e., $100/N$ agitation modes relative to the number of nodes N . For each network size N , the results are averaged over 100 Barabási-Albert random graphs.

A surrogate matrix \hat{A} predicts the dynamics on the true matrix A if (5) holds with a high accuracy. Thus, we aim to show that if the POD (2) is accurate for $m \ll N$, then the set of linear equations (5) has many solutions for the surrogate \hat{A} . First, we rewrite (5) as

$$\left(\hat{A}r_i(x(t))\right)_i = (Ar_i(x(t)))_i, \quad (12)$$

where we define the $N \times 1$ vector $r_i(x(t))$, for every node i , as

$$r_i(x(t)) = (g(x_i(t), x_1(t)), \dots, g(x_i(t), x_N(t)))^T. \quad (13)$$

Second, we aim to show that the vector $r_i(x(t))$ satisfies the POD

$$r_i(x(t)) \approx \sum_{p=1}^{\tilde{m}} \tilde{c}_{p,i}(t) \tilde{y}_{p,i} \quad (14)$$

with $\tilde{m} \ll N$ agitation modes $\tilde{y}_{p,i}$. In particular, we argue that if the POD (2) of the nodal state $x(t)$ is accurate for $m \ll N$, then the POD (14) of the vector $r_i(x(t))$ is accurate for $\tilde{m} \ll N$.

For the SIS dynamics in Example 1, the vector $r_i(t)$ simplifies to $r_i(x(t)) = x_i(t)x(t)$. Hence, for SIS dynamics, the POD (2) of the nodal state $x(t)$ yields a POD (14) of the vector $r_i(x(t))$ with the same agitation modes $\tilde{y}_{p,i} = y_p$ by identifying the scalar functions as $\tilde{c}_{p,i}(t) = x_i(t)c_p(t)$. In contrast, for the general dynamics (1), the relationship between the POD of the nodal state $x(t)$ and the vector $r_i(x(t))$ is more complicated.

To shorten the notation, we drop the time index t in the following: we formally replace the nodal state $x(t)$ by x and the functions $g(x_i(t), x_j(t))$ and $r_i(x(t))$ by $g(x_i, x_j)$ and $r_i(x)$, respectively.

To analyse the nonlinear dependency of the vector $r_i(x)$ on the nodal state x , we resort to a Taylor expansion. The function $r_i : \mathbb{R}^N \rightarrow \mathbb{R}^N$ is specified by the nonlinear interaction function g of the dynamical model (1). The Taylor expansion of the function $g(x_i, x_j)$ around the point $x_i = x_j = 0$ reads

$$g(x_i, x_j) = g(0, 0) + \sum_{k=1}^{\infty} \sum_{\alpha+\beta=k} \frac{1}{\alpha!\beta!} x_i^\alpha x_j^\beta \left. \frac{\partial^k g(x_i, x_j)}{\partial x_i^\alpha \partial x_j^\beta} \right|_{x_i=x_j=0}. \quad (15)$$

We denote the Taylor coefficient by

$$\eta(\alpha, \beta) = \frac{1}{\alpha!\beta!} \left. \frac{\partial^k g(x_i, x_j)}{\partial x_i^\alpha \partial x_j^\beta} \right|_{x_i=x_j=0}. \quad (16)$$

The indices i and j refer to the first and second argument of the function $g(x_i, x_j)$. Thus, the coefficient $\eta(\alpha, \beta)$ does not depend on the value of the indices i, j . With (16), we rewrite (15) as

$$g(x_i, x_j) = g(0, 0) + \sum_{k=1}^{\infty} \sum_{\alpha=0}^k \eta(\alpha, k-\alpha) x_i^\alpha x_j^{k-\alpha}. \quad (17)$$

With (17), we obtain the Taylor series of the function $r_i(x)$ in (13) as

$$r_i(x) = r_i(0) + \sum_{k=1}^{\infty} \sum_{\alpha=0}^k \eta(\alpha, k-\alpha) x_i^\alpha x^{k-\alpha}, \quad (18)$$

where we denote the element-wise power of the vector x as

$$x^{k-\alpha} = \left(x_1^{k-\alpha}, \dots, x_N^{k-\alpha} \right)^T.$$

We define the truncation of the series (18) until the q -th power as

$$r_{q,i}(x) = r_i(0) + \sum_{k=1}^q \sum_{\alpha=0}^k \eta(\alpha, k-\alpha) x_i^\alpha x^{k-\alpha}. \quad (19)$$

For a sufficiently large power q , it holds that $r_i(x(t)) \approx r_{q,i}(x(t))$ at every time $t \in [0, T_{\max}]$. In fact, if the interaction function $g(x_i, x_j)$ is a polynomial of degree q , then the truncation $r_{q,i}(x)$ is exact, i.e., $r_i(x) = r_{q,i}(x)$. For instance, the SIS epidemic process, whose interaction function equals $g(x_i, x_j) = (1 - x_i)x_j$, is characterised by $r_i(x) = r_{2,i}(x)$.

Proposition 2. *Suppose that the $N \times 1$ nodal state vector $x(t)$ equals the linear combination of m vectors y_p at every time t ,*

$$x(t) = \sum_{p=1}^m c_p(t) y_p \quad (20)$$

for some scalars $c_p(t) \in \mathbb{R}$. Then, the function $r_{q,i}(x(t))$ in (19) satisfies

$$r_{q,i}(x(t)) = \sum_{p=1}^{\tilde{m}} \tilde{c}_{p,i}(t) \tilde{y}_{p,i} \quad (21)$$

where the number \tilde{m} of vectors $\tilde{y}_{p,i} \in \mathbb{R}^N$ equals

$$\tilde{m} = \sum_{\beta=0}^q \binom{\beta + m - 1}{m - 1}.$$

Proof. It follows from (20) that $x(t) \in \text{span}\{y_1, \dots, y_m\}$ at every time t . We rewrite the function $r_{q,i}(x)$ in (19) as

$$r_{q,i}(x) = r_i(0) + \sum_{k=1}^q \sum_{\beta=0}^k \eta(k - \beta, \beta) x_i^{k-\beta} x^\beta.$$

Both terms $\eta(k - \beta, \beta)$ and $x_i^{k-\beta}$ are scalars. Thus, we obtain that

$$r_{q,i}(x) = r_i(0) + \sum_{\beta=0}^q \mu_\beta(x) x^\beta \quad (22)$$

for some scalars $\mu_0(x), \mu_1(x), \dots, \mu_q(x) \in \mathbb{R}$. We consider the addends x^β in (22) separately. Due to (20), for any vector $x \in \text{span}\{y_1, \dots, y_m\}$, the j -th component of x^β equals

$$x_j^\beta = \left(\sum_{p=1}^m c_p(y_p)_j \right)^\beta$$

for some scalars c_1, \dots, c_m . The multinomial theorem yields that

$$x_j^\beta = \sum_{p_1+p_2+\dots+p_m=\beta} \frac{\beta!}{p_1!p_2!\dots p_m!} \prod_{l=1}^m (c_l(y_l)_j)^{p_l}.$$

Thus, by stacking the components $x_1^\beta, \dots, x_N^\beta$, we can express the vector x^β as

$$\begin{aligned} x^\beta &= \sum_{p_1+p_2+\dots+p_m=\beta} \frac{\beta!}{p_1!p_2!\dots p_m!} \prod_{l=1}^m \begin{pmatrix} (c_l(y_l)_1)^{p_l} \\ \vdots \\ (c_l(y_l)_N)^{p_l} \end{pmatrix} \\ &= \sum_{p_1+p_2+\dots+p_m=\beta} \zeta_\beta(p_1, \dots, p_m) \nu_\beta(p_1, \dots, p_m), \end{aligned} \quad (23)$$

where we define the coefficients

$$\zeta_\beta(p_1, \dots, p_m) = \frac{\beta!}{p_1!p_2!\dots p_m!}$$

and the $N \times 1$ vectors

$$\nu_\beta(p_1, \dots, p_m) = \prod_{l=1}^m \begin{pmatrix} (c_l(y_l)_1)^{p_l} \\ \vdots \\ (c_l(y_l)_N)^{p_l} \end{pmatrix}.$$

Relation (23) shows that the vector x^β is a linear combination of all vectors $\nu_\beta(p_1, \dots, p_m)$ with $p_1 + p_2 + \dots + p_m = \beta$, of which there are

$$\binom{\beta + m - 1}{m - 1}.$$

The vector $r_{q,i}(x)$ in (22) is a linear combination of the vectors x^β for $\beta = 0, \dots, q$. Thus, for any $x(t) \in \text{span}\{y_1, \dots, y_m\}$, we can write the vector $r_{q,i}(x(t))$ as a linear combination of

$$\tilde{m} = \sum_{\beta=0}^q \binom{\beta + m - 1}{m - 1}.$$

vectors $\nu_\beta(p_1, \dots, p_m)$ with $p_1 + p_2 + \dots + p_m = \beta$, where $\beta = 0, \dots, q$. Lastly, we obtain the vectors $\tilde{y}_{p,i}$, where $p = 1, \dots, \tilde{m}$, as an orthonormal basis vectors of the subspace spanned by the \tilde{m} vectors $\nu_\beta(p_1, \dots, p_m)$. Furthermore, the scalar functions in (21) follow as $\tilde{c}_{p,i}(t) = \tilde{y}_{p,i}^T r_{q,i}(x)$, which completes the proof. \square

Proposition 2 shows that, if the nodal state $x(t)$ is agitated⁹ with only $m \ll N$ agitation modes y_p , then also the vector $r_{q,i}(x(t))$ is agitated with only $\tilde{m} \ll N$ agitation modes $\tilde{y}_{p,i}$. Suppose that the power q is sufficiently great, such that $r_i(x(t)) \approx r_{q,i}(x(t))$. Then, (12) yields that the surrogate \hat{A} predicts accurately if, for all nodes i ,

$$\left(\hat{A} \sum_{p=1}^{\tilde{m}} \tilde{c}_{p,i}(t) \tilde{y}_{p,i} \right)_i = \left(A \sum_{p=1}^{\tilde{m}} \tilde{c}_{p,i}(t) \tilde{y}_{p,i} \right)_i.$$

Equivalently, we obtain that

$$\hat{A} \sum_{p=1}^{\tilde{m}} \tilde{c}_{p,i}(t) \tilde{y}_{p,i} = A \sum_{p=1}^{\tilde{m}} \tilde{c}_{p,i}(t) \tilde{y}_{p,i},$$

which is satisfied if

$$\hat{A} \tilde{y}_{p,i} = A \tilde{y}_{p,i} \tag{24}$$

for every agitation mode $\tilde{y}_{p,i}$. The linear system (24) has N^2 unknowns \hat{a}_{ij} but only $\tilde{m}N \ll N^2$ equations. Hence, there are many surrogate matrices \hat{A} that solve (24) and provide accurate predictions of the nodal state $x(t)$.

D Approximate low-dimensional dynamics and multiple possible surrogate matrices

In Example 1 and in Proposition 2 we considered that the POD (20) is exact. However, the POD might only hold approximately for the dynamics (1), with a high approximation accuracy as demonstrated by Figure 1. In this section, we show that there are many surrogate matrices $\hat{A} \neq A$ that result in accurate predictions, also if the POD is only approximate.

⁹By (20) we considered that the POD is exact. In Appendix D, we show that there are many surrogate matrices \hat{A} that predict accurately, also if the POD only holds approximately.

First, we rewrite the set of linear equations (5), which is central to obtaining the surrogate matrix \hat{A} . For every node i , we define the $n \times N$ matrix F_i with the vector $r_i^T(x(t))$ in (13) as

$$F_i = \begin{pmatrix} r_i^T(x[0]) \\ \vdots \\ r_i^T(x[n-1]) \end{pmatrix}.$$

Then, the set of linear equations (5) of the observation times $t = 0, \Delta t, \dots, (n-1)\Delta t$ becomes

$$F_i \begin{pmatrix} \hat{a}_{i1} \\ \vdots \\ \hat{a}_{iN} \end{pmatrix} = F_i \begin{pmatrix} a_{i1} \\ \vdots \\ a_{iN} \end{pmatrix}. \quad (25)$$

Thus, we obtained a linear system (25) for every row $\hat{a}_{1i}, \dots, \hat{a}_{1N}$ of the surrogate \hat{A} . The rank of the matrix F_i is essential: If the matrix F_i is of full rank, i.e., $\text{rank}(F_i) = N$, then there is exactly one solution to (25), namely the entries a_{i1}, \dots, a_{iN} of the true adjacency matrix A . Otherwise, if $\text{rank}(F_i) < N$, then there are infinitely many solutions $\hat{A} \neq A$ to (25). If there is more than one solution to (25), then the LASSO estimation (9) results in the sparsest solution \hat{A} (with respect to the ℓ_1 -norm).

Every computer works with finite precision arithmetic. Thus, not the exact rank but the *numerical* rank of the matrix F_i is decisive to solve the system (25) in practice. The numerical rank equals the number of singular values of the matrix F_i that are greater than a small threshold, which is set in accordance to the machine precision. We compute the numerical rank for Barabási-Albert random graphs versus the network size N . Here, we consider the *best case* for inferring the network topology: The derivative $dx_i(t)/dt$ is observed exactly, without any approximation error as in (8).

Figure S4 shows that the numerical rank of the matrix F_i stagnates as the network size N grows. The linear system (25) is *severely ill-conditioned* for large networks. For example, for the SIS process on a network with $N = 1000$ nodes, we observe a 1000×1001 nodal state sequence $x[0], \dots, x[1000]$, but the numerical rank does not exceed 32. Thus, there are countless (infinitely many) surrogates $\hat{A} \neq A$ that result in accurate predictions of the dynamics.

E Details of the surrogate network inference algorithm

Here, we give details to the prediction method, which is a development of the algorithm we proposed in [19] for discrete-time epidemic models. The solution $\hat{a}_{i1}(\rho_i), \dots, \hat{a}_{iN}(\rho_i)$ to the LASSO (9) depends on the regularisation parameter ρ_i . We aim to choose the parameter ρ_i that results in the solution $\hat{a}_{i1}(\rho_i), \dots, \hat{a}_{iN}(\rho_i)$ with the greatest prediction accuracy. To assess the prediction accuracy, we apply *hold-out cross-validation* [20]: We divide the nodal state observations into a training set $x[0], \dots, x[n_{\text{train}}]$ and a validation set $x[n_{\text{train}} + 1], \dots, x[n]$. The training set is used to obtain the solution $\hat{a}_{i1}(\rho_i), \dots, \hat{a}_{iN}(\rho_i)$ in dependency of ρ_i , whose prediction accuracy is evaluated on the validation set. We choose the regularisation parameter ρ_i with the greatest prediction accuracy on the validation set.

More precisely, we define the training set as the first 80% of the nodal state observations $x[0], \dots, x[n_{\text{train}}]$, where $n_{\text{train}} = \lceil 0.8n \rceil$. By $\hat{a}_{i1}(\rho_i), \dots, \hat{a}_{iN}(\rho_i)$, we denote the solution to the LASSO *on the*

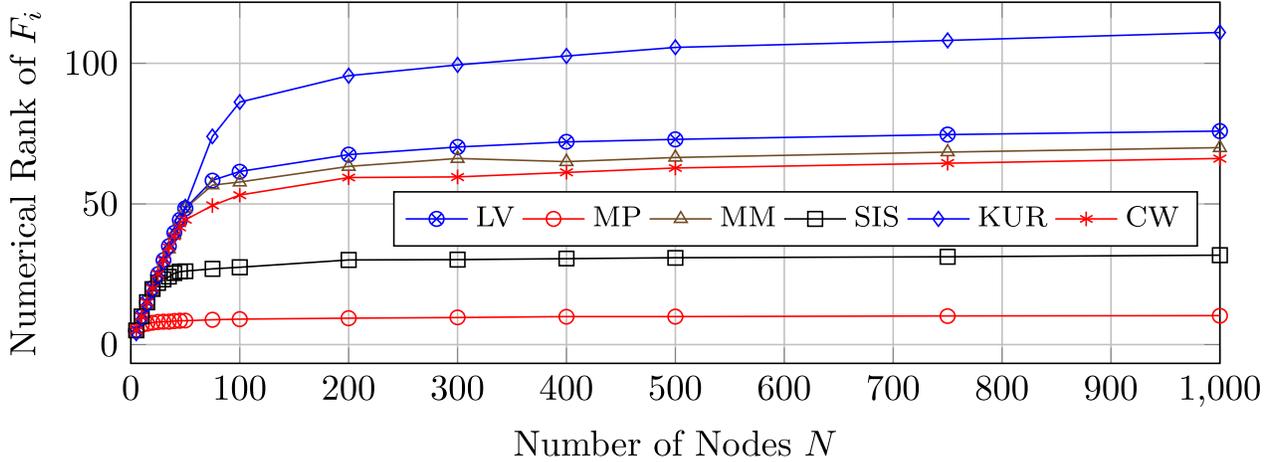


Figure S4: **Ill-condition of inferring the network topology.** The numerical rank of the matrix F_i of the linear system (25) versus the number of nodes N for Barabási-Albert random graphs. The observation time is set to $t_{\text{obs}} = T_{\text{max}}$, and the number of observations equals $n = 1000$.

training set $x[0], \dots, x[n_{\text{train}}]$, i.e., the solution to

$$\begin{aligned} \underset{\hat{a}_{i1}, \dots, \hat{a}_{iN}}{\text{argmin}} \quad & \sum_{k=0}^{n_{\text{train}}} \left(\frac{x_i[k+1] - x_i[k]}{\Delta t} - f_i(x_i[k]) - \sum_{j=1}^N \hat{a}_{ij} g(x_i[k], x_j[k]) \right)^2 + \rho_i \sum_{j=1}^N \hat{a}_{ij} \\ \text{s.t.} \quad & \hat{a}_{ij} \geq 0 \quad j = 1, \dots, N. \end{aligned} \quad (26)$$

If an entry $\hat{a}_{ij}(\rho_i)$ of the LASSO solution is smaller than the threshold 0.01, then we round off and set $\hat{a}_{ij}(\rho_i) = 0$. For every value of the regularisation parameter ρ_i , we define the prediction error $\text{MSE}(\rho_i)$ on the validation set $x[n_{\text{train}} + 1], \dots, x[n]$ as

$$\text{MSE}(\rho_i) = \sum_{k=n_{\text{train}}+1}^{n-1} \left(\frac{x_i[k+1] - x_i[k]}{\Delta t} - f_i(x_i[k]) - \sum_{j=1}^N \hat{a}_{ij} g(x_i[k], x_j[k]) \right)^2. \quad (27)$$

We iterate over a set Θ_i , specified below, of predefined candidate values for ρ_i . Every candidate value $\rho_i \in \Theta_i$ results in a different prediction error $\text{MSE}(\rho_i)$. We determine the final regularisation parameter $\rho_{\text{opt},i}$ as the candidate value $\rho_{\text{opt},i} \in \Theta_i$ with the minimal prediction error $\text{MSE}(\rho_{\text{opt},i})$. We obtain the final estimate $\hat{a}_{i1}, \dots, \hat{a}_{iN}$ as the solution to the LASSO (9) with the regularisation parameter $\rho_{\text{opt},i}$ on the whole data set $x[0], \dots, x[n_{\text{train}}]$.

We define the set Θ_i as 20 logarithmically equidistant candidate values as $\Theta_i = \{\rho_{\min,i}, \dots, \rho_{\max,i}\}$. If $\rho_i > \rho_{\text{th},i}$, where $\rho_{\text{th},i} = 2\|F_i^T V_i\|_{\infty}$, then, as demonstrated by Kim *et al.* [21], the solution to the LASSO (9) equals $\hat{a}_{ij} = 0$ for all nodes j . Here, the entries of the $(n_{\text{train}} + 1) \times 1$ vector V_i are given by $(V_i)_k = (x_i[k+1] - x_i[k])/\Delta t - f_i(x_i[k])$. Thus, we set the candidate values in the set Θ_i proportional to $\rho_{\text{th},i}$. We define

$$\rho_{\max,i} = 2 \cdot 10^{-2} \left\| F_i^T \begin{pmatrix} \frac{x_i[1] - x_i[0]}{\Delta t} - f_i(x_i[0]) \\ \vdots \\ \frac{x_i[n] - x_i[n-1]}{\Delta t} - f_i(x_i[n-1]) \end{pmatrix} \right\|_{\infty} \quad (28)$$

and $\rho_{\min,i} = 10^{-4}\rho_{\max,i}$. The surrogate inference method is given by Algorithm 1.

Algorithm 1 Obtaining the surrogate matrix

```

1: Input: nodal state time series  $x[0], x[1], \dots, x[n]$ 
2: Output: surrogate matrix  $\hat{A}$  with elements  $\hat{a}_{ij}$ 
3: for  $i = 1, \dots, N$  do
4:   Set  $\rho_{\max,i}$  to (28)
5:    $\rho_{\min,i} \leftarrow 10^{-4}\rho_{\max,i}$ 
6:    $\Theta_i \leftarrow 20$  logarithmically equidistant values from  $\rho_{\min,i}$  to  $\rho_{\max,i}$ 
7:   for  $\rho_i \in \Theta_i$  do
8:      $\hat{a}_{i1}(\rho_i), \dots, \hat{a}_{iN}(\rho_i) \leftarrow$  solution to LASSO (26) on the training set
9:      $\hat{a}_{ij}(\rho_i) \leftarrow 0$  for all  $\hat{a}_{i1}(\rho_i), \dots, \hat{a}_{iN}(\rho_i)$  smaller than 0.01
10:    Compute  $\text{MSE}(\rho_i)$  by (27) on the validation set
11:   end for
12:    $\rho_{\text{opt},i} \leftarrow \underset{\rho_i \in \Theta_i}{\text{argmin}} \text{MSE}(\rho_i)$ 
13:    $(\hat{a}_{i1}, \dots, \hat{a}_{iN}) \leftarrow$  the solution to LASSO (9) for  $\rho_i = \rho_{\text{opt},i}$  on the whole data set
14:    $\hat{a}_{ij} \leftarrow 0$  for all  $\hat{a}_{i1}, \dots, \hat{a}_{iN}$  smaller than 0.01
15: end for

```

We would like to mention two things on the way we incorporated rounding into obtaining the surrogate network. First, the *exact* solution to ℓ_1 -norm minimisation problems, such as the constrained LASSO (26), often contains $\hat{a}_{ij} = 0$ for some entries \hat{a}_{ij} of the optimisation variable [22, 23, 24]. We solve the LASSO (26) numerically, by the Matlab command `quadprog`, which may give $\hat{a}_{ij} \approx 0$ as output even though the exact solution would be $\hat{a}_{ij} = 0$. Second, we emphasise that the hold-out cross-validation in Algorithm 1 does take into account that we set $\hat{a}_{ij} = 0$ for all entries \hat{a}_{ij} that are below the threshold $\hat{a}_{ij} < 0.01$. The mean square error in line 10 of Algorithm 1 is computed after the rounding in line 9. Thus, we expect that the final output of Algorithm 1, computed in line 13, corresponds to the minimum of the mean square error that can be obtained by first solving the LASSO (26) numerically and then rounding the output, which results in either $\hat{a}_{ij} \geq 0.01$ or $\hat{a}_{ij} = 0$.

F Prediction accuracy and topological similarity for random graphs

In Sections 4 and 5, we evaluated the prediction accuracy and reconstruction accuracy for real-world networks. Here, we provide an additional evaluation for the Barabási-Albert and Erdős-Rényi random graphs with $N = 300$ nodes.

Figure S5 shows that the predictions are accurate for Barabási-Albert random graphs. However, just as for the real-world networks in Section 5, the topology of the surrogate network bears little similarity with the true network topology, since the AUC for the networks in Figure S5 equals: **(a)** AUC=0.52, **(b)** AUC=0.51, **(c)** AUC=0.53, **(d)** AUC=0.51, **(e)** AUC=0.55, **(f)** AUC=0.54.

For Erdős-Rényi, we obtain similar results, the prediction accuracy is demonstrated by Figure S6. All of the networks corresponding to Figure S6 are connected. The topological similarity of the surrogate network and the true network topology is small, with AUC values given by: **(a)** AUC=0.51,

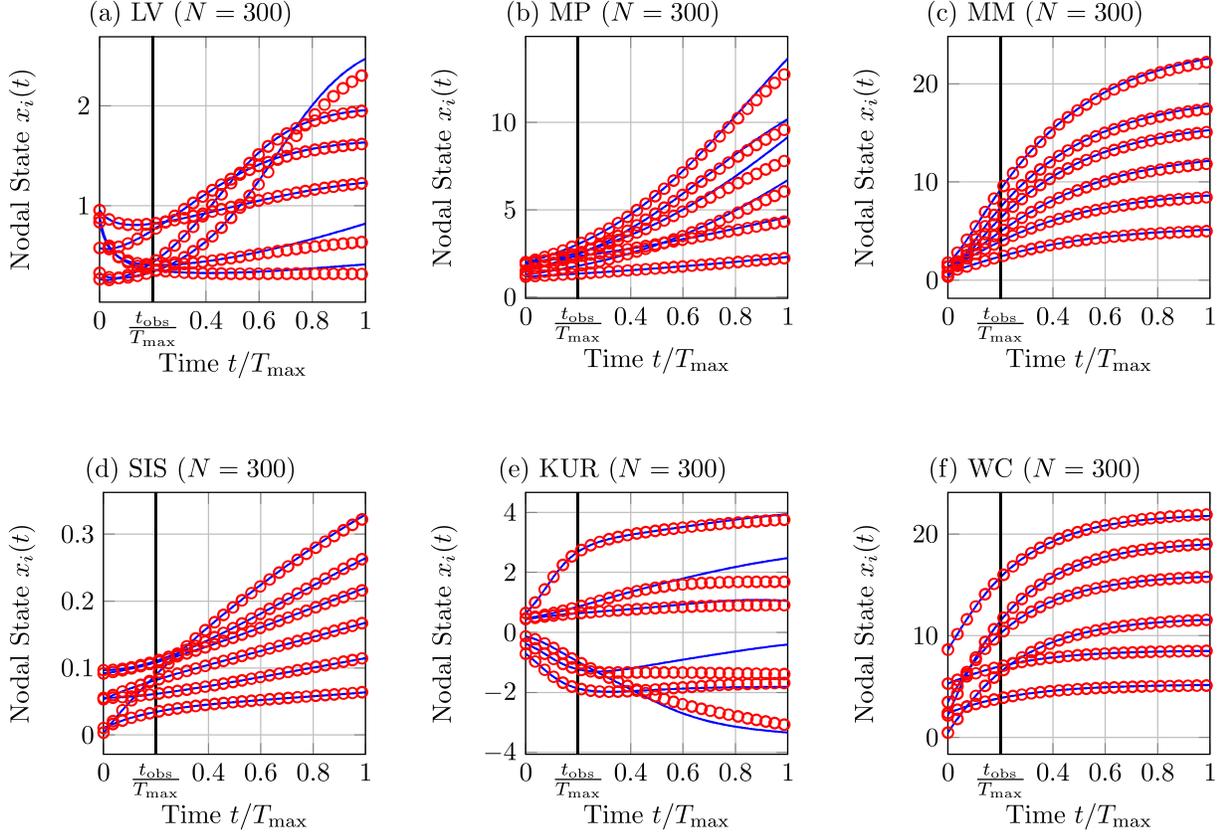


Figure S5: **Accuracy of the prediction method for Barabási-Albert random graphs.** Based on $n = 100$ nodal state observations $x(0), x(\Delta t), \dots, x(n\Delta t)$ until time $t_{\text{obs}} = n\Delta t$, the nodal state $x(t)$ is predicted at times $t > t_{\text{obs}}$. The true adjacency matrices A with $N = 300$ nodes are generated by the Barabási-Albert random graph model. The blue curves are the true nodal states $x_i(t)$. The red marks are the nodal state predictions $\hat{x}_i(t)$ based on the surrogate matrix \hat{A} , initialised as $\hat{x}(t_{\text{obs}}) = x(t_{\text{obs}})$. For clarity, only six nodal states $x_i(t)$ are depicted for each network. The average of the prediction error $\epsilon_i(t) = |x_i(t) - \hat{x}_i(t)|$, with respect to all nodes i and future times $t \in [t_{\text{obs}}, T_{\text{max}}]$, is denoted by $\bar{\epsilon}$ and equals: (a) $\bar{\epsilon} = 1.41 \cdot 10^{-2}$, (b) $\bar{\epsilon} = 6.79 \cdot 10^{-3}$, (c) $\bar{\epsilon} = 5.77 \cdot 10^{-2}$, (d) $\bar{\epsilon} = 4.28 \cdot 10^{-4}$, (e) $\bar{\epsilon} = 1.37 \cdot 10^{-1}$, (f) $\bar{\epsilon} = 3.42 \cdot 10^{-2}$.

(b) AUC=0.51, (c) AUC=0.51, (d) AUC=0.5, (e) AUC=0.51, (f) AUC=0.52.

F.1 The spectrum of the surrogate and true networks

In Section 5, we used the in-degree distribution and the AUC value as metrics to compare the surrogate and the true network topology. Here, we consider the eigenvalue spectra [25] of the surrogate matrix \hat{A} and the true matrix A , to provide an additional comparison between the two topologies. The spectrum of the matrices A is given by the set $\{\lambda_1, \lambda_2, \dots, \lambda_N\}$ of (not necessarily distinct) eigenvalues λ_i of A , where the eigenvalues are sorted as $0 \leq |\lambda_1| \leq \dots \leq |\lambda_N|$ and $0 \leq |\hat{\lambda}_1| \leq \dots \leq |\hat{\lambda}_N|$. Similarly, we denote the eigenvalues of the surrogate matrix \hat{A} by $\hat{\lambda}_1, \dots, \hat{\lambda}_N$.

For Barabási-Albert random graphs, Figure S7 demonstrates two points. First, the spectrum of

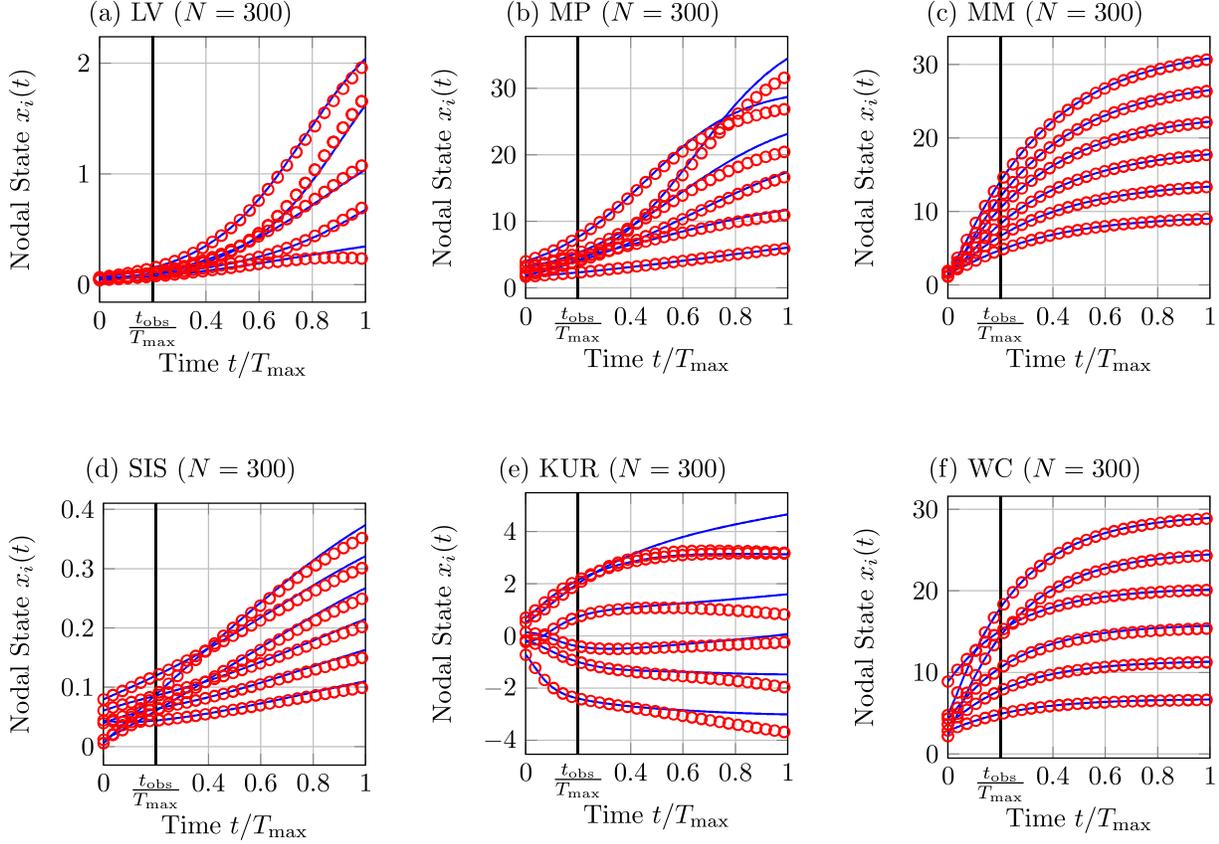


Figure S6: **Accuracy of the prediction method for Erdős-Rényi random graphs.** Based on $n = 100$ nodal state observations $x(0), x(\Delta t), \dots, x(n\Delta t)$ until time $t_{\text{obs}} = n\Delta t$, the nodal state $x(t)$ is predicted at times $t > t_{\text{obs}}$. The true adjacency matrices A with $N = 300$ nodes are generated by the Erdős-Rényi random graph model with link probability $p_{\text{ER}} = 0.05$. The blue curves are the true nodal states $x_i(t)$. The red marks are the nodal state predictions $\hat{x}_i(t)$ based on the surrogate matrix \hat{A} , initialised as $\hat{x}(t_{\text{obs}}) = x(t_{\text{obs}})$. For clarity, only six nodal states $x_i(t)$ are depicted for each network. The average of the prediction error $\epsilon_i(t) = |x_i(t) - \hat{x}_i(t)|$, with respect to all nodes i and future times $t \in [t_{\text{obs}}, T_{\text{max}}]$, is denoted by $\bar{\epsilon}$ and equals: **(a)** $\bar{\epsilon} = 1.44 \cdot 10^{-2}$, **(b)** $\bar{\epsilon} = 3.39 \cdot 10^{-1}$, **(c)** $\bar{\epsilon} = 7.23 \cdot 10^{-2}$, **(d)** $\bar{\epsilon} = 4.07 \cdot 10^{-3}$, **(e)** $\bar{\epsilon} = 1.91 \cdot 10^{-1}$, **(f)** $\bar{\epsilon} = 9.35 \cdot 10^{-2}$.

the spectra of the surrogate matrix \hat{A} is very different to the spectrum of the true matrix A . Second, almost all eigenvalues $\hat{\lambda}_i$ of the surrogate matrix \hat{A} are (close to) zero. Hence, it holds that $\hat{A}v \approx 0$ for all vectors $v \in \mathcal{V}$ in a subspace $\mathcal{V} \subset \mathbb{R}^N$. Here, the subspace \mathcal{V} equals the span of the eigenvectors of \hat{A} which correspond to eigenvalues that are (close to) zero. (If the eigenvalues are exactly 0, then the subspace \mathcal{V} equals the kernel of the matrix \hat{A} .) Figure S7 illustrates the large dimension $\dim\{\mathcal{V}\} \gg 1$ of the subspace \mathcal{V} . The large dimension of the subspace \mathcal{V} relates to the small dimension of the subspace \mathcal{X} of the nodal state $x(t)$ in Section 3.

Figure S8 demonstrates the same two observations for Erdős-Rényi graphs as for Barabási-Albert graphs Figure S7.

Lastly, one could argue that the comparison of the surrogate network \hat{A} and the true network A

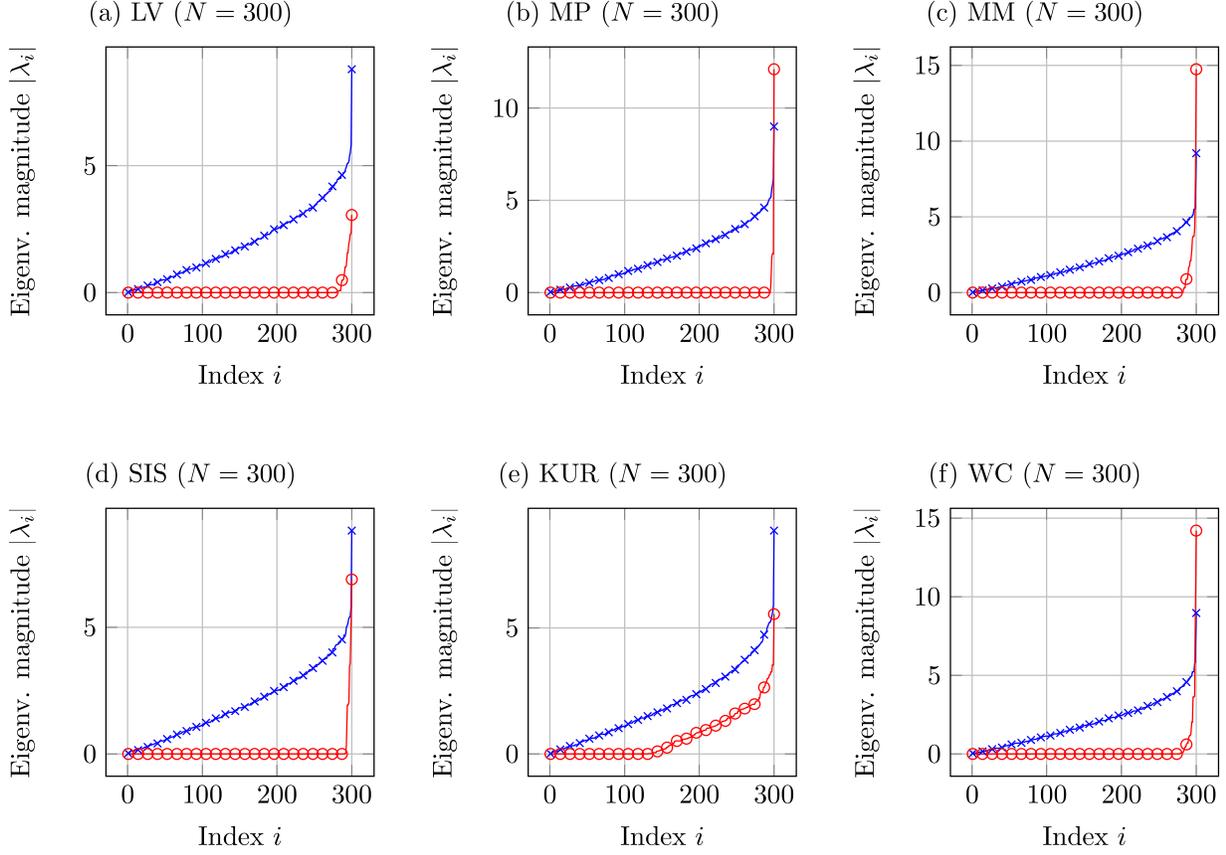


Figure S7: **The spectra of the surrogate and the true network for Barabási-Albert random graphs.** A comparison of the spectrum of the surrogate matrix \hat{A} in red and the spectrum of the true matrix A in blue. The networks are the same used for generating the dynamics in Figure S5, i.e., the true matrix A was generated by the Barabási-Albert random graph model.

should consider rotational symmetries. For instance, consider $N = 2$ nodes and that the true matrix is given by $a_{12} = a_{21} = 0$, $a_{11} = a_{22} = 1$ and the surrogate matrix is given by $\hat{a}_{11} = \hat{a}_{22} = 0$, $\hat{a}_{12} = \hat{a}_{21} = 1$. Then, the matrices A and \hat{A} would be the same, up to a rotation of $\pi/4$, but we would have an AUC=0.5. However, we emphasise that a rotation of the matrix A does have an impact on the dynamics (1). A potential rotational symmetry of the matrices A and \hat{A} does not imply that the coupling terms $\sum_{j=1}^N a_{ij}g(x_i(t), x_j(t))$ and $\sum_{j=1}^N \hat{a}_{ij}g(x_i(t), x_j(t))$ in (1) are the same, due to the non-linearity of the coupling function g and an nodal state vector $x(t)$ with generally unequal entries $x_i(t) \neq x_j(t)$. However, while comparing the eigenvalues of the surrogate matrix \hat{A} and the true matrix A in Figures S7 and S8, we found, in line with the in-degree and AUC topological comparison in Figure 4, that the eigenvalue spectra of the matrices \hat{A} and A are very different. If the surrogate matrix \hat{A} and the true matrix A were rotationally symmetric, then their spectra would be the same.

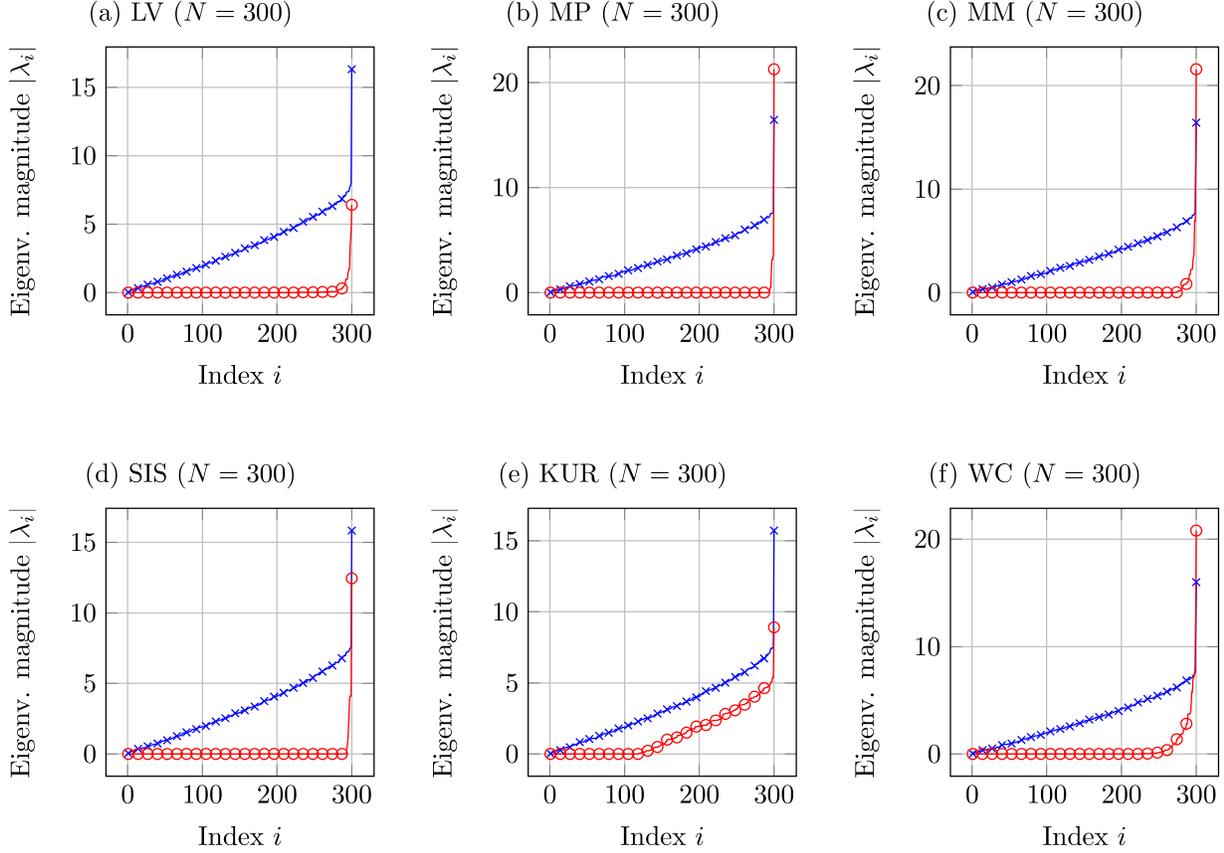


Figure S8: **The spectra of the surrogate and the true network for Erdős-Rényi random graphs.** A comparison of the spectrum of the surrogate matrix \hat{A} in red and the spectrum of the true matrix A in blue. The networks are the same used for generating the dynamics in Figure S6, i.e., the true matrix A was generated by the Erdős-Rényi random graph model.

G Sensitivity analysis of the prediction method

In this section, we perform an extensive sensitivity analysis of the prediction method. We evaluate the performance on randomly generated networks and the dynamics described in Section 2. Unless explicitly stated otherwise, we use the same set of parameters from Subsections G.1 to G.4, which we detail in the following.

We consider Barabási-Albert random graphs with $N = 100$ nodes to generate an unweighted network. From the unweighted network, we generate a weighted network as follows: If there is a link between node i and j , then we set the link weight a_{ij} to a uniform random number in $[0.5, 1.5]$.

We set the initial nodal state $x_i(0)$ to a uniform random number in $[0, 1]$, except for the SIS and Kuramoto models for which the interval is set to $[0, 0.1]$ and $[-\pi/4, \pi/4]$, respectively. For the different models, the maximum prediction time is set to: (LV) $T_{\max} = 8$, (MP) $T_{\max} = 2$, (MM) $T_{\max} = 4$, (SIS) $T_{\max} = 2$, (KUR) $T_{\max} = 2$, and (WC) $T_{\max} = 4$. We choose the model-specific prediction time T_{\max} such that the dynamics are sufficiently rich, as illustrated by Figure S9. (If the prediction time T_{\max} is too small, then there can be too little change of the nodal state $x(t)$ over the time interval

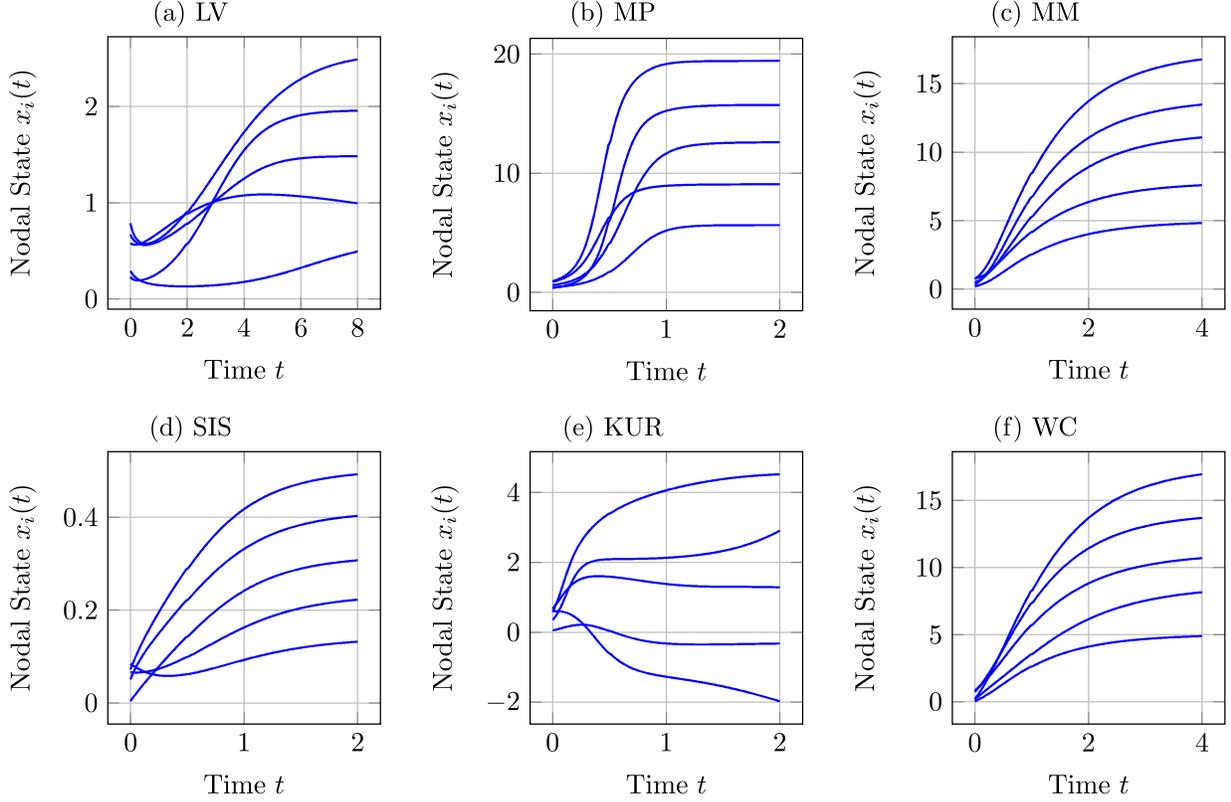


Figure S9: **The nodal state dynamics.** The blue curves show the nodal state $x_i(t)$ of the respective dynamic model on a Barabási-Albert random graph with $N = 100$ nodes. For readability, only five of the 100 nodal states $x_i(t)$ are depicted. The maximum prediction time T_{\max} is different for each dynamic model: **(a)** $T_{\max} = 8$, **(b)** $T_{\max} = 2$, **(c)** $T_{\max} = 4$, **(d)** $T_{\max} = 2$, **(e)** $T_{\max} = 2$, **(f)** $T_{\max} = 4$.

$t \in [0, T_{\max}]$. If the prediction time T_{\max} is too large, then the approximation (8) of the derivative $dx(t)/dt$ can be too inaccurate.) The sampling time is set to

$$\Delta t = T_{\max}/200. \quad (29)$$

The model parameters (e.g., the curing rates δ_i for the SIS model) are set in the same way as described in Appendix A.

G.1 Prediction time horizon

We evaluate the prediction accuracy with respect to the observation time t_{obs} and prediction time horizon $\tilde{t} = t - t_{\text{obs}}$. We define the prediction error $\epsilon(\tilde{t})$ at the prediction time horizon $\tilde{t} = t - t_{\text{obs}}$ as the average

$$\epsilon(\tilde{t}) = \frac{1}{N} \sum_{i=1}^N |x_i(\tilde{t}) - \hat{x}_i(\tilde{t})|. \quad (30)$$

The time step Δt is set as in (29), and we vary the observation time from $t_{\text{obs}} = 0.05T_{\max}$ to $t_{\text{obs}} = 3T_{\max}$. The number of nodal state observations follows as $n = t_{\text{obs}}/\Delta t$, which we round up if the result is not a natural number.

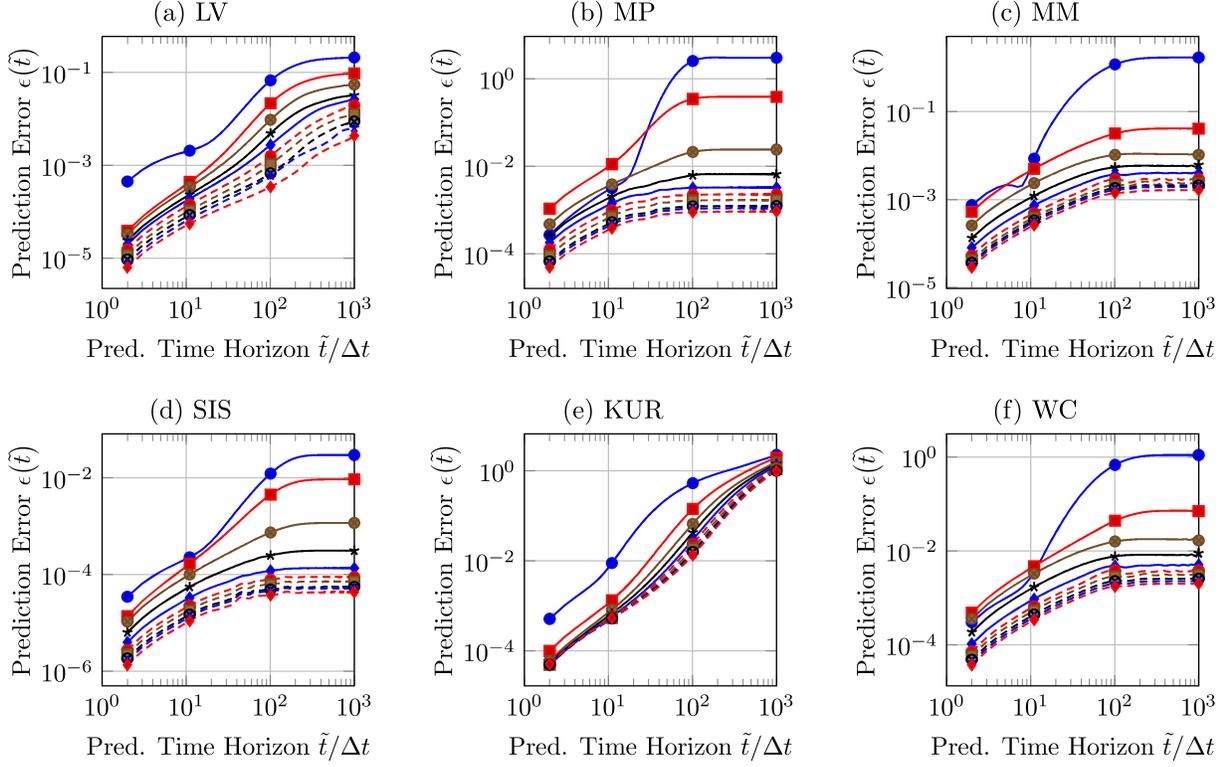


Figure S10: **The prediction accuracy versus the prediction time horizon.** The prediction error $\epsilon(\tilde{t})$ versus prediction time horizon $\tilde{t} = t - t_{\text{obs}}$, normalised by the time step Δt , on a loglog-scale. Each of the ten curves corresponds to another value of the observation time t_{obs} in an equally-spaced range from $t_{\text{obs}} = 0.05T_{\text{max}}$ to $t_{\text{obs}} = 3T_{\text{max}}$. The results are averaged over 100 Barabási-Albert random graphs with $N = 100$ nodes.

Figure S10 shows that, for a broad range observation times t_{obs} , the prediction error $\epsilon(\tilde{t})$ increases gradually as the prediction time horizon \tilde{t} increases. More specifically, the straight lines in the log-log-scale of Figure S10 suggest that the increase of prediction error $\epsilon(\tilde{t})$ seems polynomial with respect to the prediction time horizon \tilde{t} , provided that the prediction time \tilde{t} is sufficiently small. Except for the Kuramoto model, the error $\epsilon(\tilde{t})$ plateaus at a large value as the prediction time horizon \tilde{t} becomes large. Except for the Kuramoto model, the nodal state $x_i(t)$ takes only values in a limited range (compare to Figure S9), which explains that there is a plateau for the prediction error $\epsilon(\tilde{t})$.

G.2 The dependency of the prediction error on the network size

We are interested in the impact of the network size N on the prediction error $\epsilon(\tilde{t})$. We consider $n = 100$ nodal state observations with the time step Δt given in (29). (Hence, the observation time satisfies $t_{\text{obs}} = T_{\text{max}}/2$.) We range the network size from $N = 25$ to $N = 500$. We obtain the network in two steps. First, for each number of nodes N , we generate a Barabási-Albert random as described in Appendix G to obtain a weighted adjacency matrix A .

Second, we rescale the adjacency matrix A such that 2-norm of the matrix A remains comparable for different network sizes N . Without rescaling of the adjacency matrix A , the largest eigenvalue λ_1

grows for large network sizes N . Thus, the time scales of the dynamics would not be comparable, which would not be in agreement with our focus on a fixed time interval $[0, T_{\max}]$, which is independent of the number of nodes N .

To rescale the matrix A , we obtain the average largest eigenvalue λ_{target} of the networks with $N = 100$. We do not rescale networks with $N = 100$ nodes. For networks with $N \neq 100$ nodes, we rescale the adjacency matrix A with largest eigenvalue λ_1 as $A \leftarrow \lambda_{\text{target}}/\lambda_1 sA$. Here, we choose to include a small scalar perturbation s , which is generated as a uniform random number in $[0.9, 1.1]$.

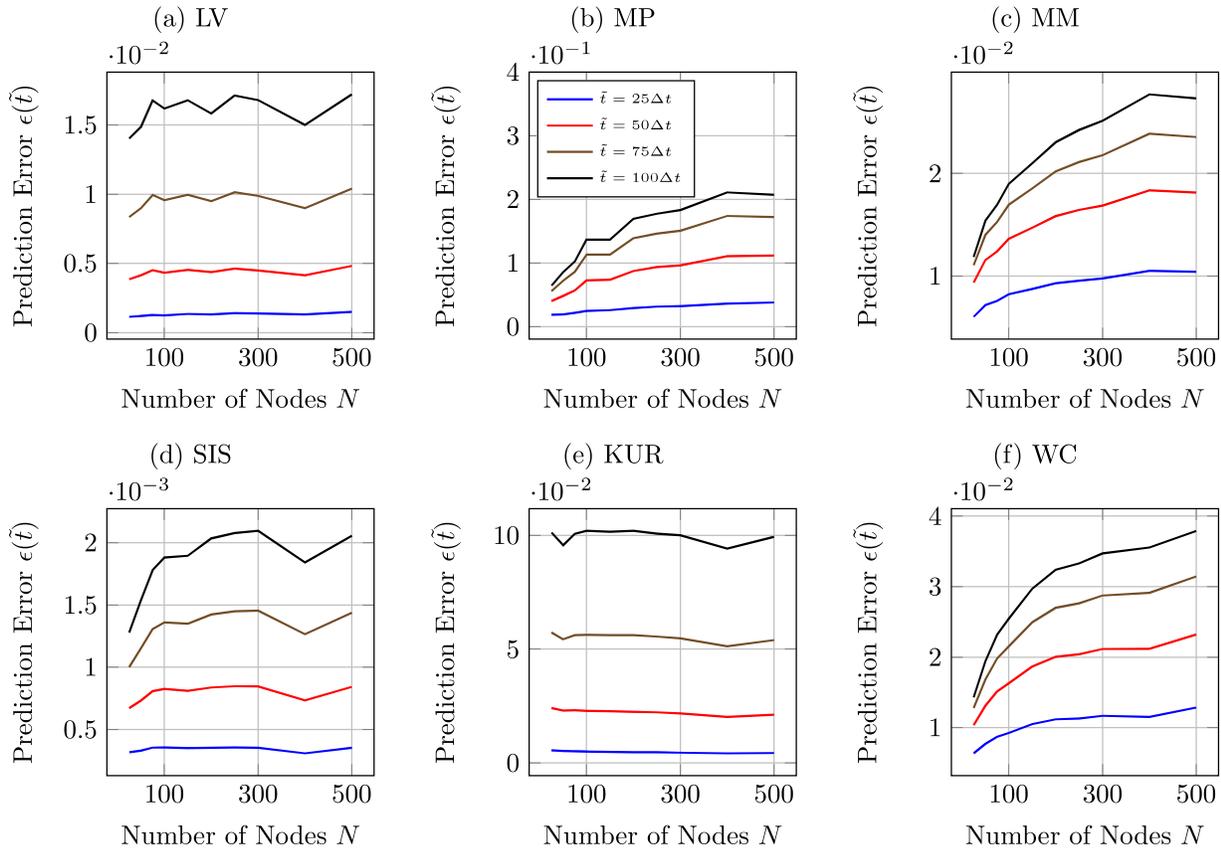


Figure S11: **The prediction error versus the network size.** The prediction error $\epsilon(\tilde{t})$ versus the network size N . Each of the four curves corresponds to another value of the prediction time horizon $\tilde{t} = t - t_{\text{obs}}$. The results are averaged over 100 and 20 Barabási-Albert random graphs for $N < 400$ and $N \geq 400$ nodes, respectively.

Figure S11 shows that the prediction error $\epsilon(\tilde{t})$ increases for larger network sizes N for the MP, the MM and the WC dynamics. Interestingly, for the other models (LV, SIS and KUR), there does not seem to be a substantial increase of the prediction error $\epsilon(\tilde{t})$ for larger networks. As expected, the smaller the prediction horizon \tilde{t} , the smaller the prediction error $\epsilon(\tilde{t})$ for all models. Furthermore, the smaller the prediction horizon \tilde{t} , the smaller the effect of the network size N on the prediction error $\epsilon(\tilde{t})$.

G.3 The impact of errors on the prediction accuracy

For many applications, the nodal state $x[k]$ is not observed exactly but is subject to noise. We consider that the input of the surrogate network Algorithm 1 is subject to erroneous observations $x[k] + \varpi_i[k]x_{i,\max}$, where $x_{i,\max} = \max_{k=0,\dots,n} x_i[k]$ denotes the maximum nodal state of node i until time $k = n$. Here, the error $\varpi_i[k]$ is drawn from a Gaussian distribution with zero mean and variance σ^2 , independently and identically distributed for all nodes i and times k . The approximation (8) for the difference $(x[k+1] - x[k])/\Delta t$, which is at the core of obtaining the surrogate matrix \hat{A} , is affected by the error $\varpi_i[k]$.

We consider $n = 100$ nodal state observations with the time step Δt given in (29). We explore the impact of different error levels by varying the standard deviation from $\sigma = 10^{-4}$ to $\sigma = 10^{-2}$. We consider the prediction error $\bar{\epsilon}$, which is the average of the time-dependent prediction error $\epsilon(t)$ over all prediction times $t \in [t_{\text{obs}}, T_{\text{max}}]$, and we set the observation time to $t_{\text{obs}} = T_{\text{max}}/2$.

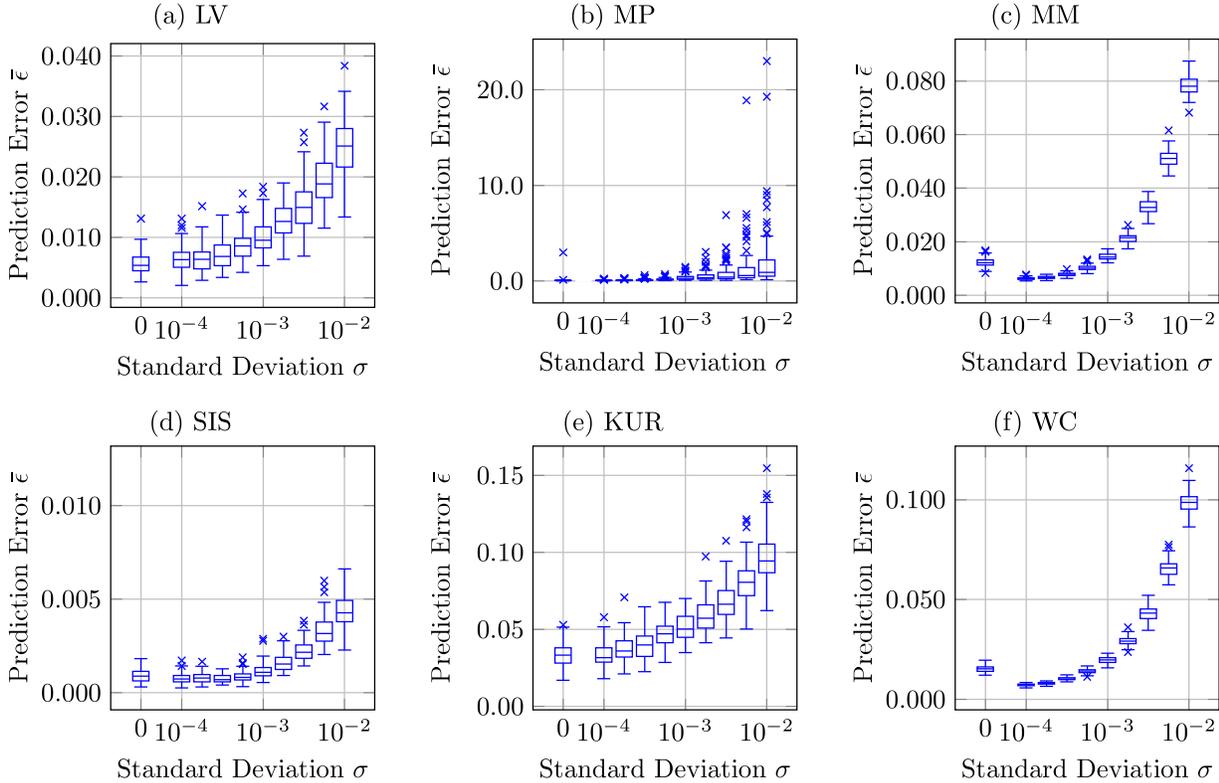


Figure S12: **The prediction error versus the random error level.** The prediction error $\bar{\epsilon}$ versus the standard deviation σ of the errors $\varpi_i[k]$ on a semilogarithmic scale. The leftmost boxplot corresponds to no noise ($\sigma = 0$). The boxplots capture the variation with respect to 100 Barabási-Albert random graphs with $N = 100$ nodes.

Figure S12 illustrates that the prediction method is robust in the presence of errors $\varpi_i[k]$ with a sufficiently small standard deviation σ . The sensitivity of the prediction method depends on the dynamic model and is particularly strong for the MM and WC models. Curiously, for the MM and WC models, and to a lesser extent for the SIS model, the prediction is more accurate in the presence of errors $\varpi_i[k]$ with a small standard deviation σ . Since the beneficial effect of small errors does not

hold for all models, we believe that the effect is due to the specific nonlinear dynamics rather than the prediction method.

G.4 Heterogeneous interaction functions

We study the prediction accuracy of our approach for interaction functions $g_i(x_i(t), x_j(t))$ that, in contrast to the original formulation (1), depend on the node i . We consider the self-interaction function

$$f_i(x_i(t)) = -x_i(t),$$

such that, if there were no interactions ($a_{ij} = 0$ for all nodes i, j), then the dynamics (1) would be stable. We describe the interaction function $g_i(x_i(t), x_j(t))$ as a third-degree polynomial

$$g_i(x_i(t), x_j(t)) = \frac{1}{N} (\theta_{1i} + \theta_{2i}x_i(t) + \theta_{3i}x_j(t) + \theta_{4i}x_i^2(t) + \theta_{5i}x_i(t)x_j(t) + \theta_{6i}x_j^2(t) + \theta_{7i}x_i^3(t) + \theta_{8i}x_i^2(t)x_j(t) + \theta_{9i}x_i(t)x_j^2(t) + \theta_{10i}x_j^3(t)), \quad (31)$$

which is specified by the coefficients θ_{si} , $s = 1, \dots, 10$. With (31), we can quantify the heterogeneity of the coupling functions g_i in terms of the coefficients θ_{si} . We denote the mean of the coefficients θ_{si} with respect to all nodes i as $\bar{\theta}_s$. For all $s = 1, \dots, 10$, the mean $\bar{\theta}_s$ is drawn from a standard normal distribution. Based on the mean $\bar{\theta}_s$, we generate the coefficients θ_{si} as

$$\theta_{si} = \bar{\theta}_s + \tilde{\theta}_{si},$$

where the term $\tilde{\theta}_{si}$ is drawn from a normal distribution with zero mean and variance σ_g^2 . We emphasise that, as throughout this work, we consider the coupling functions g_i in (31) to be fully known to the surrogate network Algorithm 1. The maximum prediction time is set to $T_{\max} = 2$. We consider $n = 50$ observations of the nodal state $x(t)$ until the observation time $t_{\text{obs}} = 0.5$, and we set the $x_i(0)$ to a uniform random number in $[0, 1]$ for every node i .

Figure S13 shows that the standard deviation σ_g that determines the coefficients θ_{si} has no clear impact on the prediction accuracy. Hence, the results suggest that the prediction accuracy is unaffected by the degree of heterogeneity of the coupling functions g_i .

H Connected components of the surrogate networks

In Algorithm 1, we obtain the surrogate network \hat{A} without imposing a connectivity constraint. Hence, the surrogate network \hat{A} may be disconnected. Here, we are interested in the connectivity of the surrogate matrix \hat{A} , more specifically the number of connected components. We use the same parameter setting as in Appendix G.1.

Figure S14 shows that the distribution of the connected components in the surrogate network \hat{A} substantially varies across different dynamics models, ranging from almost a point-distribution for the Kuramoto model to a widely-spread distribution for the LV model. On average, across all six models, a fraction of 35.82% of the networks are disconnected (i.e., have more than one connected component). Remarkably, despite the high prediction accuracy demonstrated for the same parameters in Figure S10, Figure S14 suggests that also disconnected surrogate networks \hat{A} can yield accurate predictions.

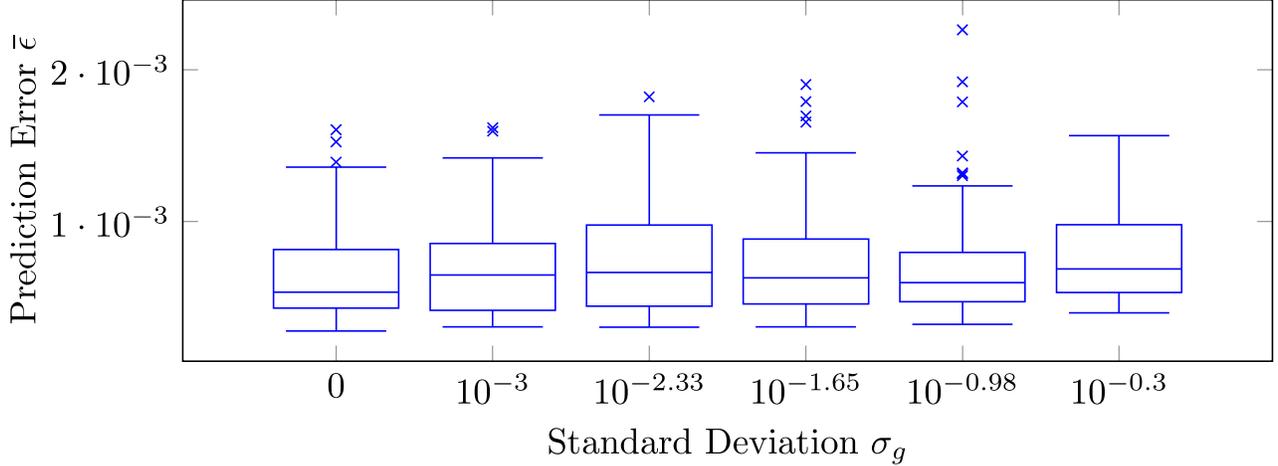


Figure S13: **The prediction accuracy versus the heterogeneity of the coupling functions.** The average $\bar{\epsilon}$ of the prediction error $\epsilon_i(t) = |x_i(t) - \hat{x}_i(t)|$, with respect to all nodes i and future times $t \in [t_{\text{obs}}, T_{\text{max}}]$, versus the standard deviation σ_g which determines the heterogeneity of the coupling functions g_i . The boxplots capture the variation with respect to 100 Barabási-Albert random graphs with $N = 100$ nodes. The leftmost boxplot ($\sigma_g = 0$) shows the error $\bar{\epsilon}$ for homogeneous coupling functions $g_i = g_j$, and the five boxplots to the right correspond to a standard deviation ranging from $\sigma_g = 10^{-3}$ to $\sigma_g = 0.5$, logarithmically equally spaced.

I Observing multiple trajectories on the same network

In the previous sections, we considered that only one trajectory of the nodal state $x(t)$ is observed, from some initial state $x(0)$ to the nodal state $x(t_{\text{obs}})$ at the observation time t_{obs} . In the following, we consider observing multiple nodal state trajectories on *the same* network, with the same and fixed adjacency matrix A , starting from different initial conditions $x(0)$. Intuitively, we expect that the more trajectories are observed, the more accurate our prediction approach is.

Suppose we observe S nodal state trajectories $x^{(s)}(t)$, where $s = 1, \dots, S$, on the same adjacency matrix A . For every trajectory $x^{(s)}(t)$, we consider $n = 50$ nodal state observations with the time step Δt given in (29). (Hence, the observation time satisfies $t_{\text{obs}} = T_{\text{max}}/4$.) Then, the LASSO (9) can be adjusted in a straightforward manner as

$$\begin{aligned}
 & \underset{\hat{a}_{i1}, \dots, \hat{a}_{iN}}{\operatorname{argmin}} \quad \sum_{s=1}^S \sum_{k=0}^{n-1} \left(\frac{x_i^{(s)}[k+1] - x_i^{(s)}[k]}{\Delta t} - f_i(x_i^{(s)}[k]) - \sum_{j=1}^N \hat{a}_{ij} g(x_i^{(s)}[k], x_j^{(s)}[k]) \right)^2 + \rho_i \sum_{j=1}^N \hat{a}_{ij} \\
 & \text{s.t.} \quad \hat{a}_{ij} \geq 0 \quad j = 1, \dots, N.
 \end{aligned} \tag{32}$$

Algorithm 1 is adjusted to multiple trajectories in an analogous manner.

We evaluate the impact of $S = 1, \dots, 20$ trajectories on the prediction accuracy and the similarity of the resulting surrogate network \hat{A} with the true network A . We generate weighted Barabási-Albert random graphs with $N = 100$ in the same way as described in the beginning of Appendix G. For every trajectory $x^{(s)}(t)$ and every node i , we set the initial nodal state $x_i^{(s)}(0)$ randomly, as detailed in the beginning of Appendix G.

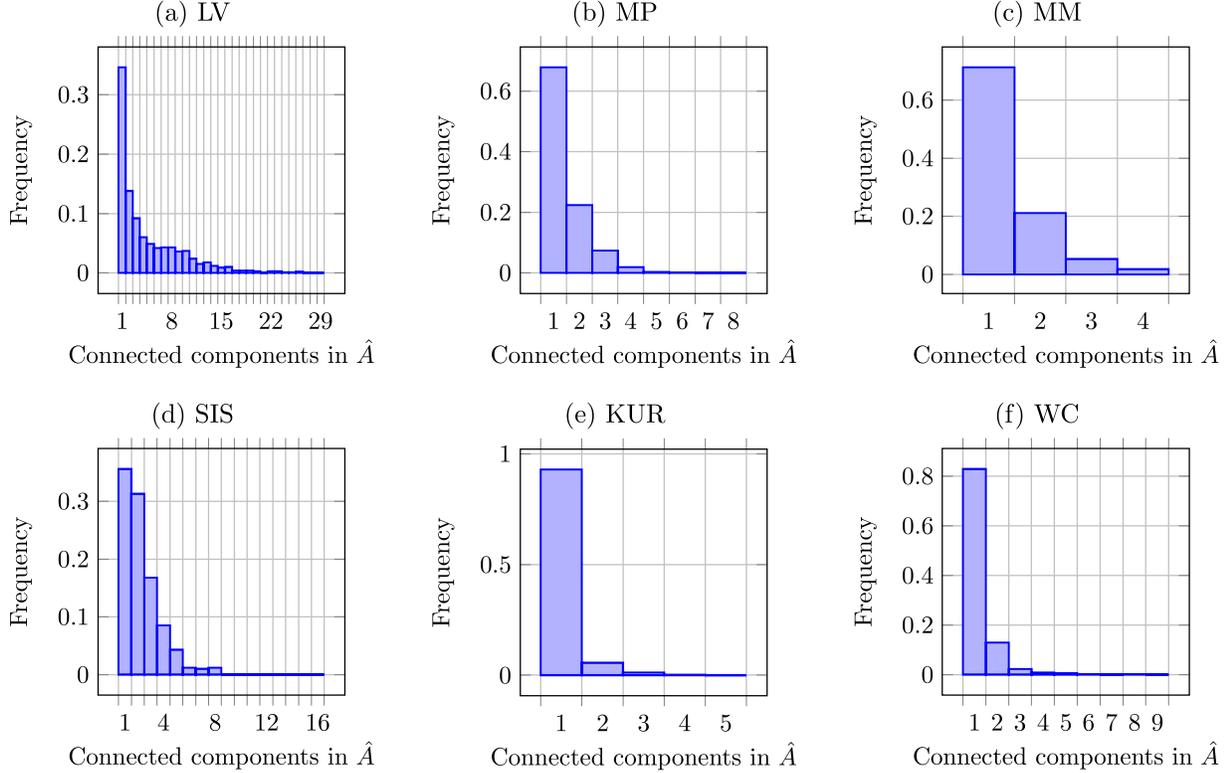


Figure S14: **The distribution of the number of connected components in the surrogate network.** The frequency of the number of connected components of the surrogate matrix \hat{A} . The results capture the distribution of the number of connected components with respect to a range of ten observation times t_{obs} and 100 Barabási-Albert random graphs with $N = 100$ nodes. The fraction of connected surrogate networks \hat{A} (i.e., with exactly one connected component) equals: **(a)** 34.6%, **(b)** 67.8%, **(c)** 71.3%, **(d)** 35.6%, **(e)** 92.9%, **(f)** 82.9%.

Figure S15 shows that the prediction accuracy for the trajectory $x^{(s)}(t)$, where $s = S$, generally decreases as the number S of observed trajectories increases. There are variations across the different dynamic models. The prediction error of the mutualistic population and Kuramoto model starts at the largest value for $S = 1$ observed trajectories and decreases more rapidly than for the other models. The prediction error for the SIS model shows a clear plateau for more trajectories, even with a slight increase from $S = 10$ to $S = 20$, which might be due to numerical inaccuracies when solving the adjusted LASSO (32) with many terms.

Figure S16 shows that the topology of the surrogate network \hat{A} does converge to the true topology, provided sufficiently many independent trajectories S . By comparing Figures S15 and S16, we observe that if more trajectories S are observed, then both the prediction accuracy and the topological similarity of the true network A and the surrogate \hat{A} increases.

In the sequel, we investigate the relation of the prediction error on the topological similarity more closely, across any number of observed trajectories $S = 1, \dots, 20$. Figure S17 shows the relation of the prediction error $\bar{\epsilon}$ and the topological similarity of the networks \hat{A} and A . (Figure S17 is based on the same data points as Figures S15 and S16.) As expected, a larger similarity between the networks \hat{A} and

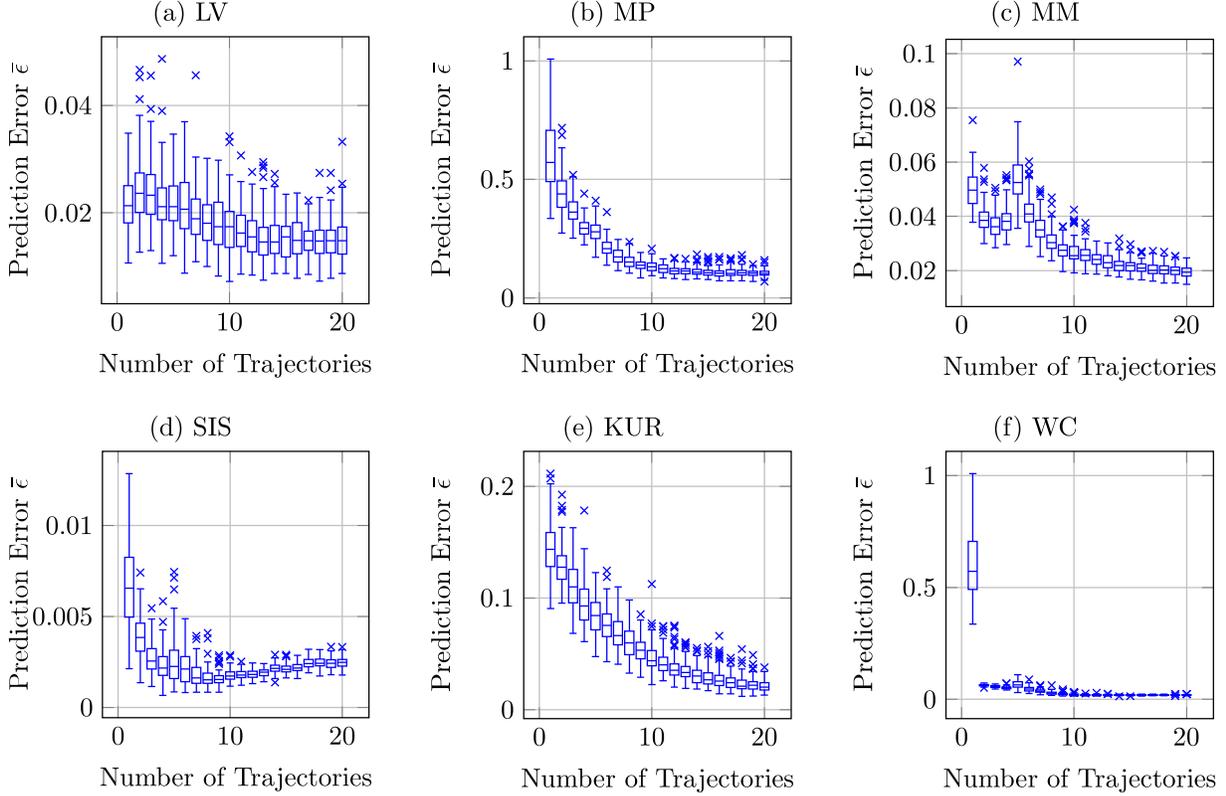


Figure S15: **The prediction accuracy from observing multiple trajectories on the same network.** The average $\bar{\epsilon}^{(s)}$ of the prediction error $\epsilon_i^{(s)}(t) = |x_i^{(s)}(t) - \hat{x}_i^{(s)}(t)|$, with respect to all nodes i and future times $t \in [t_{\text{obs}}, T_{\text{max}}]$, for trajectory s provided that $S = s$ trajectories $x^{(s)}(t)$ have been observed on the same network. The boxplots capture the variation with respect to 100 Barabási-Albert random graphs with $N = 100$ nodes.

A does result in a lower prediction error $\bar{\epsilon}$, across all models. We stress that, while the area under the curve AUC and the prediction error $\bar{\epsilon}$ are correlated, the crucial confounding variable is the number of observed trajectories. The prediction error $\bar{\epsilon}$ in Figure S17 decreases rather gradually over the broad AUC-range of the area under the curve, from no topological similarity at all (AUC = 0.5) to identical networks (AUC = 1) and demonstrates, for most dynamical models, the remarkable insensitivity of the underlying graph on the prediction. Intuitively, we had anticipated a stronger negative correlation between the prediction error and the topological accuracy AUC of the underlying graph as in e.g. (b) MP in Figure S17, but most dynamic models only show such negative correlation in an AUC-range near to 1.

J Computational complexity of the prediction method

We aim to assess the runtime of Algorithm 1 on the network size N . We use the same parameter setting as in Appendix G.2.

Figure S18 shows that the impact of the network size N on the runtime of the prediction method is consistent across all models. Furthermore, the runtime of Algorithm 1 closely follows a fitted

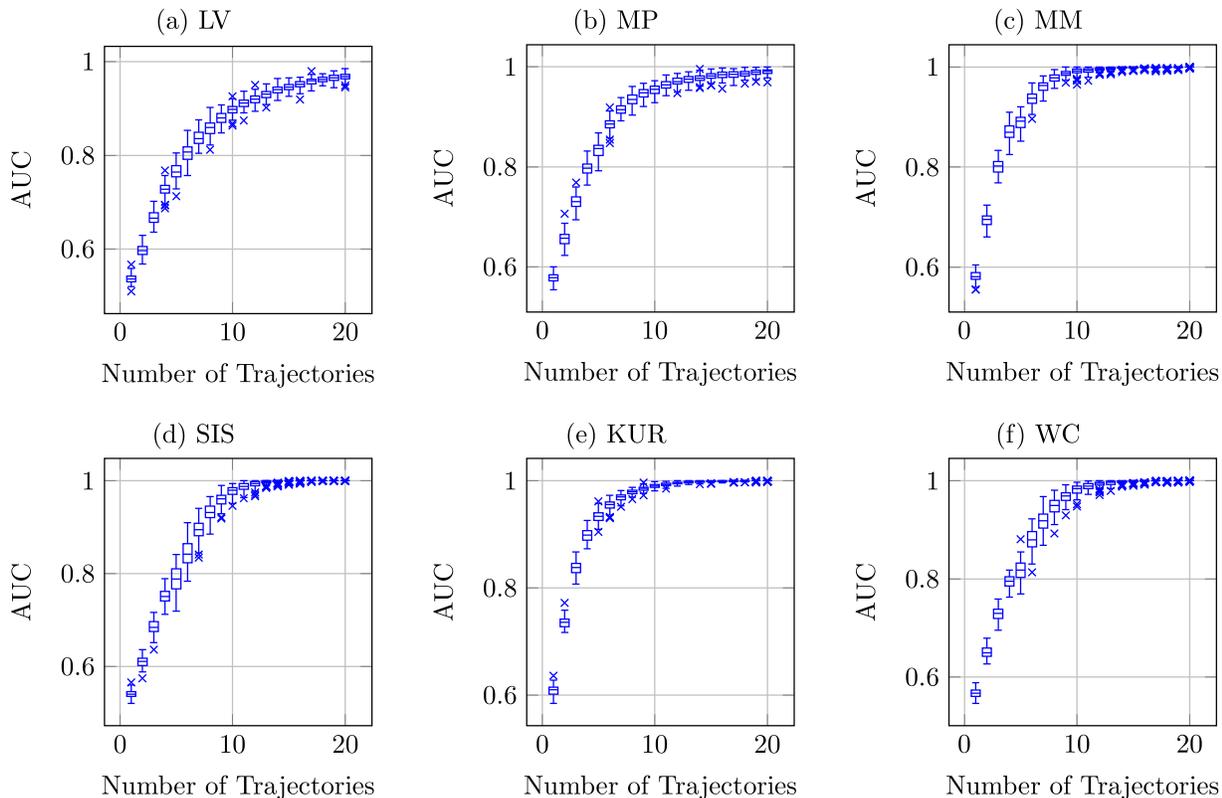


Figure S16: **The similarity of the surrogate network and the true network from observing multiple trajectories on the same network.** The area under the curve AUC versus the number S of observed trajectories $x^{(s)}(t)$ on the same network. The boxplots capture the variation with respect to 100 Barabási-Albert random graphs with $N = 100$ nodes.

second-degree polynomial $f_{\text{runtime}}(N)$ on a log-log scale

$$\log(f_{\text{runtime}}(N)) = s_0 + s_1 \log(N) + s_2 \log^2(N)$$

with coefficients s_0, s_1, s_2 . The fitted values for the leading coefficient are (LV) $s_2 = 2.23$, (MP) $s_2 = 2.43$, (MM) $s_2 = 2.4$, (SIS) $s_2 = 2.15$, (KUR) $s_2 = 3.01$, (WC) $s_2 = 2.44$. Thus, Figure S18 suggests that the computational complexity of Algorithm 1 is quasi-polynomial with respect to the network size N . We emphasise that Algorithm 1 is very amenable for heavy parallelisation, because the commands from lines 4–14 in Algorithm 1 can be run in parallel for each node i .

K Analysing the predictability of low-dimensional dynamics on networks

We aim to analyse the predictability of the network dynamics (2), building upon the central observation that the nodal state $x(t)$ is low-dimensional. Our approach follows a variation of the Lyapunov Exponent [26].

We assume that the nodal state $x(t)$ is observed from the initial time $t = 0$ until the observation time $t = t_{\text{obs}}$. During the observation time interval $t \in [0, t_{\text{obs}}]$, we assume that the POD (2) is exact

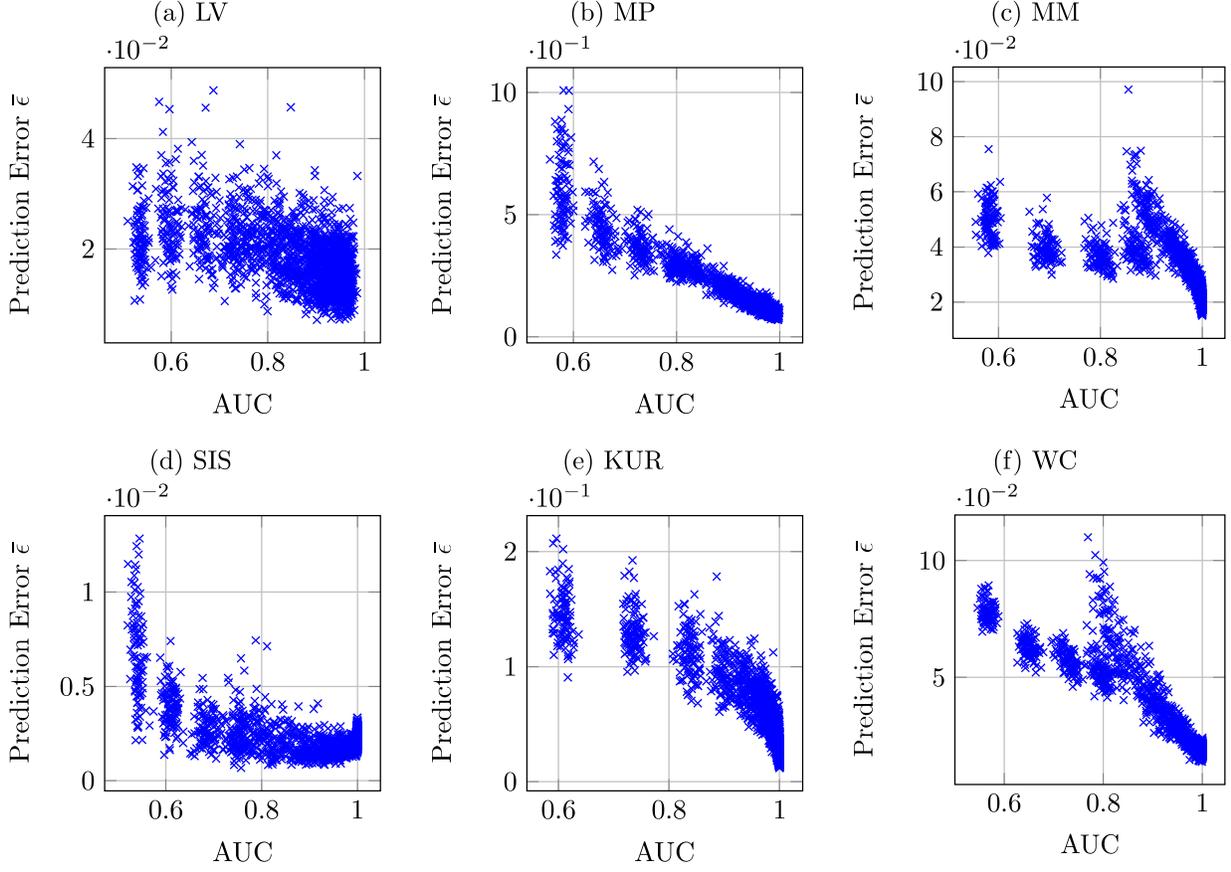


Figure S17: **The prediction accuracy versus the topological similarity.** The average prediction error $\bar{\epsilon}$ versus the area under the curve AUC of the surrogate network \hat{A} . The boxplots capture the variation with respect to 100 Barabási-Albert random graphs with $N = 100$ nodes and across observing $S = 1, \dots, 20$ trajectories.

for some number $m < N$ of agitation modes.

While we consider that the nodal state $x(t)$ is an element of the subspace $\mathcal{X} = \{y_1, \dots, y_m\}$ until the observation time t_{obs} , the POD (2) might not be exact any more at times $t > t_{\text{obs}}$. At $t > t_{\text{obs}}$, additional agitation modes can emerge that were not present for $t \leq t_{\text{obs}}$.

To analyse the predictability of the nodal state $x(t)$, we focus on a perturbed nodal state¹⁰ $\tilde{x}(t)$. The perturbed nodal state $\tilde{x}(t)$ is intended as a *best-case* prediction of the nodal state $x(t)$. The prediction $\tilde{x}(t)$ is obtained by fitting to the nodal state $x(t)$ on the observation time interval $t \in [0, t_{\text{obs}}]$. We consider that the fit is exact. Hence, not only the nodal state $x(t)$ is an element of the subspace \mathcal{X} , but also the prediction $\tilde{x}(t)$ satisfies

$$\tilde{x}(t) = \sum_{p=1}^m \tilde{c}_p(t) y_p \quad (33)$$

at all times $t \in [0, t_{\text{obs}}]$, where the scalar functions $\tilde{c}_p(t) = c_p(t)$ for all agitation modes p at times $t \in [0, t_{\text{obs}}]$.

¹⁰With the notation $\tilde{x}(t)$ for the perturbed nodal state, we stress that the perturbed nodal state $\tilde{x}(t)$ does not exactly equal the nodal state predictions $\hat{x}(t)$ based on Algorithm 1.

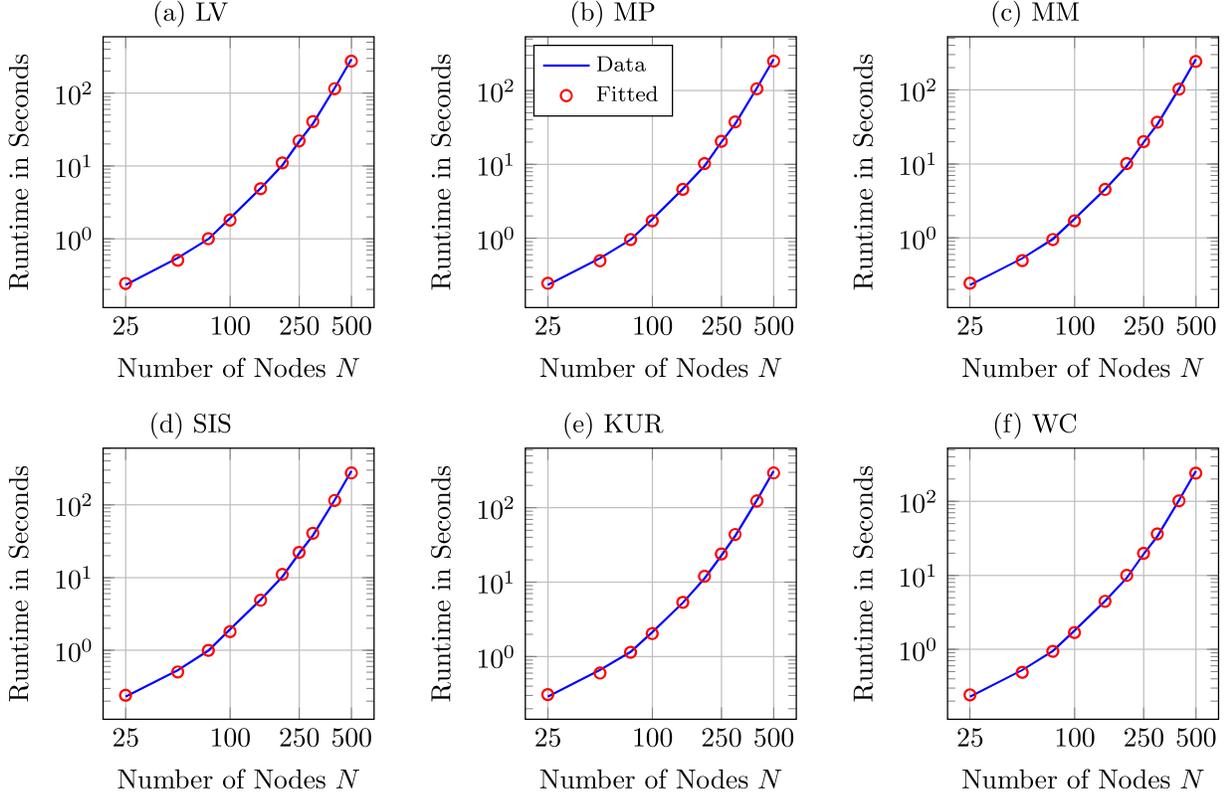


Figure S18: **The computational complexity versus the network size.** The runtime of Algorithm 1 (in seconds) versus the network size N on a loglog-scale. The blue curve is the measured runtime and the red circles correspond to a fitted second-degree polynomial $f_{\text{runtime}}(N)$ in the log-log domain. The results are averaged over 100 and 20 Barabási-Albert random graphs for $N < 400$ and $N \geq 400$ nodes, respectively.

We make two assumptions on the prediction $\tilde{x}(t)$. First, we assume that the POD (33) also holds at future times $t > t_{\text{obs}}$. We do not assume $\tilde{c}_p(t) = c_p(t)$ at future times $t \geq t_{\text{obs}}$. The first assumption is motivated as follows: If we observe the dynamics of the nodal state $x(t)$ only on the subspace \mathcal{X} , then it is hardly possible to make accurate predictions of the nodal state projections $y^T x(t)$ for any vector y in the orthogonal complement \mathcal{X}^\perp of the subspace \mathcal{X} .

Second, we assume that the derivatives $dc_p(t)/dt$ and $d\tilde{c}_p(t)/dt$ are the same for all $p = 1, \dots, m$, provided that $x(t) = \tilde{x}(t)$. To be more precise, consider the definition of the functions $c_p(t)$ in (3), which yields that

$$\begin{aligned} \frac{dc_p(t)}{dt} &= y_p^T \frac{dx(t)}{dt} \\ &= y_p^T f_{\text{dyn}}(x(t)), \end{aligned} \quad (34)$$

where the $N \times 1$ function $f_{\text{dyn}}(x(t)) = (f_{\text{dyn},1}(x(t)), \dots, f_{\text{dyn},N}(x(t)))^T$ is given by the right-hand side of the dynamical model (1), i.e.,

$$f_{\text{dyn},i}(x(t)) = f_i(x_i(t)) + \sum_{j=1}^N a_{ij} g(x_i(t), x_j(t)).$$

Then, the second assumption translates to

$$\frac{d\tilde{c}_p(t)}{dt} = y_p^T f_{\text{dyn}}(\tilde{x}(t)) \quad (35)$$

for all $p = 1, \dots, m$ at every time $t \geq 0$. We emphasise that the second assumption (35) is a *best-case* scenario: The projection of the vector $f_{\text{dyn}}(\tilde{x}(t))$ on the subspace \mathcal{X} is exact for any prediction $\tilde{x}(t) \in \mathcal{X}$. Hence, if $x(t)$ belongs to the subspace \mathcal{X} at all future times $t > t_{\text{obs}}$, then the prediction $\tilde{x}(t)$ would be exact, $\tilde{x}(t) = x(t)$.

Finally, after introducing the two assumptions above, we assess the predictability of the dynamics (1) by studying the deviation

$$\sigma(t) = \tilde{x}(t) - x(t). \quad (36)$$

The sensitivity of the deviation $\sigma(t)$ at future times $t > t_{\text{obs}}$, with respect to small initial perturbations $\sigma(t_{\text{obs}}) \neq 0$, is quantified by the ratio

$$\epsilon_\sigma(t) = \frac{\|\sigma(t)\|_2}{\|\sigma(t_{\text{obs}})\|_2}. \quad (37)$$

for some small initial condition $\sigma(t_{\text{obs}})$ with $\|\sigma(t_{\text{obs}})\|_2 \approx 0$.

Suppose that $\sigma(t_{\text{obs}}) \approx 0$ and $\epsilon_\sigma(t) > 1$. Then, the small initial deviation $\sigma(t_{\text{obs}})$ is amplified from the observation time t_{obs} to time t , which indicates that the prediction $\tilde{x}(t)$ diverges from the nodal state $x(t)$. On the other hand, a ratio satisfying $\epsilon_\sigma(t) < 1$ would indicate that the prediction $\tilde{x}(t)$ converges to the nodal state $x(t)$. Hence, the ratio $\epsilon_\sigma(t)$ serves as a quantification of the predictability of the dynamics (1).

Lemma 3 gives the evolution of the deviation $\sigma(t)$, provided that the deviation $\sigma(t)$ is small. Lemma 3 follows from describing the evolution of the prediction $\tilde{x}(t)$ by linearising the dynamics (1) around the trajectory of the nodal state $x(t)$.

Lemma 3. *Provided that the deviation $\sigma(t)$ is small, the evolution of the deviation $\sigma(t)$ obeys*

$$\frac{d\sigma(t)}{dt} \approx P_{\mathcal{X}} J(x(t)) \sigma(t) - P_{\mathcal{X}^\perp} f_{\text{dyn}}(x(t)). \quad (38)$$

where the $N \times N$ matrix $J(x(t))$ denotes the Jacobian of the function f_{dyn} around the point $x(t)$, the $N \times N$ projection matrix $P_{\mathcal{X}}$ onto the subspace \mathcal{X} is

$$P_{\mathcal{X}} = \sum_{p=1}^m y_p y_p^T \quad (39)$$

and the $N \times N$ projection matrix $P_{\mathcal{X}^\perp}$ onto the orthogonal complement $\mathcal{X}^\perp = \text{span}\{y_{m+1}, \dots, y_N\}$ is

$$P_{\mathcal{X}^\perp} = \sum_{p=m+1}^N y_p y_p^T. \quad (40)$$

Proof. From the definition (36), it follows that the deviation $\sigma(t)$ evolves as

$$\frac{d\sigma(t)}{dt} = \frac{d\tilde{x}(t)}{dt} - \frac{dx(t)}{dt}. \quad (41)$$

First, we express the evolution of the nodal state $x(t)$, i.e., its derivative $dx(t)/dt$, in terms of the agitation modes y_p . The POD (2) is exact if the number of agitation modes satisfies $m = N$. Hence, without loss of generality, the nodal state $x(t)$ can be written as

$$x(t) = \sum_{p=1}^N c_p(t) y_p,$$

and the derivative of the nodal state $x(t)$ equals

$$\frac{dx(t)}{dt} = \sum_{p=1}^N \frac{dc_p(t)}{dt} y_p. \quad (42)$$

With (34), we obtain from (42) that the derivative of the nodal state $x(t)$ equals

$$\frac{dx(t)}{dt} = \sum_{p=1}^N (y_p^T f_{\text{dyn}}(x(t))) y_p.$$

Analogously, with (35), it follows that the prediction $\tilde{x}(t)$ evolves as

$$\frac{d\tilde{x}(t)}{dt} = \sum_{p=1}^m (y_p^T f_{\text{dyn}}(\tilde{x}(t))) y_p.$$

Thus, (41) becomes

$$\begin{aligned} \frac{d\sigma(t)}{dt} &= \sum_{p=1}^m (y_p^T f_{\text{dyn}}(\tilde{x}(t)) - y_p^T f_{\text{dyn}}(x(t))) y_p - \sum_{p=m+1}^N (y_p^T f_{\text{dyn}}(x(t))) y_p \\ &= \sum_{p=1}^m y_p y_p^T (f_{\text{dyn}}(\tilde{x}(t)) - f_{\text{dyn}}(x(t))) - \sum_{p=m+1}^N y_p y_p^T f_{\text{dyn}}(x(t)). \end{aligned} \quad (43)$$

We consider that $\tilde{x}(t) \approx x(t)$ or, equivalently, that $\sigma(t) \approx 0$. Then, at any time t , linearising the dynamics (1) around the point $x(t)$ yields that

$$f_{\text{dyn}}(\tilde{x}(t)) \approx f_{\text{dyn}}(x(t)) + J(x(t)) (\tilde{x}(t) - x(t)),$$

where we denote the $N \times N$ Jacobian matrix at $x(t)$ by $J(x(t))$, whose entries follow from (1) as

$$(J(x(t)))_{ij} = \frac{\partial f_i(x_i(t))}{\partial x_j} + \sum_{j=1}^N a_{ij} \frac{\partial g(x_i(t), x_j(t))}{\partial x_j}, \quad i, j = 1, \dots, N.$$

Hence, the derivative of the deviation $\sigma(t)$ in (43) becomes

$$\begin{aligned} \frac{d\sigma(t)}{dt} &\approx \sum_{p=1}^m y_p y_p^T (f_{\text{dyn}}(x(t)) + J(x(t)) (\tilde{x}(t) - x(t)) - f_{\text{dyn}}(x(t))) - \sum_{p=m+1}^N y_p y_p^T f_{\text{dyn}}(x(t)) \\ &= \sum_{p=1}^m y_p y_p^T J(x(t)) \sigma(t) - \sum_{p=m+1}^N y_p y_p^T f_{\text{dyn}}(x(t)), \end{aligned}$$

where the second equality follows from the definition of the deviation $\sigma(t)$ in (36). The definition of the matrices $P_{\mathcal{X}}$ and $P_{\mathcal{X}^\perp}$ in (39) and (40) completes the proof. \square

The first term in (38) captures the network dynamics (1) on the POD subspace \mathcal{X} , linearised around the trajectory of the nodal state $x(t)$. The second term corresponds to the exact (i.e., not linearised) changes of the nodal state $x(t)$ that are orthogonal to the POD subspace \mathcal{X} .

We study the predictability of the dynamics (1) by the ratio $\epsilon_\sigma(t)$ in (37), based on solving (38) numerically. To obtain the initial condition $\sigma(t_{\text{obs}})$, we randomly generate the initial prediction state $\tilde{x}(t_{\text{obs}})$ from a small a perturbation of the nodal state $x(t_{\text{obs}})$. For every node i , the POD $\tilde{x}_i(t_{\text{obs}})$ is set to $\tilde{x}_i(t_{\text{obs}}) = \xi_i x_i(t_{\text{obs}})$, where ξ_i is drawn from a normal distribution with mean $\text{E}\{\xi_i\} = 1$ and variance $\text{Var}\{\xi_i\} = 10^{-6}$. Thus, the mean and variance of the initial deviation $\sigma_i(t_{\text{obs}})$ are given by $\text{E}\{\sigma_i(t_{\text{obs}})\} = 0$ and $\text{Var}\{\sigma_i(t_{\text{obs}})\} = 10^{-6}$, independently of the node i . The evolution of the deviation $\sigma(t)$ is computed by (37). We choose the number of agitation modes m , to define the matrices $P_{\mathcal{X}}$ and $P_{\mathcal{X}^\perp}$, by the Matlab command `rank` of the nodal state matrix $(x[1], \dots, x[n])$. Hence, the nodal states $x(t)$ and $\tilde{x}(t)$ are virtually equal on the observation time interval $t \in [0, t_{\text{obs}}]$.

We randomly generate 100 initial states $\tilde{x}_i(t_{\text{obs}})$, which results in 100 ratio $\epsilon_\sigma(t)$. We denote the average of the 100 ratios as $\bar{\epsilon}_\sigma(t)$ and define the logarithm as predictability parameter

$$\Lambda(t) = \log(\bar{\epsilon}_\sigma(t)).$$

The sign of the predictability parameter $\Lambda(t)$ indicates either one of two cases. If $\Lambda(t) < 0$, then we consider the dynamics (1) to be predictable until time¹¹ t , since small deviations $\sigma(t_{\text{obs}})$ decrease from the observation time t_{obs} until the time t . Otherwise, if $\Lambda(t) \geq 0$, then we call the dynamics (1) not predictable until time t .

The predictability parameter $\Lambda(t)$ is closely related to the (maximum) Lyapunov exponent [26], which is the inspiration for our approach. More specifically, suppose that the subspace \mathcal{X} of the dynamics $x(t)$ equals \mathbb{R}^N . Then, the maximum Lyapunov exponent equals the maximum of

$$\lim_{t \rightarrow \infty} \frac{1}{t} \log \left(\frac{\|\sigma(t)\|_2}{\|\sigma(0)\|_2} \right)$$

with respect to the initial deviation $\sigma(0)$. In contrast to the maximum Lyapunov exponent, the predictability parameter $\Lambda(t)$ is not normalised by the time t and is evaluated for a finite time t , subject to dynamics in the subspace \mathcal{X} .

We are interested if the predictability parameter $\Lambda(t)$ is meaningful in understanding the accuracy of our prediction method based on Algorithm 1. In particular, we are curious about the dependence of the prediction error $\epsilon(t)$, defined in (30), on the predictability parameter $\Lambda(t)$. For obtaining the error $\epsilon(t)$ of our prediction method, we consider the same parameter setting as detailed in the beginning of Appendix G to generate nodal states $x(t)$ and predictions $\hat{x}(t)$ based on the surrogate network \hat{A} of Algorithm 1. However, we make one change in the parameters: To capture the variations of the predictability with respect to the observation time t_{obs} , we generate the time t_{obs} as a uniform random number in $[0.05T_{\text{max}}, 0.5T_{\text{max}}]$ for each simulation run. Hence, the observation time t_{obs} is variable for every run, but the maximum prediction time T_{max} is constant.

¹¹The predictability parameter $\Lambda(t)$ depends not only on the dynamics (1) but also on the initial nodal state $x(0)$, the observation time t_{obs} and the number agitation modes m . Hence, a more accurate, but cumbersome, formulation would be “predictable until time t given the initial nodal state $x(0)$, the observation time t_{obs} and the number agitation modes m ”.

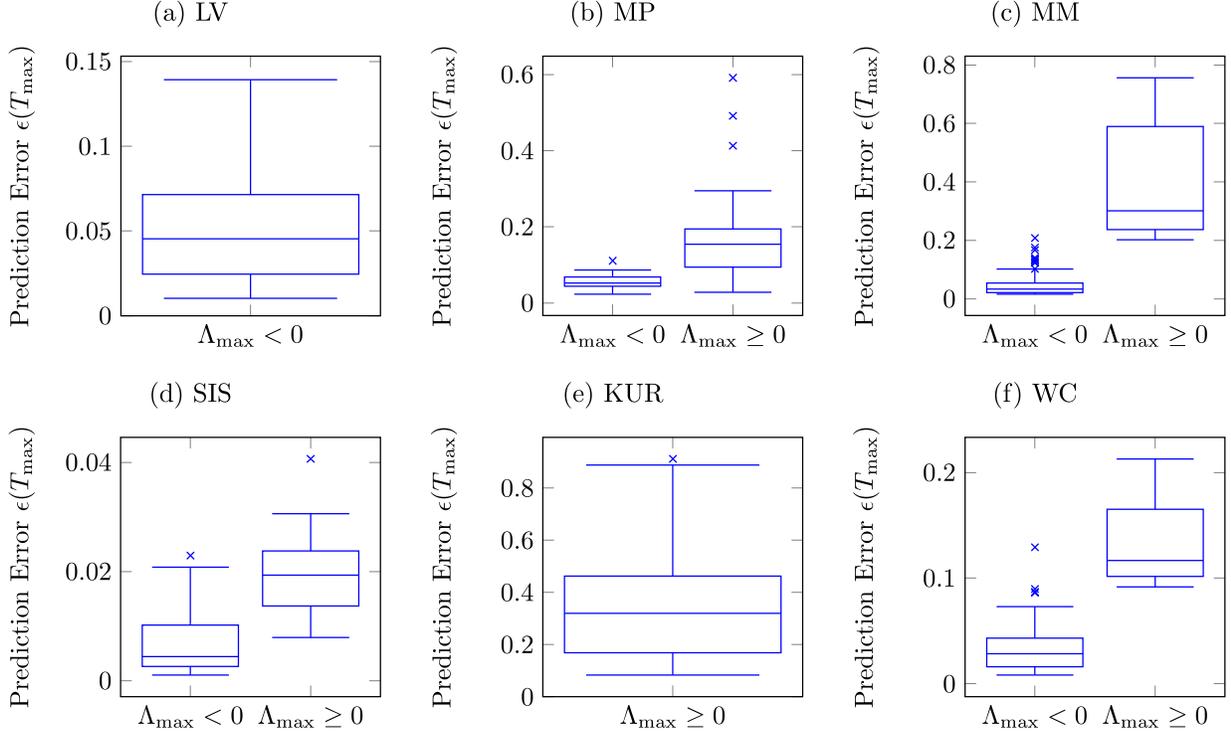


Figure S19: **The prediction error versus the sign of the predictability parameter.** The prediction error $\epsilon(T_{\max})$ at the end of the prediction interval T_{\max} versus the maximum Λ_{\max} of the prediction parameter $\Lambda(t)$ on the prediction interval $t \in [t_{\text{obs}}, T_{\max}]$. The boxplots capture the variation with respect to 100 Barabási-Albert random graphs with $N = 100$ nodes. For the LV and the KUR models, we obtained $\Lambda_{\max} \geq 0$ for every run.

Figure S19 shows that the prediction accuracy can indeed be assessed the sign of the predictability parameter $\Lambda(t)$. We emphasise the predictions $\tilde{x}(t)$ are the output of the specific Algorithm 1, whereas the predictability parameter $\Lambda(t)$ is agnostic to the specifics of the prediction algorithm. The parameter $\Lambda(t)$ only depends on the dynamics (1) and the prediction parameters t_{obs}, T_{\max} .

Figure S20 shows that, conditioned on the realisations with $\Lambda_{\max} \geq 0$, there is a high correlation between the prediction error $\log(\epsilon(t))$ and the parameter $\Lambda(t)$ across time t . The realisations that resulted in $\Lambda_{\max} < 0$ were excluded, since the parameter $\Lambda(t)$ rapidly converges to zero if $\Lambda_{\max} < 0$, which is in contrast to the increasing prediction error $\epsilon(t)$, see Figure S10. Thus, given that $\Lambda_{\max} < 0$, we expect that the prediction error $\epsilon(t)$ cannot be assessed by a parameter $\Lambda(t)$ that is agnostic to the specifics of Algorithm 1.

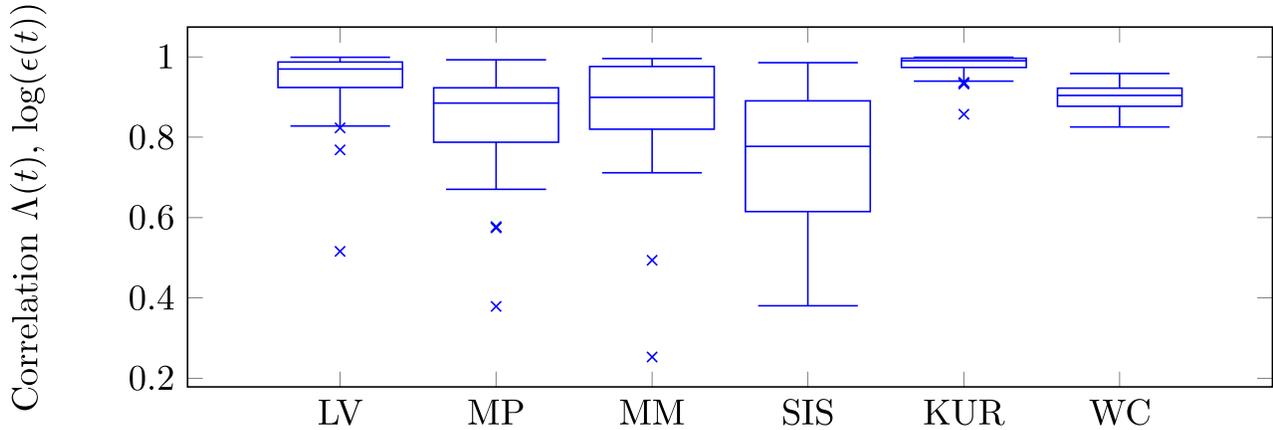


Figure S20: **The correlation of the prediction error and the predictability parameter.** The correlation of the logarithm of the prediction error $\log(\epsilon(t))$ and the prediction parameter $\Lambda(t)$ on the time interval $t \in [t_{\text{obs}}, T_{\text{max}}]$. Only realisations with $\Lambda_{\text{max}} \geq 0$ are included. The boxplots capture the variation with respect to 100 Barabási-Albert random graphs with $N = 100$ nodes.

References

- [1] Neo D Martinez. “Artifacts or attributes? Effects of resolution on the Little Rock Lake food web”. In: *Ecological Monographs* 61.4 (1991), pp. 367–392.
- [2] Jérôme Kunegis. “Konect: the Koblenz network collection”. In: *Proceedings of the 22nd International Conference on World Wide Web*. ACM. 2013, pp. 1343–1350.
- [3] Makoto Kato et al. “Insect-flower relationship in the primary beech forest of Ashu, Kyoto: an overview of the flowering phenology and the seasonal pattern of insect visits”. In: *Contributions from the Biological Laboratory Kyoto University* 27 (1990), pp. 309–375.
- [4] Enrico L Rezende et al. “Non-random coextinctions in phylogenetically structured mutualistic networks”. In: *Nature* 448.7156 (2007), p. 925.
- [5] Uzi Harush and Baruch Barzel. “Dynamic patterns of information flow in complex networks”. In: *Nature Communications* 8.1 (2017), p. 2181.
- [6] Ron Milo et al. “Network motifs: simple building blocks of complex networks”. In: *Science* 298.5594 (2002), pp. 824–827.
- [7] Baruch Barzel and Albert-László Barabási. “Universality in network dynamics”. In: *Nature Physics* 9.10 (2013), p. 673.
- [8] Lorenzo Isella et al. “What’s in a crowd? Analysis of face-to-face behavioral networks”. In: *Journal of Theoretical Biology* 271.1 (2011), pp. 166–180.
- [9] Pauline Van den Driessche and James Watmough. “Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission”. In: *Mathematical Biosciences* 180.1-2 (2002), pp. 29–48.
- [10] Ana Lajmanovich and James A Yorke. “A deterministic model for gonorrhoea in a nonhomogeneous population”. In: *Mathematical Biosciences* 28.3-4 (1976), pp. 221–236.

- [11] Michael Breakspear, Stewart Heitmann, and Andreas Daffertshofer. “Generative models of cortical oscillations: neurobiological implications of the Kuramoto model”. In: *Frontiers in Human Neuroscience* 4 (2010), p. 190.
- [12] Nathalie Tzourio-Mazoyer et al. “Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain”. In: *NeuroImage* 15.1 (2002), pp. 273–289.
- [13] David C Van Essen et al. “The WU-Minn Human Connectome Project: an overview”. In: *NeuroImage* 80 (2013), pp. 62–79.
- [14] Prejaas Tewarie et al. “How do spatially distinct frequency specific MEG networks emerge from one underlying structural connectome? The role of the structural eigenmodes”. In: *NeuroImage* 186 (2019), pp. 211–220.
- [15] Edward Laurence et al. “Spectral dimension reduction of complex dynamical networks”. In: *Physical Review X* 9.1 (2019), p. 011042.
- [16] John G White et al. “The structure of the nervous system of the nematode *Caenorhabditis elegans*”. In: *Philos. Trans. Roy. Soc. London* 314.1165 (1986), pp. 1–340.
- [17] Beth L Chen, David H Hall, and Dmitri B Chklovskii. “Wiring optimization can relate neuronal structure and function”. In: *Proceedings of the National Academy of Sciences* 103.12 (2006), pp. 4723–4728.
- [18] Lav R Varshney et al. “Structural properties of the *Caenorhabditis elegans* neuronal network”. In: *PLoS Computational Biology* 7.2 (2011).
- [19] B Prasse and P Van Mieghem. “Network Reconstruction and Prediction of Epidemic Outbreaks for General Group-Based Compartmental Epidemic Models”. In: *IEEE Transactions on Network Science and Engineering* (2020).
- [20] Christoph Bergmeir and José M Benitez. “On the use of cross-validation for time series predictor evaluation”. In: *Information Sciences* 191 (2012), pp. 192–213.
- [21] Seung-Jean Kim et al. “An Interior-Point Method for Large-Scale l_1 -Regularized Least Squares”. In: *IEEE Journal of Selected Topics in Signal Processing* 1.4 (2007), pp. 606–617.
- [22] Emmanuel J Candès and Michael B Wakin. “An introduction to compressive sampling”. In: *IEEE signal processing magazine* 25.2 (2008), pp. 21–30.
- [23] Trevor Hastie, Robert Tibshirani, and Martin Wainwright. *Statistical learning with sparsity: the lasso and generalizations*. CRC press, 2015.
- [24] Scott Shaobing Chen, David L Donoho, and Michael A Saunders. “Atomic decomposition by basis pursuit”. In: *SIAM review* 43.1 (2001), pp. 129–159.
- [25] Piet Van Mieghem. *Graph spectra for complex networks*. Cambridge University Press, 2010.
- [26] Steven H Strogatz. *Nonlinear Dynamics and Chaos: With applications to Physics, Biology, Chemistry, and Engineering*. CRC press, 2018.