Autonomous inference of complex network dynamics from incomplete and noisy data

Ting-Ting Gao$^{1,2}$ and Gang Yan$^{1,2,3,*}$

1. MOE Key Laboratory of Advanced Micro-Structured Materials and School of Physics Science and Engineering, Tongji University, Shanghai 200092, P. R. China
2. Frontiers Science Center for Intelligent Autonomous Systems, Tongji University, Shanghai 200092, P. R. China
3. Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai 200031, P. R. China

* Correspondence to Gang Yan (gyan@tongji.edu.cn).

The availability of empirical data that capture the structure and behavior of complex networked systems has been greatly increased in recent years, however a versatile computational toolbox for unveiling a complex system’s nodal and interaction dynamics from data remains elusive. Here we develop a two-phase approach for autonomous inference of complex network dynamics, and its effectiveness is demonstrated by the tests of inferring neuronal, genetic, social, and coupled oscillators dynamics on various synthetic and real networks. Importantly, the approach is robust to incompleteness and noises, including low resolution, observational and dynamical noises, missing and spurious links, and dynamical heterogeneity. We apply the two-phase approach to inferring the early spreading dynamics of H1N1 flu upon the worldwide airline network, and the inferred dynamical equation can also capture the spread of SARS and COVID-19 diseases. These findings together offer an avenue to discover the hidden microscopic mechanisms of a broad array of real networked systems.

Introduction

From two-photon calcium imaging of neuronal activities$^{1,2}$, high-throughput genetic experiments$^{3,4}$ to digital recordings of human mobility$^{5–7}$, our ability to observe the dynamic behavior of nodes in complex biological, social and technological systems has advanced spectacularly in the past years. The collected observations, often in the form of time-series data, allow us to extract the dynamic patterns of a system’s individual nodes. To gain meaningful insights on the system, however, such a reductionist approach of tracking all individual nodes is insufficient. Indeed, complex system behavior emerges not just from the single nodes, but rather from the dynamic interactions between the nodes$^{6,8–18}$. This requires us to infer complex network dynamics, i.e. to retrieve both self nodal dynamics and interaction dynamics from the accumulating data of network topological structure and nodes’ activities.

The balance of self vs. interaction dynamics is most naturally captured by a general equation that tracks the activities of all nodes via$^9$

$$\frac{dx_i(t)}{dt} = F(x_i(t)) + \sum_{j=1}^{n} A_{ij} G(x_i(t), x_j(t)),$$

(1)

where $x_i(t) \equiv (x_{i,1}(t), \ldots, x_{i,d}(t))^T$ is node $i$’s $d$-dimensional activity, representing, e.g. the membrane potential of a neuron in a brain network$^{9,12}$, the proportion of infected people in a country or region$^{5–7}$.
or the state of a component in an oscillator network\textsuperscript{19}. These activities are driven by the self-regulation function $\mathbf{F}(\mathbf{x}_i) \equiv (F_1(\mathbf{x}_i), \ldots, F_d(\mathbf{x}_i))^\top$, designed to describe the dynamics of all nodes in isolation, and by the pairwise function $\mathbf{G}(\mathbf{x}_i(t), \mathbf{x}_j(t)) \equiv (G_1(\mathbf{x}_i, \mathbf{x}_j), \ldots, G_d(\mathbf{x}_i, \mathbf{x}_j))^\top$ which captures the dynamic mechanisms of interaction between the nodes. Finally, the network $A_{ij}$, an $n \times n$ adjacency matrix, denotes the influence or flow from node $j$ to $i$, where $n$ is the number of nodes in the system. As shown by Barzel and Barabási, with appropriate choices of the nonlinear functions $\mathbf{F}$ and $\mathbf{G}$, Eq. (1) is able to describe a broad range of complex systems\textsuperscript{9}. However, for most real systems, the functions $\mathbf{F}$ and $\mathbf{G}$ are unknown. Hence, a pressing lacuna in the study of complex systems is a versatile computational toolbox for automatically inferring Eq. (1) from the observed data of network topology $A_{ij}$ and node activities $\mathbf{x}_i(t)$.

Complex biological, social or technological systems lack the fundamental physical rules that govern particle systems, so we do not have a\textit{ priori} knowledge of their internal microscopic mechanisms\textsuperscript{30}. Therefore, the goal is not to only identify the model’s parameters but rather to retrieve the forms of $\mathbf{F}$ and $\mathbf{G}$ and infer the explicit model itself. Despite the recent significant progress in developing methods to infer the governing equations of single- or few-body dynamics\textsuperscript{21–27}, the task of inferring network dynamics poses particular challenges. For example, $\mathbf{F}$ and $\mathbf{G}$ are usually of different types hence one cannot obtain their compact forms if using only orthogonal basis functions\textsuperscript{22,23,28,29}; Node activities data are noisy and the mappings of network topologies are usually incomplete\textsuperscript{30,31}; Collective behavior, such as synchronization and consensus\textsuperscript{19}, can conceal the specific forms of microscopic mechanisms in interaction dynamics. To overcome these challenges we propose here a two-phase inference approach. Our analysis indicates that the two-phase strategy allows us to achieve efficient and, most importantly, highly accurate inference, even in the face of unfavorable scenarios, such as noisy or low-resolution data or an only partially mapped topology (Fig. 1a).

**Results**

**Overview of the two-phase inference approach**

Lacking a\textit{ priori} knowledge of the structure of $\mathbf{F}$ and $\mathbf{G}$, a natural approach is to pre-construct two extensive libraries $L_F$ and $L_G$ that contain a variety of elementary functions. The combinations of these elementary functions can potentially generate the true network dynamics. In this work, the libraries contain not only orthogonal basis functions but include polynomials, trigonometric, exponential, fractional, rescaling, sigmoid and other activation functions that frequently used in various domains (see Supplementary Tables 1 and 2). Large libraries are helpful for finding a compact and optimal model to capture network dynamics but also make the inference problem more difficult, because, due to the lack of orthogonality, the elementary functions can be similar with each other and thus less discriminative.

Introducing the time series data $\mathbf{x}_i(t)$, where $i = 1, 2, \ldots, n$, into $L_F$ and $L_G$, we obtain two time-varying matrices $\Theta_F(t) \equiv L_F(\mathbf{x}_i(t))$ and $\Theta_G(t) \equiv L_G(\mathbf{x}_i(t), \mathbf{x}_j(t))$ that encode the patterns of node activities imposed by the elementary functions in $L_F$ and $L_G$ (Fig. 1b). Then the inference problem can be recast to selecting appropriate patterns in $\Theta_F(t)$ and $\Theta_G(t)$ that best match the evolution of observed system state $\dot{\mathbf{x}}(t)$, i.e. to inferring the sparse coefficients $\xi_F$ and $\xi_G$ that best solve

$$
\dot{\mathbf{x}}(t) = \tilde{\Theta}_F(t)\xi_F + \tilde{A}\tilde{\Theta}_G(t)\xi_G,
$$

where $\tilde{A} \equiv A \otimes I_d$, $\tilde{\Theta}_F \equiv \Theta_F \otimes I_d$, $\tilde{\Theta}_G \equiv \Theta_G \otimes I_d$, symbol $\otimes$ denotes Kronecker product and $I_d$ is the $d$-dimensional identity matrix. Here, we consider the general setting where each node state is $d$-dimensional and the network is directed and heterogeneous. Consequently, the problem of inferring complex network
dynamics is high-dimensional and irreducible. Indeed, the number of elementary functions in $L_F$ and $L_G$ is approximately 25, 80 or 140 when node activity itself has one, two, or three dimensions respectively in the simulation validations below (see Supplementary Tables 1 and 2).

Our approach is a two-phase procedure consisting of global regression and local fine-tuning. In phase I, we approximate the derivatives $\dot{x}(t)$ (see Methods) and calculate the matrices $\tilde{\Theta}_F(t)$ and $\tilde{\Theta}_G(t)$, and then normalize each column of them (Fig. 1b). These normalized data are used to identify, through regression, the leading elementary functions that are most probably constituents of the true $F$ and $G$ (Fig. 1c, and see also Methods). Phase I is able to significantly narrow down the model space but the dynamical equation inferred by such regression alone lacks generative power (Fig. 1d). Next, in phase II, we perform fine-tuning with the original values of $\dot{x}(t)$, $\tilde{\Theta}_F(t)$ and $\tilde{\Theta}_G(t)$, i.e. without normalization.

We use topological samplings (see Methods) and the weighted Akaike’s information criterion (wAIC, see Methods) to sequentially remove the elementary functions with smallest inferred coefficients (Fig. 1e). The final sets of elementary functions and their coefficients $\hat{\xi}_F$ and $\hat{\xi}_G$ compose $\hat{F}$ and $\hat{G}$, leading to the inferred dynamics of complex networks (Fig. 1f).

Inferring complex network dynamics

To validate the effectiveness of our approach we apply it to infer five network dynamics, including Hindmarsh-Rose ($HR, d = 3$) and FitzHugh-Nagumo ($FHN, d = 2$) neuronal systems, social balance dynamics ($SB, d = 1$), Kuramoto dynamics ($d = 1$), and coupled heterogeneous Rössler oscillators ($d = 3$), here $d$ is the dimension of each node activity. To obtain the nodes’ activities data we respectively simulate these dynamics (Supplementary Table 4) on a variety of topologies, including Erdős-Rényi (ER) and scale-free (SF) synthetic networks, and five empirical networks – cellular-level brain networks of $C. elegans$ and $Drosophila$, Advogato social network, power grids of North-Europe and United States. The time series of node activities and each network topology are the input data to our approach. The five specific equations governing these dynamics are the ground truths that we aim to infer. These dynamical models and networks are widely used in various domains and exhibit different properties (see Supplementary Information Sections II and III), accounted for the diversity of our tests.

Figure 2 illustrates the procedure of inferring FitzHugh-Nagumo neuronal network dynamics. Through the global regression, Phase I identifies ten most relevant elementary functions for each dimension of FHN (Fig. 2b), and then, by local fine-tuning, Phase II autonomously learns the compact and optimal form of the dynamical equation as well as the most appropriate coefficient for each of the necessary elementary functions (Fig. 2c). The form of the inferred equation in Fig. 2c perfectly matches the ground truth in Fig. 2a and the learnt coefficients are also highly accurate. Indeed, the relative errors $\Delta = (\xi - \hat{\xi})/\xi$, where $\xi$ and $\hat{\xi}$ are true and learnt coefficients respectively, are smaller than 3% (Fig. 2d). The dynamical equation inferred by our approach exhibits generative power, being able to generate nodes’ activities and trajectories that agree well with observation data (Fig. 2e,f).

Our approach also successfully infers the equations governing other four network dynamics. Regarding the accuracy of learnt coefficients, the relative errors $|\Delta| < 3\%$ for the Hindmarsh-Rose dynamics (Fig. 3a) and the edge dynamics (Fig. 3c) on both synthetic and empirical networks. In Kuramoto dynamics and coupled heterogeneous Rössler oscillators the self-dynamics are nonidentical, i.e. each node’s dynamics has its own form (see Supplementary Information Section III). Hence we aim to infer an effective form of Eq. (1) that minimizes the inconsistency between inferred and true nodes’ activities. Even for these more challenging cases, the two-phase approach still succeeds with relative coefficient errors $|\Delta| < 5\%$ or $< 20\%$ (Fig. 3e,g)). Both activities and trajectories generated by the effective equations exhibit high agreement with the true averaging dynamics (Fig. 3f,h,i)).
Inferrability of network dynamics

Whether a network dynamics is inferrable depends on several factors. Here we explore three key factors, namely synchronized dynamics, dynamical heterogeneity, and deficient libraries.

Synchronized dynamics. If a network is completely synchronized, i.e., all of its nodes behave in the same manner\textsuperscript{19,34,35}, distinguishing the activities of a node and its neighbors becomes impossible and the microscopic interacting mechanism $G(x_i, x_j)$ between nodes will be cloaked and undiscoverable. In other words, more synchronized is a network, more difficult to infer its dynamics. Here we tune the coupling strength between nodes to change the degree of network synchronization (i.e. the order parameter $\langle R \rangle$, see Supplementary Information Section IV), and test the capability of our two-phase approach in inferring partially synchronized network dynamics. As shown in Fig. 4a, although the inference inaccuracy increases when the system becomes more synchronized, our approach still can infer the true FHN equation even when the network is highly synchronized ($\langle R \rangle \approx 0.7$). The inference inaccuracy is quantified by symmetric mean absolute percentage error (sMAPE, see Methods). The more accurate the inference result, the closer the sMAPE value to zero.

Dynamical heterogeneity. Equation (1) assumes that nodes have the same form $F$ of self-dynamics, yet this is not always true. For instance, although the self-dynamics of Kuramoto model is simply one elementary function $\omega$ representing the natural frequency of a node, different nodes can have different values of $\omega$. For such nonidentical self-dynamics it is difficult, if not impossible, to infer a specific form $F_i(x_i)$ for each node $i$ due to an $n$-fold increase in the dimensionality of potential model space ($n$ is network size). Therefore we aim to infer an effective equation that best captures the averaging dynamics, as shown in Fig. 3(e,g). Here we further explore to what extent of dynamical heterogeneity our approach can tolerate. To do so we assign each node a value of $\omega$ randomly drawn from a normal distribution $N(0, \sigma)$ and increase the standard deviation $\sigma$. The inference inaccuracy indeed increases when $\sigma$ becomes larger, and the two-phase approach can tolerate dynamical heterogeneity $\sigma \leq 0.5$ (Fig. 4b).

Deficient libraries. Although two rather comprehensive libraries of elementary functions are built, it is still possible that some elementary functions of the true unknown dynamics are missing. Another possibility is that the compact form of true dynamics cannot be composed by these elementary functions. For these cases, our two-phase approach will infer an alternative equation to capture the system behaviors. We test such capability in gene regulation and Hindmarsh-Rose neuronal dynamics whose true coupling functions are intentionally removed from $L_G$. As shown in Fig. 4c, the trajectories generated by the inferred and the true equations are close to each other, and the discrepancy is small for all nodes (see Methods and also Supplementary Information Section IV-B).

Inferring from incomplete and noisy data

Incompleteness of mapped network topology and noises of observed nodes’ activities are inevitable in real data\textsuperscript{30,31}. Hence, here we validate the robustness of our two-phase approach against low resolution, dynamical and observational noises, spurious and missing links, as well as through comparisons with previous methods\textsuperscript{23,36,37}.

Low resolution. Experimental and digitally recording technologies often have limited measurement frequencies, inducing low resolution of observed time series. To validate our approach’s robustness against low resolution we numerically simulate the five nonlinear network dynamics in Figs. 2 and 3 with step size 0.01, and then regularly down-sample the activities data. We calculate the failure ratios in inferring the form of true equations (Supplementary Figure 14a) and also the inference inaccuracies (Fig. 5a). The results show that the two-phase strategy requires only a proportion of 5\% to 50\% data for the inference.

Observational and dynamical noises. Observational noises are induced by the measuring process and
dynamical noises represent the intrinsic stochasticity in dynamics. To produce the former we add Gaussian
noises to nodes' activity data and quantify the intensity of observational noise with signal-to-noise-ratio
(SNR, see Supplementary Information Section V-A); To imitate the latter we add a stochastic term of
Gaussian white noise with intensity \( \eta \) into the true dynamical equations and generate the nodes’ activities
data by numerical simulations of these stochastic differential equations (see Supplementary Information
Section V-A). We test the impact of these two types of noises on the performance of the two-phase
inference approach, without any denoising preprocess. As shown in Fig. 5b and Supplementary Figure
14b, the approach can tolerate dynamical noise with \( \eta \leq 0.15 \), meaning that it successfully reconstructs
the hidden equations when the stochastic intensity is not higher than 15% of the average amplitude of
ture deterministic dynamics. Moreover, the approach can tolerate 30-dB observational noise (Fig. 5c and
Supplementary Figure 14c).

Spurious and missing links. Spurious and missing links in real data induce an incomplete network topology
\( A_{ij} \), which further lead to an inaccurate interaction matrix \( \Theta_G \). To test the impact of these erroneous
links we randomly add or remove a fraction of links from the true network topology that was used to
simulate the nodes’ activities. Owing to the topological sampling in Phase II, our approach is able to
tolerate 25% spurious and 30% missing links (Fig. 5(d,e) and Supplementary Figure 14(d,e)).

Comparison with previous methods. Two most illuminating and effective methods for dynamics inference
are SINDy\(^{23}\) and ARNI\(^{37}\). Note that ARNI originally aimed at inferring network topology but can be
transferred to infer network dynamics by minor modification (see Supplementary Information Section V-
C). Here we compare our approach with SINDy and ARNI from different aspects, including the amount
of required data (Fig. 5f), the robustness against observational noise (Fig. 5g), correlated dynamical noise
(Fig. 5h and also Supplementary Information Section V-A), missing links (Fig. 5i), and different network
sizes (Fig. 5j). While ARNI needs fewer data points if network topologies are complete and nodal activities
do not have any noise (Fig. 5f), the two-phase approach outperforms both SINDy and ARNI in inferring
complex network dynamics from incomplete and noisy data (Fig. 5g-j). We also perform comparisons with
SINDy’s variant\(^{36}\) regarding partially synchronized or heterogeneous dynamics (Supplementary Figures
13 and 17). These results indicate that our approach can better handle high-dimensional networked
systems and better cope with incompleteness and noises in data.

Ablation studies. Besides the two-phase strategy, our approach also involves three important components,
namely, normalization in the first phase yet non-normalization in the second for solving the issue raised by
highly skewed observations at different nodes, topological sampling for imitating the feature of observed
incomplete topologies, and optimal selection by wAIC for determining the most appropriate complexity of
inferred dynamics. The essentiality of the two-phase strategy and the three components is demonstrated
by ablation studies. Specifically, we ablate each phase or component and then assess the performance
of degenerated approaches. As shown in Fig. 5(k,l) and Supplementary Information Section V-B, the
inference inaccuracy sMAPE indeed significantly increases if the phases or components are individually
ablated.

Inference of empirical systems

To demonstrate the approach’s ability of handling empirical systems, we apply it to infer the spreading
dynamics of infectious disease H1N1. The network underlying this diffusion system is the worldwide
airline network, which captures the human mobility between different countries or regions and plays a
dominant role for global disease spreading\(^{5,6}\). Each entry \( A_{ij} \) of the weighted network’s adjacency matrix
\( A \) represents the traffic volume from node \( j \) to \( i \), where each node denotes a country or region. The total
passengers daily are approximately \( \Phi = 8.9 \times 10^6 \) and, taking into account the population \( P_i \) of each
node \( i \), the adjacency matrix is modified to
\[
\hat{A}_{ij} = \frac{\Phi}{\sum_{i=1}^{n} P_i} A_{ij}.
\] (3)

The magnitude order of entries in matrix \( \hat{A} \) is around \( 10^{-2} \) to \( 10^{-3} \). The nodal activities \( x_i(t) \) are extracted from the daily reports of infected cases in each country or region. Here we consider the nodes whose accumulated H1N1 cases are more than 100 and focus on the early spreading dynamics, i.e. within the 45 days since the first case was reported in each node, which captures the system behaviors before government control.

Based on these empirical data, our approach successfully infers a concise effective dynamical equation
\[
\frac{dx_i}{dt} = ax_i + b \sum_{j=1}^{N} \hat{A}_{ij} \frac{1}{1 + e^{-(x_j - x_i)}}.
\] (4)

where \( a = 0.074 \) and \( b = 7.130 \) (see Supplementary Information Section VI and Supplementary Figure 18). It is interesting that our approach infers a sigmoid (nonlinear) form, rather than the linear form of epidemic models, to better capture the interaction dynamics. This might be caused by the fact that people usually consciously travel less if their countries/regions or the destinations have high infection risk. While Eq. 4 describes the dynamics of all nodes with the same parameters \( a \) and \( b \), we also extend it by taking into account of dynamical heterogeneity in the nodes, i.e. to obtain \( a_i \) and \( b_i \) from each node \( i \)'s activity data (see Fig. 6(b-e) and also Supplementary Figure 18).

Because empirical systems lack ground-truths, we verify the inferred Eq. 4 by testing its generalizability to SARS and COVID-19 diseases. Based on the daily reported numbers within the first 45 days in each node, we find that Eq. 4 is also able to capture the early spread of SARS and COVID-19 upon the worldwide airline network. Indeed, as shown in Fig. 6(f-i) and Supplementary Figures 19-20, the evolution of cumulative numbers of SARS cases (for nodes whose eventual infected cases are more than 100) and COVID-19 cases (for nodes whose eventual infected cases are more than 2000) agree well with the activities generated by Eq. 4 with heterogeneous parameters \( a_i \) and \( b_i \).

Discussion

Many real networks have been mapped so far but there are still complex systems whose network structure information is totally missing. For the latter cases, a possible scheme is inferring their topological structure, especially directed or causal networks, from nodes’ activities data first and then applying our approach to infer system dynamics. It is worth noting that inferring network structure from nodes’ activities data is also challenging, especially when the number of nodes is large, because the number of parameters needing to be estimated is about \( n^2 \) where \( n \) is network size. Therefore, how to simultaneously infer both structure and dynamics of large complex systems is still an outstanding problem.

Our work also raises several questions worthy of future pursuit. First, stochasticity in the dynamics of some real complex systems might be stronger than that we considered in this work. Such highly stochastic systems are better described by stochastic differential equations. Second, our approach does not account for discrete or boolean dynamics, or systems that contain thresholding terms or exhibit irregular dynamics with instability properties. Third, when nodal activity is multi-dimensional, experimental access might be limited to a sub-dimension of the activity vector. The Koopman operator and time-delay embedding techniques are helpful for capturing the dynamical properties of sub-dimension observable systems. Yet, the problem remains unsolved for complex networked systems. Finally, the nodes in a complex system can have higher-order, beyond pairwise, couplings and such higher-order interactions
may significantly impact the dynamics of networked systems\textsuperscript{49,50}. Hence it is an interesting direction to extend the approach to inferring higher-order network dynamics.

Methods

Two-phase inference approach

The left side of Eq. (1) represents the time-varying derivative of each node’s activity, which can be numerically obtained from $x_i(t)$ through the five-point approximation\textsuperscript{51}

$$\dot{x}_i \approx \frac{x_{t-2\delta t} - 8x_{t-\delta t} + 8x_{t+\delta t} - x_{t+2\delta t}}{12\delta t},$$

where $\delta t$ is the time step. Hence the specific goal is to infer both the exact structure and the corresponding coefficients of the self-dynamics function $F(x_i(t))$ and the interaction dynamics function $G(x_i(t), x_j(t))$.

Because we lack a priori knowledge of the forms of $F$ and $G$, hence we construct two comprehensive libraries, $L_F$ and $L_G$, for self- and interaction dynamics respectively, including polynomial, trigonometric, exponential, fractional, rescaling, and various activation functions as listed in Supplementary Tables 1 and 2. Introducing the observed time series of nodes’ activities to the elementary functions in $L_F$ and $L_G$, we obtain two matrices $\Theta_F(t) = L_F(x_i(t))$ and $\Theta_G(t) = L_G(x_i(t), x_j(t))$ that describe the corresponding behaviors of these elementary functions (Supplementary Figure 1). To infer the compact forms that best match Eq. (2) we propose a two-phase approach.

Phase I, global regression. The purpose of this phase is to assess the relevance of each elementary function in $L_F$ and $L_G$ to the true, yet unknown, network dynamics. Given the observations of $x_i(t)$ for all $i$ at time $t$ we approximate the derivatives $\dot{x}(t)$ and calculate the matrices $\hat{\Theta}_F(t)$ and $\hat{\Theta}_G(t)$. These values are highly skewed and can span several orders of magnitude (Supplementary Figure 3) due to the skewness of node degrees and the nonlinearity of system dynamics, which could induce overestimation of the importance for inherently low-value constituents. To eliminate this severe effect it is crucial to normalize each column in $\dot{x}(t)$, $\hat{\Theta}_F(t)$ and $\hat{\Theta}_G(t)$. Then, the inference problem described by Eq. (2) is further recast to an optimization formula

$$\arg\min_{\xi_F, \xi_G} \int_0^T \left( \|\hat{\Theta}_F(t)\xi_F + \hat{\Theta}_G(t)\xi_G - \dot{x}(t)\|^2 \right) dt + \lambda (\|\xi_F\| + \|\xi_G\|),$$

where $\lambda > 0$ is a hyper-parameter that regulates the sparsity of coefficient vectors $\xi_F$ and $\xi_G$. We employ the regression analysis method of least absolute shrinkage and selection operator (i.e. lasso) to solve (6) and perform 5-fold validation to obtain the most appropriate value of $\lambda$ (see Supplementary Information Section I-B Algorithm 1). The resultant $\xi_F$ and $\xi_G$ capture the relevance of each elementary function in $L_F$ and $L_G$, enabling the identification of leading elementary functions that are most probably constituents of the true $F$ and $G$ (Fig. 1c). Consequently, Phase I is able to significantly narrow down the model space. However, the dynamical equation inferred by such regression alone lacks generative power. For instance, as shown in Fig. 1d, the trajectory generated by an inferred dynamical equation of Phase I deviates from that of the true network dynamics.

Phase II, local fine-tuning. In order to reconstruct generative and concise expressions for $F$ and $G$ we next perform fine-tuning in the reduced model space (see Supplementary Information Section I-B Algorithm 2). In contrast to Phase I we now use the original values of $\dot{x}(t)$, $\hat{\Theta}_F(t)$ and $\hat{\Theta}_G(t)$, i.e. without normalization, to further identify the necessary elementary functions and learn their precise coefficients.

Since spurious or missing links in the observed network topology have an adverse effect on the learning,
we perform topological sampling (see Methods) that imitates the feature of observed, usually incomplete, topologies. Another issue is to determine the minimal number of elementary functions required for reconstructing $\mathbf{F}$ and $\mathbf{G}$. To do so, we sequentially remove the elementary functions with smallest inferred coefficients and calculate, using a weighted version of Akaike’s information criterion (wAIC, see Methods), the information inconsistency between the observed nodes’ activities and the remaining set of elementary functions. This process stops when removing a certain elementary function consistently increases the value of wAIC. As shown in Fig. 1e, each curve in a plot at the left column represents the information inconsistency vs. model complexity for one topological sample. And we find that, indeed, the joint operation with wAIC and topological sampling are helpful for inference from noisy and incomplete data (Fig. 5(k,l)).

The final sets of elementary functions and their coefficients $\hat{\xi}_F$ and $\hat{\xi}_G$ compose the forms $\hat{\mathbf{F}}$ and $\hat{\mathbf{G}}$, leading to the successful inference of network dynamics described by Eq. (1). Indeed, as demonstrated in Fig. 1f, the trajectory generated by the inferred dynamical equation agrees well with the numerical simulations of the true network dynamics. It is worth noting that the ground truth, i.e. the form of the true equation, keeps unknown during the whole procedure and is only used to assess the accuracy of the final inferred results, hence our approach works in an autonomous unsupervised way.

wAIC

The original Akaike’s Information Criterion (AIC)\(^\text{52}\) is a frequently used method to balance the fitting and the complexity of a model with respect to the observed data, defined as $AIC = n \log MSE + 2p$, where $n$ is the number of observations, MSE is mean squared error of regression result of the model, and $p$ is the number of variables. By AIC, one aims to select an optimal model that best fits the observations with the fewest variables from the model candidates. However, we find that the original AIC does not work well in the inference problem we aim to solve in the present work. Hence we introduce a weighted version of AIC (i.e. wAIC),

$$
\text{wAIC} = \begin{cases} 
  w \cdot (n \log MSE + 2p), & (n \log MSE + 2p) \geq 0, \\
  (n \log MSE + 2p)/w, & (n \log MSE + 2p) < 0,
\end{cases}
$$

(7)

to balance the fitting accuracy and the model complexity, where $w$ is the inferred coefficient of a term from Phase I. A term with a larger $w$ inferred by Phase I is more likely to be able to capture the properties of the underlying unknown dynamics. Thus, multiplying $w$ or $1/w$ with AIC amplifies the impact of removing this term from the equation. The smaller the wAIC, the more consistent is the composition of the elementary functions with observed data and less important is this removed term.

To be specific, to evaluate the relevance of a term $i$, we remove this term from the equation inferred by Phase I and calculate the value of wAIC$_i$ of the new shorter equation (Supplementary Figure 2). We repeat this process to obtain the wAIC for each term. Then we sort these terms based on their wAIC values, and remove terms one by one with wAIC values from small to large. This operation gives a shorter equation at each step, and we calculate these shortened equations’ AIC values. The optimal equation is determined at the tuning point where the curve starts consistently increasing, as indicated by pink stars in Fig. 1e.

Topological sampling

We perform topological sampling in phase II as follows. We randomly choose $S$ nodes from all $n$ nodes, and obtain the activities of these $S$ nodes’ partial neighbors. Introducing the sampled ego structures and nodes’ activities into libraries $L_F$ and $L_G$ allows us to construct the self and interaction matrices $\hat{\Theta}_F$ and $\hat{\Theta}_G$, and to further distill the elementary functions as well as their coefficients. We repeat the process to
obtain $K$ sets of samples, average the coefficients of the elementary functions inferred from the $K$ sample sets. In the present work we set $S = 10$ and $K = 20$.

**sMAPE**

The inference inaccuracy is quantified by symmetric mean absolute percentage error (sMAPE)\(^5^3\),

$$
sMAPE = \frac{1}{m} \sum_{i=1}^{m} \frac{|I_i - R_i|}{(|I_i| + |R_i|)},
$$

where $m$ is the cardinal number of the set that contains both inferred and true elementary functions, $I_i$ and $R_i$ are inferred and true coefficients respectively. The range of sMAPE is $[0, 1]$. The more accurate the inferred equation, the lower the value of sMAPE. Note that if an inferred elementary function should not exist in the true equation or a true elementary function is not successfully inferred, the value of sMAPE will significantly increase. Therefore, the metric sMAPE captures not only the errors of inferred coefficients but also the incorrectness of the inferred equation form.

**Normalized Euclidean distance (NED)**

To evaluate the discrepancy between the inferred and true dynamics, we used the metric of normalized Euclidean distance (NED) that represents the distance between the two trajectories generated by the inferred and the true dynamical equations respectively. That is,

$$
NED(x_i, \hat{x}_i) = \frac{1}{D_{\text{max}}(T - t_0)} \sum_{t=t_0}^{T} \sqrt{(x_i(t) - \hat{x}_i(t))^2 + (\dot{x}_i(t) - \dot{\hat{x}}_i(t))^2}.
$$

Here $x_i$ is the true trajectory and $\hat{x}_i$ is the trajectory generated by the inferred equation, $t_0$ and $T$ are the beginning and ending times respectively, and $D_{\text{max}}$ is the longest Euclidean distance between a pair of points of true trajectory.

**Data availability**

The empirical networks data include *C. elegans* connectome\(^5^4–5^6\), the mushroom-body region of *Drosophila*\(^5^7\), the North-Europe power grid\(^5^8\), the U.S. power grid\(^5^9\), Advogato social network\(^6^0\) retrieved from website http://networkrepository.com, and the worldwide airline network retrieved from OpenFights data (https://openflights.org/data.html). The empirical data of epidemic spreading include daily reported numbers of H1N1 and SARS cases available at Kaggle (https://www.kaggle.com/lnunes/a-brief-comparative-study-of-epidemics/data), and the daily reported numbers of COVID-19 cases\(^6^1\). Source Data are available at the Code Ocean capsule\(^6^2\).

**Code availability**

All source codes are publicly available at the Code Ocean capsule\(^6^2\).

**Acknowledgements**

TTG and GY are supported by the National Key Research and Development Program of China (grant no. 2021ZD0204500), National Natural Science Foundation of China (grant no. 12161141016 and 11875043), Shanghai Municipal Science and Technology Major Project (grant no. 2021SHZDZX0100), Shanghai
Municipal Commission of Science and Technology Project (grant no. 18ZR1442000 and 19511132101), and Fundamental Research Funds for the Central Universities. The authors are also grateful for the helpful discussion with Prof. Baruch Barzel, Dr. Jack Moore, Dr. Xiaolei Ru and Mr. Tongyu Li.

Author contributions

GY conceived the research, GY and TTG designed the research, TTG performed the research, TTG and GY analyzed the results, GY and TTG wrote this manuscript.

Competing interests

The authors declare no competing interests.

References

[1] Grewe, B. F., Langer, D., Kasper, H., Kampa, B. M. & Helmchen, F. High-speed in vivo calcium imaging reveals neuronal network activity with near-millisecond precision. *Nature Methods* 7, 399–405 (2010).

[2] Stetter, O., Battaglia, D., Soriano, J. & Geisel, T. Model-free reconstruction of excitatory neuronal connectivity from calcium imaging signals. *PLoS Comput. Biol.* 8, e1002653 (2012).

[3] Reuter, J. A., Spacek, D. V. & Snyder, M. P. High-throughput sequencing technologies. *Mol. Cell.* 58, 586–597 (2015).

[4] Levy, S. E. & Myers, R. M. Advancements in next-generation sequencing. *Annu. Rev. Genomics Hum. Genet.* 17, 95–115 (2016).

[5] Colizza, V., Barrat, A., Barthélémy, M. & Vespignani, A. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proc. Natl. Acad. Sci. USA* 103, 2015–2020 (2006).

[6] Brockmann, D. & Helbing, D. The hidden geometry of complex, network-driven contagion phenomena. *Science* 342, 1337–1342 (2013).

[7] Chang, S. et al. Mobility network models of covid-19 explain inequities and inform reopening. *Nature* 589, 82–87 (2021).

[8] Newman, M., Barabási, A.-L. & Watts, D. J. *The Structure and Dynamics of Networks* (Princeton University Press, Princeton and Oxford, 2011).

[9] Barzel, B. & Barabási, A.-L. Universality in network dynamics. *Nature Phys.* 9, 673–681 (2013).

[10] Harush, U. & Barzel, B. Dynamic patterns of information flow in complex networks. *Nature Commun.* 8, 1–11 (2017).

[11] Stankovski, T., Pereira, T., McClintock, P. V. & Stefanovska, A. Coupling functions: universal insights into dynamical interaction mechanisms. *Rev. Mod. Phys.* 89, 045001 (2017).

[12] Breakspear, M. Dynamic models of large-scale brain activity. *Nature Neurosci.* 20, 340–352 (2017).

[13] Santolini, M. & Barabási, A.-L. Predicting perturbation patterns from the topology of biological networks. *Proc. Natl. Acad. Sci. USA* 115, E6375–E6383 (2018).
[14] Buldyrev, S. V., Parshani, R., Paul, G., Stanley, H. E. & Havlin, S. Catastrophic cascade of failures in interdependent networks. *Nature* **464**, 1025–1028 (2010).

[15] Yang, Y., Nishikawa, T. & Motter, A. E. Small vulnerable sets determine large network cascades in power grids. *Science* **358** (2017).

[16] Pastor-Satorras, R., Castellano, C., Van Mieghem, P. & Vespignani, A. Epidemic processes in complex networks. *Rev. Mod. Phys.* **87**, 925 (2015).

[17] Castellano, C., Fortunato, S. & Loreto, V. Statistical physics of social dynamics. *Rev. Mod. Phys.* **81**, 591–646 (2009).

[18] Becker, J., Brackbill, D. & Centola, D. Network dynamics of social influence in the wisdom of crowds. *Proc. Natl. Acad. Sci. USA* **114**, E5070–E5076 (2017).

[19] Arenas, A., Díaz-Guilera, A., Kurths, J., Moreno, Y. & Zhou, C. Synchronization in complex networks. *Phys. Rep.* **469**, 93–153 (2008).

[20] Barzel, B., Liu, Y.-Y. & Barabási, A.-L. Constructing minimal models for complex system dynamics. *Nature Commun.* **6**, 1–8 (2015).

[21] Schmidt, M. & Lipson, H. Distilling free-form natural laws from experimental data. *Science* **324**, 81–85 (2009).

[22] Wang, W.-X., Yang, R., Lai, Y.-C., Kovanis, V. & Grebogi, C. Predicting catastrophes in nonlinear dynamical systems by compressive sensing. *Phys. Rev. Lett.* **106**, 154101 (2011).

[23] Brunton, S. L., Proctor, J. L. & Kutz, J. N. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc. Natl. Acad. Sci. USA* **113**, 3932–3937 (2016).

[24] Rudy, S. H., Brunton, S. L., Proctor, J. L. & Kutz, J. N. Data-driven discovery of partial differential equations. *Sci. Adv.* **3**, e1602614 (2017).

[25] Udrescu, S.-M. & Tegmark, M. AI Feynman: a physics-inspired method for symbolic regression. *Sci. Adv.* **6**, eaay2631 (2020).

[26] Raissi, M. & Karniadakis, G. E. Hidden physics models: Machine learning of nonlinear partial differential equations. *J. Comput. Phys.* **357**, 125–141 (2018).

[27] Iten, R., Metger, T., Wilming, H., Del Rio, L. & Renner, R. Discovering physical concepts with neural networks. *Phys. Rev. Lett.* **124**, 010508 (2020).

[28] Frishman, A. & Ronceray, P. Learning force fields from stochastic trajectories. *Phys. Rev. X* **10**, 021009 (2020).

[29] Brückner, D. B., Ronceray, P. & Broedersz, C. P. Inferring the dynamics of underdamped stochastic systems. *Phys. Rev. Lett.* **125**, 058103 (2020).

[30] Shandilya, S. G. & Timme, M. Inferring network topology from complex dynamics. *New J. Phys.* **13**, 013004 (2011).

[31] Newman, M. E. J. Network structure from rich but noisy data. *Nature Phys.* **14**, 542–545 (2018).

[32] Rabinovich, M. I., Varona, P., Selverston, A. I. & Abarbanel, H. D. Dynamical principles in neuroscience. *Rev. Mod. Phys.* **78**, 1213 (2006).
[33] Marvel, S. A., Kleinberg, J., Kleinberg, R. D. & Strogatz, S. H. Continuous-time model of structural balance. *Proc. Natl. Acad. Sci. USA* **108**, 1771–1776 (2011).

[34] Strogatz, S. H. Exploring complex networks. *Nature* **410**, 268–276 (2001).

[35] Barahona, M. & Pecora, L. M. Synchronization in small-world systems. *Phys. Rev. Lett.* **89**, 054101 (2002).

[36] Mangan, N. M., Kutz, J. N., Brunton, S. L. & Proctor, J. L. Model selection for dynamical systems via sparse regression and information criteria. *Proc. Math. Phys. Eng. Sci.* **473**, 20170009 (2017).

[37] Casadiego, J., Nitzan, M., Hallerberg, S. & Timme, M. Model-free inference of direct network interactions from nonlinear collective dynamics. *Nature Commun.* **8**, 1–10 (2017).

[38] Runge, J., Nowack, P., Kretschmer, M., Flaxman, S. & Sejdinovic, D. Detecting and quantifying causal associations in large nonlinear time series datasets. *Sci. Adv.* **5**, eaau4996 (2019).

[39] Sugihara, G. *et al.* Detecting causality in complex ecosystems. *Science* **338**, 496–500 (2012).

[40] Sun, J., Taylor, D. & Bollt, E. M. Causal network inference by optimal causation entropy. *SIAM J. Appl. Dyn. Syst.* **14**, 73–106 (2015).

[41] Kralemann, B., Pikovsky, A. & Rosenblum, M. Reconstructing effective phase connectivity of oscillator networks from observations. *New J. Phys.* **16**, 085013 (2014).

[42] Frässle, S. *et al.* Regression dcm for fmri. *Neuroimage* **155**, 406–421 (2017).

[43] Gilson, M., Moreno-Bote, R., Ponce-Alvarez, A., Ritter, P. & Deco, G. Estimation of directed effective connectivity from fmri functional connectivity hints at asymmetries of cortical connectome. *PLoS Comput. Biol.* **12**, e1004762 (2016).

[44] Deco, G., Rolls, E. T. & Romo, R. Stochastic dynamics as a principle of brain function. *Prog. Neurobiol.* **88**, 1–16 (2009).

[45] Genkin, M., Hughes, O. & Engel, T. A. Learning non-stationary langevin dynamics from stochastic observations of latent trajectories. *Nature Commun.* **12**, 1–9 (2021).

[46] Zhao, H. Inferring the dynamics of black-box systems using a learning machine. *Sci. China: Phys. Mech. Astron.* **64**, 1–10 (2021).

[47] Jahnke, S., Memmesheimer, R.-M. & Timme, M. Stable irregular dynamics in complex neural networks. *Phys. Rev. Lett.* **100**, 048102 (2008).

[48] Champion, K. P., Brunton, S. L. & Kutz, J. N. Discovery of nonlinear multiscale systems: Sampling strategies and embeddings. *SIAM J. Appl. Dyn. Syst.* **18**, 312–333 (2019).

[49] Battiston, F. *et al.* The physics of higher-order interactions in complex systems. *Nature Phys.* **17**, 1093–1098 (2021).

[50] Lambiotte, R., Rosvall, M. & Scholtes, I. From networks to optimal higher-order models of complex systems. *Nature Phys.* **15**, 313–320 (2019).

[51] Sauer, T. Numerical solution of stochastic differential equations in finance. In *Handbook of computational finance*, 529–550 (Springer, 2012).

[52] Akaike, H. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* **19**, 716–723 (1974).
[53] Flores, B. E. A pragmatic view of accuracy measurement in forecasting. *Omega* **14**, 93–98 (1986).

[54] White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of the nematode caenorhabditis elegans. *Philos Trans R Soc Lond B Biol Sci* **314**, 1–340 (1986).

[55] Varshney, L. R., Chen, B. L., Paniagua, E., Hall, D. H. & Chklovskii, D. B. Structural properties of the *Caenorhabditis elegans* neuronal network. *PLoS Comput. Biol.* **7**, e1001066 (2011).

[56] Yan, G. *et al.* Network control principles predict neuron function in the caenorhabditis elegans connectome. *Nature* **550**, 519–523 (2017).

[57] Scheffer, L. K. *et al.* A connectome and analysis of the adult drosophila central brain. *Elife* **9**, e57443 (2020).

[58] Menck, P. J., Heitzig, J., Kurths, J. & Schellnhuber, H. J. How dead ends undermine power grid stability. *Nature communications* **5**, 1–8 (2014).

[59] Kunegis, J. Konect: the koblenz network collection. In *Proceedings of the 22nd international conference on world wide web*, 1343–1350 (2013).

[60] Rossi, R. & Ahmed, N. The network data repository with interactive graph analytics and visualization. In *Twenty-ninth AAAI conference on artificial intelligence* (2015). http://networkrepository.com.

[61] Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track covid-19 in real time. *Lancet Infect. Dis.* **20**, 533–534 (2020).

[62] Gao, T.-T. & Yan, G. A two-phase approach for inferring complex network dynamics. (Code Ocean, 2022) https://doi.org/10.24433/C0.4774495.v1.
Figure 1: Overview of the two-phase inference approach. 

a. Observation data of network topology $A_{i,j}$, including spurious and missing links, and low-resolution and noisy data of nodal activities $x_i(t)$. 

b. Mapping the normalized observation data into two matrices $\Theta_F$ and $\Theta_G$ that represent the time-varying patterns of elementary functions.

c. Phase I that narrows down the model space by identifying several leading elementary functions through global regression for each dimension of $\dot{x}_i(t)$.

d. Comparison of trajectories generated respectively by the true network dynamics and by the dynamical equation inferred by Phase I alone.

e. Phase II that performs local fine-tuning, by using topological sampling and weighted Akaike’s information criterion (wAIC), to further determine the optimal number (indicated by pink stars) of elementary functions for $\hat{F}(x_i(t))$ and $\hat{G}(x_i(t), x_j(t))$.

f. Comparison of trajectories generated respectively by the true and the inferred dynamical equations. The example illustrated in (c-f) is Hindmarsh-Rose neuronal dynamics on a directed BA network with size $n = 100$, average degree $\langle k \rangle = 5$. 

\[ \frac{dx_i(t)}{dt} = F(x_i(t)) + \sum_{j=1}^{N} A_{ij} G(x_i(t), x_j(t)) \]
**Figure 2: Inferring FitzHugh-Nagumo neuronal network dynamics on synthetic and real topologies.**

**a.** True FitzHugh-Nagumo (FHN) dynamics used to simulate nodes’ activities data on various topologies. $F_d$ and $G_d$ are self- and interaction dynamics of the $d$-th dimension respectively, $x_{i,d}$ is $d$-th dimension’s state of node $i$, and $x_{p,i,d}$ is the polynomial with order $p$.

**b.** Ten leading elementary functions identified by Phase I for each dimension.

**c.** The necessary elementary functions and their coefficients further inferred through Phase II on two synthetic (directed ER and undirected SF) and one empirical (Drosophila mushroom body) networks, where $g^{\text{FHN}}_F$ denotes the term $(x_j - x_i)/k^{\text{FHN}}$. 

**d.** Relative errors $\Delta$ of the inferred elementary functions and their coefficients.

Note that the elementary functions ruled out from $\Theta_F$ and $\Theta_G$ by our approach (i.e. whose coefficients are inferred as zero) are not shown.

**e,f.** Nodes’ activities and trajectories generated respectively by the true and the inferred equations.
Figure 3: Inference accuracy for other four typical nonlinear network dynamics. a,b. Similar to Fig. 2 but for inferring HR neuronal dynamics, where the interaction dynamics $G(x_i, x_j)$ are composed of $g_{HR}^1 \equiv 1/(1 + e^{10(x_j - x_i)})$ and $g_{HR}^2 \equiv x_i/(1 + e^{10(x_j - x_i)})$. c,d. Relative errors and six edges’ activities of the inferred edge dynamics of social balance. e-i. Relative errors of the inferred effective equations for network dynamics of Kuramoto model and coupled Rössler oscillators. In both cases the self-dynamics are heterogeneous, i.e. the intrinsic frequency of each node is not identical but follows a normal distribution $N(1, \sigma)$ with $\sigma = 0.1$. Grey curves represent the activity of individual nodes and black curves represents the averaging activity of systems. Symbols $g_{kura}$ and $g_{ross}$ denote terms $\sin(x_j - x_i)$ and $(x_j - x_i)$ respectively. The details of these dynamics and empirical networks are shown in Supplementary Tables 3 and 4.
Figure 4: Inferrability of network dynamics. a. Inference inaccuracy represented by sMAPE and synchronization represented by order parameter $\langle R \rangle$ vs. coupling strength between nodes. b. Inaccuracy of inferred effective equation for Kuramoto network dynamics where the natural frequency $\omega$ of each node follows a normal distribution $\mathcal{N}(1, \sigma)$. Larger $\sigma$ indicates higher dynamical heterogeneity. c. Normalized Euclidean distance (NED, see Methods) when some true elementary functions were deliberately removed from libraries $L_F$ and $L_G$. The error bars represent the mean ± the standard deviation, and the sample size is 100. The networks are scale-free with size $n = 100$ and average degree $\langle k \rangle = 5.0$. Simulation details are shown in Supplementary Table 4.
Figure 5: Inference robustness against incompleteness and noises. a-e. Inference inaccuracies sMAPE when the nodes’ activities data are low-resolution, have dynamical noises (Gaussian white noise with intensity $\eta$) or observational noises (intensity quantified by signal-to-noise ratio SNR) or when the topologies data have spurious and missing links. f-j. Comparisons of inference inaccuracies between SINDy, ARNI and our approach for inferring HR neuronal network dynamics, with varying amount of time-points, observational noise, correlated dynamical noise, missing links, and different network sizes. Simulation details are shown in Supplementary Table 4. k, l. Comparison results of ablation studies. The box-whisker plots are visualized with the Tukey method (i.e. the box represents the interquartile range (IQR) and the line in the box shows the median, with whiskers that extend 1.5 times the IQR from the box edges; The outliers are also plotted) and the sample size is 20. Five ablation studies were performed: ① removing topological sampling, ② using original AIC instead of wAIC, ③ removing Phase II, ④ removing Phase I, or ⑤ without normalization to $\Theta_F$ and $\Theta_G$. Statistical significance is obtained through multiple Mann-Whitney tests. Three or four asterisks indicate $p$-value $< 10^{-3}$ or $10^{-4}$, and n.s. means not significant.
Figure 6: Inference of early spreading dynamics from empirical data. 

a. Worldwide airline network (partial) used for the inference. Each node represents a country or region, and line thickness represents the amount of passenger flow. The form describes the dynamical equation inferred by the two-phase approach. 

b-e. Comparisons between the empirical cumulative number of H1N1 cases for different nodes (dashed lines) and the cumulative number generated by the inferred equation (solid lines). For better visualization, the comparisons are displayed in four plots (from b to e) and the dates when the first case is reported for each node are all shifted to the first day ($t = 1$). 

f-i. Comparisons between the empirical and the inferred cumulative numbers in different nodes for SARS (f, g) and COVID-19 (h, i).
Supplementary Information

Autonomous inference of complex network dynamics from incomplete and noisy data

Contents

I Approach 21
   A Function libraries .............................................................. 21
   B Pseudocode .............................................................. 21

II Networks 24
   A Synthetic networks .............................................................. 25
   B Empirical networks .............................................................. 25

III Dynamics and inferred results 26
   A Brain dynamics .............................................................. 26
   B Gene regulation dynamics .............................................................. 28
   C Coupled nonlinear oscillators .............................................................. 28
   D Edge dynamics of social balance .............................................................. 33
   E Reproducibility description .............................................................. 34

IV Details of inferrability 34
   A Quantifying synchronization .............................................................. 34
   B Inferring alternative equations .............................................................. 35
   C Inferrability comparison studies .............................................................. 36

V Details of robustness analyses 36
   A Noisy data .............................................................. 37
   B Ablation studies .............................................................. 38
   C Robustness comparison studies .............................................................. 39

VI Details of empirical system inference 42
I Approach

A Function libraries

We construct two comprehensive libraries, \( L_F \) and \( L_G \), for self- and interaction dynamics respectively, including polynomial, trigonometric, exponential, fractional, rescaling, and various activation functions as listed in Supplementary Tables 1 and 2.

### Supplementary Table 1: Library \( L_F \) for self-dynamics

| Functions          | \( \Theta_F = L_F(x_i) \) |
|--------------------|---------------------------|
| Polynomial         | \( x_i, x_i^2, x_i^3, \ldots \) |
| Trigonometric      | \( \sin(x_i), \cos(x_i), \tan(x_i) \) |
| Exponential        | \( e^{x_i} \) |
| Fractional         | \( \frac{1}{x_i} \) |
| Activation         | \{ \text{sigmoid}(x_i) \}^\dagger, \tan(h(x_i)), \{ \frac{x_i}{x_i+1} \}^\dagger |
| Rescaling          | \( \frac{x_i}{k_i^{10}} \) |

### Supplementary Table 2: Library \( L_G \) for interaction dynamics

| Functions          | \( \Theta_G = L_G(x_i) \) | \( \Theta_G = L_G(x_i x_j) \) | \( \Theta_G = L_G(x_j - x_i) \) | \( \Theta_G = x_i L_G(x_j) \) |
|--------------------|---------------------|---------------------|---------------------|---------------------|
| Polynomial         | \( x_j, x_j^2, \ldots \) \( x_i x_j, (x_i x_j)^2, \ldots \) \( x_j - x_i, (x_j - x_i)^2, \ldots \) |
| Trigonometric      | \( \sin(x_j), \cos(x_j), \ldots \) \( \sin(x_i x_j), \cos(x_i x_j), \ldots \) \( \sin(x_j - x_i), \cos(x_j - x_i), \ldots \) \( x_i \sin(x_j), x_i \cos(x_j), \ldots \) |
| Exponential        | \( e^{x_i} \) \( e^{x_i x_j} \) \( e^{x_j - x_i} \) \( x_i e^{x_j} \) |
| Fractional         | \( \frac{1}{x_i} \) \( \frac{1}{x_i x_j} \) \( \frac{1}{x_j - x_i} \) \( \frac{x_i}{x_j} \) |
| Activation         | \( \{ \text{sigmoid}(x_j) \}^\dagger, \tan(h(x_j)) \) \( \{ \text{sigmoid}(x_i x_j) \}^\dagger, \tan(h(x_i x_j)) \) \( \{ \text{sigmoid}(x_j - x_i) \}^\dagger, \tan(h(x_j - x_i)) \) \( \{ x_i, \text{sigmoid}(x_j) \}^\dagger, x_i \tan(h(x_j)) \) |
| Rescaling          | \( \frac{x_i}{k_i^{10}} \) \( \frac{x_i x_j}{k_i^{10}} \) \( \frac{x_j - x_i}{k_i^{10}} \) \( \frac{x_i x_j}{k_i^{10}} \) |

B Pseudocode

Phase I. As shown in Supplementary Figure 3, the values of the terms (i.e. elementary functions) in libraries \( L_F \) and \( L_G \) can span several orders of magnitude. Such a wide spread has detrimental effects on evaluating the relevance of each term to the true unknown dynamics. For example, if a term is inherently low-valued, a regression method will probably assign a relatively large coefficient to it, which might indicate the term is more relevant. However, this is not always true since it could be caused by the fact that the scale of this term is inherently smaller than that of other terms. To remove such detrimental effect, we hence do not use the original data but rather perform normalization for the derivatives \( \dot{x} \) and each column of the matrices \( \Theta_F \) and \( \Theta_G \). To be specific, the normalization of the \( i \)-th column of the matrix \( \Theta \) is

\[
\|\Theta_i\| = \frac{\Theta_i}{\sum_{k=1}^{n \times T} \Theta_{i,k}}, \tag{S10}
\]

\(^1\text{sigmoid: indicates a variety of sigmoid functions } \frac{1}{1 + e^{-\alpha(x-\beta)}}, \text{ where } \alpha \in [1 : 1 : 10] \text{ and } \beta \in [0 : 1 : 10].
\]

\(^2\gamma \in [0 : 1 : 10].\)
Supplementary Figure 1: An example for the construction of matrices $\Theta_F$ and $\Theta_G$. a. A network, and the adjacency vector $A_{i=3}$ of node 3. b. Constructing self-dynamics matrix $\Theta_{F,i=3}$ by mapping $x_3(t)$ to library $L_F$ that contains $m_1$ elementary functions. c. Constructing interaction dynamics matrix $\Theta_{G,i=3}$ by mapping $x_3(t)$ and $\{x_j(t)\}$, where $j$ is in the set of node $i$’s neighbors, to library $L_G$ that contains $m_2$ interaction elementary functions. Through the Kronecker product $I_d$ of the identity matrix and $\Theta_{F,i=3}$ (or $\Theta_{G,i=3}$), we obtain matrices $\tilde{\Theta}_{F,i=3}$ and $\tilde{\Theta}_{G,i=3}$ capturing the whole $d$-dimensional nodes’ activities on all elementary functions. Hence the optimization is to search for $\xi_{F,i=3}$ and $\xi_{G,i=3}$ that best solve the equation exhibited in d.

Supplementary Figure 2: An example for the calculation of wAIC. To evaluate the relevance of the term $x_{1,1}$ to the true dynamics, we remove this term from the set of leading terms inferred by Phase I and then calculate the MSE$_1$ between the dynamics composed of the remaining $p$ terms and the observation data. Then we obtain the value of wAIC$_1$ that represents the relevance of this removed term (the higher the more relevant).

and the normalization of the $d$-th dimension of the derivative is

$$\|\dot{x}_d\| = \frac{\dot{x}_d}{\sum_{k=1}^{n \times T} \dot{x}_{d,k}}.$$

Therefore Phase I aims to obtain the coefficients $\xi_m$ in

$$\dot{x}_d = \sum_{m=1}^{m_1+m_2} \Theta_m (\xi_m \frac{\sum_{k=1}^{n \times T} \dot{x}_{d,k}}{\sum_{k=1}^{n \times T} \Theta_{m,k}}),$$

where $m_1$ and $m_2$ are the numbers of terms in libraries $L_F$ and $L_G$ respectively. The pseudocode of
Phase I is described in Algorithm 1.

Algorithm 1 Phase I

Input: $A$: network topology
$x$: numerical derivatives
$\tilde{\Theta}_F, \tilde{\Theta}_G$: libraries’ matrices, for self- and interaction dynamics respectively
$k$: the number of leading terms to be identified

Output: $L^*_F, L^*_G$: reduced libraries for self- and interaction dynamics respectively
$w$: weight vector

1: $\#$ data normalization
2: normalizing of each column of $\dot{x}$, $\tilde{\Theta}_F$ and $\tilde{\Theta}_G$;
3: $\#$ global regression
4: minimizing of formula (3) in main paper via lasso to obtain $\xi^*_F$, $\xi^*_G$, and intercept value $c$;
5: $\#$ Judging whether the constant term $c$ should exist
6: Temporally construct two reduced libraries: one contains only the $k$ leading terms with the largest absolute coefficients $\xi^*$, the other contains the $k$ leading terms and the constant term;
7: compute $\text{AIC}_{w.c.}$ and $\text{AIC}_{w.o.c.}$ for libraries with and without the constant term respectively.
8: if $\text{AIC}_{w.c.} - \text{AIC}_{w.o.c.} > \beta_1 \text{AIC}_{w.c.}$ $\# \beta_1$: a threshold for the judgement
9: the constant term is necessary;
10: else
11: the constant term is unnecessary;
12: end

By Phase I, the comprehensive libraries containing a large number of elementary functions that potentially compose the self- and interaction dynamics are significantly narrowed down to $L^*_F$ and $L^*_G$. Because, for most systems of interest, the underlying dynamical equation often consists of only a few terms, the reduced libraries $L^*_F$ and $L^*_G$ contain top $k$ terms with the largest absolute coefficients described by the weight vector $w$. To determine whether the hidden equation should contain a constant term, we compare the AIC values of the equations with and without constant. If the AIC value of the equation with constant is far smaller than that without constant, the hidden equation probably has a constant term. In the present paper we set $k = 10$, i.e. obtain top ten leading elementary functions for $L^*_F$ and $L^*_G$ by algorithm I.

Note that the hyper-parameter $\lambda$ in Phase I is automatically selected by 5-fold cross validation, a widely-
used trick to estimate the most appropriate value of hyper-parameter. Specifically, to choose the best \( \lambda \) from a set \( \{ \lambda_1, \lambda_2, \cdots, \lambda_r \} \), we parameterize the loss function \( (L(\lambda_1), L(\lambda_2), \cdots, L(\lambda_r)) \), and randomly divide the nodes’ activities data into five sets, with one set for test and the other four sets for training. We apply each value of \( \lambda \) and calculate the corresponding score of lasso regression. Finally the hyper-parameter with the highest regression score is selected as the optimal value of \( \lambda \).

**Phase II.** Through Phase I we have obtained a reduced model space that represents the set of most relevant elementary functions. With the reduced libraries \( L^*_F, L^*_G \) at hand, next we propose Phase II to infer a concise form as well as appropriate coefficients for the hidden dynamical equation. The procedure of Phase II has two key sub-steps, *i.e.* topological sampling, and fine-tuning with wAIC (see Methods). The pseudocode of Phase II is described in Algorithm 2.

### Algorithm 2 Phase II

**Input:** \( \dot{x} \): numerical derivatives  
\( L^*_F, L^*_G \): the reduced libraries by Phase I  
\( w \): weight vector obtained by Phase I  

**Output:** \( \hat{L}_F, \hat{L}_G \): the final inferred terms for self- and interaction dynamics respectively  
\( \hat{\xi}_F, \hat{\xi}_G \): the final inferred coefficients for self- and interaction dynamics respectively

1: extract the reduced matrices \( \hat{\Theta}_F \) and \( \hat{\Theta}_G \) from \( \tilde{\Theta}_F \) and \( \tilde{\Theta}_G \) based on \( L^*_F \) and \( L^*_G \);
2: \# topological sampling
3: for \( i = 1; i < K; i ++ \)  
4: randomly chose \( S \) nodes  
5: merge the numerical derivatives and libraries matrices of \( S \) nodes: \( \dot{x}^S, \hat{\Theta}_F^S \) and \( \hat{\Theta}_G^S \);
6: \# assessment with wAIC
7: for \( j = 1; j < \kappa; j ++ \)  
8: remove the term \( j \)
9: compute AIC \( j \) for the combination of terms without term \( j \);
10: \[ w\text{AIC}_j \leftarrow w_j \cdot \text{AIC}_j \];
11: end
12: for \( l = 1; l < \kappa; l ++ \)  
13: remove \( l \) terms that have smallest wAIC value;
14: compute AIC \( \kappa-l \) for the combination of all the remaining terms;
15: if AIC \( \kappa-l-1 - \text{AIC}_\kappa-l > \beta_2 \)  
16: obtain the inferred elementary functions \( \hat{L}_{F,i}, \hat{L}_{G,i} \) and their coefficients of the \( i \)-th topological sample by deleting \( l \) terms;
17: end
18: end
19: end
20: obtain \( \hat{L}_F \) and \( \hat{L}_G \) by merging the identified terms of all \( K \) topological samples;
21: obtain \( \hat{\xi}_F \) and \( \hat{\xi}_G \) by averaging the inferred coefficients across all \( K \) topological samples;

**II Networks**

In the present paper we used both synthetic and empirical networks to validate our two-phase inference approach. The synthetic networks are generated by directed Erdős-Rényi (ER) model\[^{S2}\] for random topologies, as well as Barabási-Albert (BA)\[^{S3}\] and the static model\[^{S4}\] for scale-free (SF) topologies. The empirical network datasets are obtained from neuronal, social and technological domains.
A Synthetic networks

Erdős-Rényi (ER) networks are generated by `erdos_renyi_graph(n, p)` in `networkX` [S5], where \( n = 100, \ p = 0.05 \). The resulting average degree \( \langle k \rangle = 5.2 \). For each link between node \( i \) and \( j \), we assign the link direction, i.e. from \( i \) to \( j \), from \( j \) to \( i \), or a reciprocal connection between them, with equal probability.

Directed SF networks are generated by Barabási-Albert model, i.e. using `barabasi_albert_graph (n, m)` in `networkX` [S5], where \( n = 100, m = 5 \). We set each link bidirectional, and then randomly delete a proportion of these unidirectional links while ensuring that the network is weakly-connected. Finally we obtain a directed network with average total-degree \( \langle k \rangle = 5.1 \).

Directed scale-free networks are generated by the static model [S4]. The procedure is as follows: we assign a weight \( \omega_i = i^{-\eta} \) to the node with index \( i \), where \( i = 1, 2, \cdots, n \) and \( 0 < \eta \leq 1 \), then randomly select a pair of nodes \( i \) and \( j \) (\( i \neq j \)) with probability proportional to \( \omega_i \) and \( \omega_j \) respectively. If there the link from node \( i \) to \( j \) already exists, we randomly select another pair of nodes with proportional to their weights; Otherwise, we connect node \( i \) to \( j \). Repeat this process until the number of links reaches \( mn \). Hence the average total-degree \( \langle k \rangle = 2m \), and the power-law exponent \( \gamma = 1 + \frac{\eta}{\eta} \), which can be tuned by \( \eta \).

B Empirical networks

The empirical networks used in this work include `C. elegans` connectome with 279 neurons [S6]–[S8], the mushroom-body region of `Drosophila` with 1832 neurons [S9], the North-Europe (N.E.) power grid with 236 nodes [S10], the U.S. power grid with 4941 nodes [S11] and a community in the weighted social network of Advogato with 623 nodes [S12]. The properties of these empirical networks are described in Supplementary Table (3).

### Supplementary Table 3: Synthetic and empirical networks used in the paper. For each network, we show its type, name and reference, number of nodes \( n \), directed or undirected, average degree \( \langle k \rangle \), maximum \( (k_{\max}) \) and minimal \( (k_{\min}) \) degree, and degree heterogeneity \( \langle k^2 \rangle / \langle k \rangle^2 \).

| Type       | Name                        | \( n \) | Dir/Undir | \( \langle K \rangle \) | \( k_{\max} \) | \( k_{\min} \) | \( \langle k^2 \rangle / \langle k \rangle^2 \) |
|------------|-----------------------------|---------|-----------|--------------------------|----------------|----------------|-----------------------------------------------|
| Synthetic  | SF (BA) [S3]                | 100     | Dir       | 5.1                      | 40             | 3              | 1.5                                           |
|            | SF, (\( \gamma = 2.5 \))   | 100     | Undir     | 2.6                      | 34             | 1              | 4.4                                           |
|            | ER [S2]                     | 100     | Dir       | 5.2                      | 21             | 4              | 1.1                                           |
| Empirical  | `C. elegans` [S6]–[S8]*     | 279     | Dir       | 10.7                     | 137            | 2              | 1.7                                           |
|            | `Drosophila` [S9]**         | 1832    | Dir       | 6.6                      | 56             | 1              | 5.4                                           |
|            | North-Europe power grid [S10]| 236   | Undir     | 2.7                      | 13             | 1              | 1.2                                           |
|            | U.S. power grid [S11]       | 4941    | Undir     | 2.7                      | 19             | 1              | 1.5                                           |
|            | Advogato [S12]***           | 623     | Dir       | 6.0                      | 251            | 1              | 4.4                                           |

* Neuronal connectome of the nematode worm `Caenorhabditis elegans`, which is defined anatomically at a cellular network as 2990 synaptic connections between 279 neurons.

** The cellular-level connectome of `Drosophila` mushroom-body region that contains neurons, which project their axons within tracts resembling pairs of mushrooms, accessed via https://neuprint-examples.janelia.org.

*** A subgraph of the directed weighted network of trust among developers in Advogato platform. The edges can have positive or negative weights representing the amount of trust or distrust. We divide the whole network into communities and use a large community in this work.
III Dynamics and inferred results

In this paper we demonstrated the effectiveness of our two-phase approach in inferring a wide range of complex network dynamics pertaining to brain, social and coupled nonlinear oscillators systems.

A Brain dynamics

Hindmarsh-Rose dynamics

Neurons exhibit spiking activities which is believed to be an essential component in information processing in brains\[^{S13}\]. To simulate such activities we employed the 3-dimensional Hindmarsh-Rose (HR) model\[^{S14}\], a moderately simplified version of the Hodgkin-Huxley\[^{S13}\] model. The true equations governing the HR network dynamics are

\[
\begin{align*}
\frac{dx_{i,1}}{dt} &= x_{i,2} - ax_{i,1}^3 + bx_{i,1}^2 - x_{i,3} + I_{\text{ext}} + \epsilon(V_{\text{syn}} - x_{i,1}) \sum_{j=1}^{N} A_{ij} \mu(x_{j,1}), \\
\frac{dx_{i,2}}{dt} &= c - u x_{i,1}^2 - x_{i,2}, \\
\frac{dx_{i,3}}{dt} &= r[s(x_{i,1} - x_0) - x_{i,3}],
\end{align*}
\]

where the coupling term

\[
\mu(x_{j,1}) = \frac{1}{1 + e^{-\lambda(x_{j,1} - \Omega_{\text{syn}})}}.
\]

Here $x_{i,1}$ is the membrane potential of neuron $i$, $x_{i,2}$ is the transport rate of sodium and potassium ions across the membrane through the ion channels, and $x_{i,3}$ is the adaptation current. The $x_{i,1}$ is a simplified notation for $x_{i,1}(t)$. Parameters $a = 1$, $b = 3$, $c = 1$, $u = 5$, $s = 4$, $r = 0.005$, $x_0 = -1.6$, coupling strength $\epsilon = 0.15$, $V_{\text{syn}} = 2$, $\lambda = 10$, $\Omega_{\text{syn}} = 1$, and $I_{\text{ext}}$ is external current which is set to 3.24 for all neurons. The coupling strength $\epsilon$ is tunable according to size and topology of networks (see sec. IV).

Applying our approach to the neuronal activities data generated by HR dynamics on a directed scale-free network, we obtain the inferred equations typically as

\[
\begin{align*}
\frac{\hat{dx}_{i,1}}{dt} &= 0.994x_{i,2} - 0.993x_{i,1}^3 + 2.977x_{i,1}^2 - 0.977x_{i,3} + 3.218 \\
&\quad + \sum_{j=1}^{N} A_{ij}(0.296g_1^{HR}(x_{j,1}) - 0.146g_2^{HR}(x_{i,1}, x_{j,1})) \\
\frac{\hat{dx}_{i,2}}{dt} &= 1.000 - 4.994x_{i,1}^2 - 0.999x_{i,2}, \\
\frac{\hat{dx}_{i,3}}{dt} &= 0.032 + 0.020x_{i,1} - 0.005x_{i,3},
\end{align*}
\]

where $g_1^{HR} = \frac{1}{1 + e^{-10(x_{i,1} - x_0)}}$ and $g_2^{HR} = \frac{x_{i,1}}{1 + e^{-10(x_{i,1} - x_0)}}$. Note that all true elementary functions have been successfully inferred, and inaccuracies of their coefficients are lower than 3% (as also shown in Fig. 3a in the main paper). The neuronal activities and trajectories generated by the true Eq. (S12, S13) and the inferred Eq. (S14) are shown in Supplementary Fig. 4.

FitzHugh-Nagumo dynamics

We also tested our approach by applying it to the neuronal activities data generated by FitzHugh-Nagumo (FHN) dynamics\[^{S15}\] on various networks. The equations governing the FHN neuronal network dynamics
The FHN dynamics capture the firing behaviors of neurons with two components. The first component represents the membrane potential containing self- and interaction dynamics, where $k_i^{in}$ is the in-degree of neuron $i$, and $\epsilon = 1$. The second component represents a recovery variable where $a = 0.28$, $b = 0.5$, and $c = -0.04$.

The equations inferred by our approach from the neuronal activities data generated on a directed scale-free network are shown as Eq. (S16). The activities and trajectories generated by the true Eq. (S15) and the inferred Eq. (S16) are shown in Supplementary Fig. 5.

$$
\left\{ \begin{array}{l}
\frac{dx_{i,1}}{dt} = x_{i,1} - x_{i,1}^3 - x_{i,2} - \epsilon \sum_{j=1}^{N} A_{ij} \frac{(x_{j,1} - x_{i,1})}{k_i^{in}}, \\
\frac{dx_{i,2}}{dt} = a + bx_{i,1} + cx_{i,2}. 
\end{array} \right.
$$

(S15)

$$
\left\{ \begin{array}{l}
\frac{dx_{i,1}}{dt} = 0.989x_{i,1} - 0.993x_{i,1}^3 - 0.996x_{i,2} - 0.996 \sum_{j=1}^{N} A_{ij} \frac{(x_{j,1} - x_{i,1})}{k_i^{in}}, \\
\frac{dx_{i,2}}{dt} = 0.279 + 0.499x_{i,1} - 0.040x_{i,2}. 
\end{array} \right.
$$

(S16)
Supplementary Figure 5: Neuronal activities and trajectories generated by the true FHN network dynamics and the inferred dynamical equations.

B Gene regulation dynamics

We also demonstrated the effectiveness of our approach in inferring gene regulation (GR) dynamics. We generate the gene activities data according to the equation\cite{S16}

\[
\frac{dx_{i,1}}{dt} = bx_{i,1} + \epsilon_{ij} \sum_{j=1}^{N} A_{i,j} \frac{x_{j,1}^{\alpha} x_{j,1}}{x_{j,1} + 1}, \tag{S17}
\]

where \(x_{i,1}(t)\) is the concentration of gene \(i\) at time \(t\), \(\alpha = 2\) denoting the Hill coefficient, and \(b = -0.2\). Parameters \(\epsilon_{ij}\) denotes the regulation strength of gene \(j\) on gene \(i\) which is set as \(\epsilon_{ij} = 0.1\) or 0.05 on different networks. The dynamical interaction equation inferred by our approach from the data generated on an undirected scale-free network with \(\epsilon_{ij} = 0.05\) is

\[
\frac{dx_{i,1}}{dt} = -0.197x_{i,1} + 0.049 \sum_{j=1}^{N} A_{i,j} \frac{x_{j,1}^2}{x_{j,1}^2 + 1}. \tag{S18}
\]

The interaction function is automatically inferred from library \(L_G\) that contains activation function \(\{x_{j,1}^2/(x_{j,1}^2 + 1)\}\). The activities and trajectories generated by the true and the inferred equations are shown in Supplementary Fig. 6.

C Coupled nonlinear oscillators

We validated the two-phase inference approach for two types of coupled nonlinear oscillators, in which the self-dynamics are heterogeneous.
**Supplementary Figure 6:** a. True and inferred activities for 10 typical genes. b. The distribution of the discrepancy between true and inferred trajectories.

**Coupled Kuramoto dynamics**

The first one is Kuramoto network dynamics\[^{[S17],[S18]}\]:

\[
\frac{dx_{i,1}}{dt} = \omega_i + \epsilon \sum_{j=1}^{N} A_{ij} \sin(x_{j,1} - x_{i,1}) \tag{S19}
\]

It describes the dynamics of \( n \) oscillators with natural frequencies following a normal distribution \( \omega \sim \mathcal{N}(1, \sigma) \) with \( \sigma = 0.1 \), and the oscillators are coupled via the links \( A_{ij} \). Parameter \( \epsilon \) denotes the coupling strength of the interactions.

The equation inferred by our approach from the activities of oscillators generated on a directed scale-free network with \( \epsilon = 0.025 \) is

\[
\frac{d\hat{x}_{i,1}}{dt} = 1.027 + 0.025 \sum_{j=1}^{N} A_{ij} \sin(x_{j,1} - x_{i,1}), \tag{S20}
\]

where constant 1.027 is the inferred effective frequency \( \hat{\omega} \) (i.e. the average value \( \langle \omega \rangle \) of nodes’ natural frequencies). The activities and trajectories generated by the inferred equation are shown in Supplementary Fig. 7.

**Supplementary Figure 7:** a. Activities denoted by \( \sin x_{i,1} \), generated by the true and the inferred equations. Light blue curves show 10 typical individuals’ trajectories in the network, while the dark blue curve is the average of all nodes’ trajectories. Dotted curve is the inferred activity. b. The distribution of the discrepancy between the true and inferred trajectories.
Coupled Rössler dynamics

Another case for coupled nonlinear oscillator system is Rössler network dynamics. We first generated chaotic activities data according to the true equations governing heterogeneous Rössler dynamics

\[
\begin{align*}
\frac{dx_{i,1}}{dt} &= -\omega_i x_{i,2} - x_{i,3} + \epsilon \sum_{j=1}^{n} A_{ij} (x_{j,1} - x_{i,1}), \\
\frac{dx_{i,2}}{dt} &= \omega_i x_{i,1} + ax_{i,2}, \\
\frac{dx_{i,3}}{dt} &= b + x_{i,3}(x_{i,1} + c),
\end{align*}
\]  

(S21)

where coupling strength \( \epsilon = 0.15 \) or 0.2 on different networks, \( a = 0.2 \), \( b = 0.2 \), and \( c = -6 \). Oscillators’ natural frequencies follow a normal distribution \( \omega \sim \mathcal{N}(1, \sigma) \) with \( \sigma = 0.1 \). The equations inferred by our approach from the data generated on a directed scale-free network with \( \epsilon = 0.15 \) are

\[
\begin{align*}
\frac{d\hat{x}_{i,1}}{dt} &= -1.011x_{i,2} - 1.009x_{i,3} + 0.126 \sum_{j=1}^{n} A_{ij} (x_{j,1} - x_{i,1}), \\
\frac{d\hat{x}_{i,2}}{dt} &= 1.019x_{i,1} + 0.197x_{i,2}, \\
\frac{d\hat{x}_{i,3}}{dt} &= 0.196 + 0.993x_{i,3}x_{i,1} - 5.653x_{i,3},
\end{align*}
\]  

(S22)

where the coefficient 1.011 is the inferred effective frequency \( \hat{\omega} \) (i.e. the average value \( \langle \omega \rangle \) of nodes’ natural frequencies). The activities and trajectories generated by the true and the inferred dynamical equations are shown in Supplementary Fig. 8.

We also generated homogeneous Rössler dynamics according to the true equations

\[
\begin{align*}
\frac{dx_{i,1}}{dt} &= -x_{i,2} - x_{i,3} + \epsilon \sum_{j=1}^{n} A_{ij} (x_{j,1} - x_{i,1}), \\
\frac{dx_{i,2}}{dt} &= x_{i,1} + ax_{i,2}, \\
\frac{dx_{i,3}}{dt} &= b + x_{i,3}(x_{i,1} + c),
\end{align*}
\]  

(S23)

where \( \epsilon = 0.1 \), \( a = 0.35 \), \( b = 0.2 \) and \( c = -5.7 \). The equations inferred by our approach from the data generated on a directed scale-free network are

\[
\begin{align*}
\frac{d\hat{x}_{i,1}}{dt} &= -0.999x_{i,2} - 0.999x_{i,3} + 0.100 \sum_{j=1}^{n} A_{ij} (x_{j,1} - x_{i,1}), \\
\frac{d\hat{x}_{i,2}}{dt} &= 0.999x_{i,1} + 0.349x_{i,2}, \\
\frac{d\hat{x}_{i,3}}{dt} &= 0.181 + x_{i,3}(0.993x_{i,1} - 5.645),
\end{align*}
\]  

(S24)

To illustrate the generative power of the inferred equations obtained by our approach, we set different initial states to validate if the inferred equations can reproduce the dynamics of the original model. The results in Supplementary Fig. 10 show that the inferred equations indeed resemble the dynamics of original models. It is worth mentioning that although our approach infers the true equations with high precision, the distance between the inferred and the true trajectories increases for longer time due to the chaoticity of the original model.
Supplementary Figure 8: Oscillators’ activities and trajectories generated by the true Rössler network dynamics and the inferred dynamical equations.

Supplementary Figure 9: a. Relative errors ∆ of the inferred elementary functions and their coefficients. b,c. Nodes’ activities and trajectories generated by the true and the inferred dynamical equations.
**Supplementary Figure 10:**

**a.** A node’s $x$-dimension trajectories generated by the true and the inferred dynamical equations for an initial state that is different from the initial state of original data.

**b.** Divergence of trajectories generated by the true and the inferred dynamical equations.
D Edge dynamics of social balance

We also tested our approach with a social balance dynamics described by edge dynamical equation [S33], i.e. the relation between two individual nodes is represented by \( x_{ij} \). Entry \( x_{ij} \) represents the strength of trust or distrust between \( i \) and \( j \). The strength of trust or distrust between nodes evolves according to

\[
\dot{X} = X^2.
\]  

(S25)

We rewrite Eq. (S25) explicitly in terms of entries \( x_{ij} \) as

\[
\frac{dx_{ij}}{dt} = \sum_k x_{ik}x_{kj}.
\]  

(S26)

The edge activities generated by the true equation are shown in Supplementary Fig. 11. To infer such an equation, we construct libraries by a variety of functions with temporal topologies data \( X(t) \), for example, \( X^2, X^3, \sin(X) \), etc. By our approach, we automatically infer the exact term and its coefficient, leading to the equation

\[
\frac{d\hat{x}_{ij}}{dt} = 0.999 \sum_k x_{ik}x_{kj}.
\]  

(S27)

Supplementary Figure 11: Activities of six typical edges generated by the inferred Eq. (S27).
E Reproducibility description

To ensure the reproducibility of our results we list the parameters of dynamics and networks for each plot in the main paper. The simulations of dynamics are from time 0 to $T$ with step-size $\delta t$.

**Supplementary Table 4:** Parameters setting in this paper.

| Plots   | Dynamics                  | Networks   | Simulation data |
|---------|---------------------------|------------|-----------------|
| Fig. 1b-f | $\epsilon^H = 0.30, \epsilon^R = -0.15$ | BA ($\gamma = 3$) | $T = 500$, $\delta t = 0.01$ |
| Fig. 2  | FHN, $\epsilon^{FHN} = 1.00$ | ER, SF ($\gamma = 2.5$), Drosophila | $T = 140$, $\delta t = 0.01$ |
| Fig. 3a | HR, $\epsilon^H = 0.30, \epsilon^R = -0.15$ | ER, SF ($\gamma = 2.5$), C.elegans | $T = 500$, $\delta t = 0.01$ |
| Fig. 3c | S.B., $\epsilon^S = 1.00$ | ER, SF ($\gamma = 2.5$), Advogato | $T = 0.7$, $\delta t = 0.001$ |
| Fig. 3e | Kuramoto, $\epsilon^{kura} = 0.015$ | ER | $T = 100$, $\delta t = 0.01$ |
| Fig. 3e | Kuramoto, $\epsilon^{kura} = 0.03$ | SF ($\gamma = 2.5$), U.S. power grid | $T = 100$, $\delta t = 0.01$ |
| Fig. 3g | Rössler, $\epsilon^{ross} = 0.15, \sigma = 0.1$ | ER, SF ($\gamma = 2.5$) | $T = 100$, $\delta t = 0.01$ |
| Fig. 3g | Rössler, $\epsilon^{ross} = 0.10, \sigma = 0.1$ | N.E. power grid | $T = 100$, $\delta t = 0.01$ |
| Fig. 4a | FHN, $\epsilon^{FHN} = (0.5, 20)$ | BA ($\gamma = 3$) | $T = 100$, $\delta t = 0.01$ |
| Fig. 4b | Kuramoto, $\epsilon^{kura} = 0.03, \sigma = (0.1, 1.1)$ | BA ($\gamma = 3$) | $T = 100$, $\delta t = 0.01$ |
| Fig. 4c | HR, $\epsilon^H = 0.30, \epsilon^R = -0.15$ | SF ($\gamma = 2.5$) | $T = 500$, $\delta t = 0.01$ |
| Fig. 4c | GR, $\epsilon^{GR} = 0.05$ | BA ($\gamma = 3$) | $T = 500$, $\delta t = 0.01$ |
| Fig. 5a-j | HR, $\epsilon^H = 0.30, \epsilon^R = -0.15$ | BA ($\gamma = 3$) | $T = 500$, $\delta t = 0.01$ |
| Fig. 5k | HR, $\epsilon^H = 0.30, \epsilon^R = -0.15$ | BA ($\gamma = 3$) | $T = 500$, $\delta t = 0.01$ |
| Fig. 5l | Rössler, $\epsilon^{ross} = 0.20, \sigma = 0.1$ | BA ($\gamma = 3$) | $T = 100$, $\delta t = 0.01$ |

In Fig. 5 of the main paper $\sigma = 0.1$ for Kuramoto and Rössler network dynamics, and the simulations of dynamics for each model are same as those in Figs. 2-3.

IV Details of inferrability

A Quantifying synchronization

To quantify the degree of synchronization of a dynamical network, the sequence of peaks in $x_i(t)$ is extracted for defining the geometric phase $\theta_i(t)$ of each node $i$. For example, in the neuronal spiking activities data, we use a Poincaré’s section at $x_i = 0.5$ to determine the time of the start (upward sense) and the end (downward sense) of a spike (Supplementary Fig. 12(a,c)). Between the $l$-th spike and $(l+1)$-th spike, the phase $\theta_i(t)$ is increased by $2\pi$, such that an interpolation between the two spikes can define the continuous time-varying phase as \cite{S23,S24}

$$\theta_i(t) = 2\pi t_i + 2\pi \frac{t - t_{l,i}}{t_{l+1,i} - t_{l,i}}, \quad t_{l,i} \leq t < t_{l+1,i},$$

(S28)
where $t$ is the current time, and $t_{i,l}$ is the time at which node $i$ starts the $l$-th spike. The degree of synchronization of the dynamical network is then quantified by order parameter

$$R(t) = \left| \frac{1}{N} \sum_{i=1}^{N} e^{i\theta_i(t)} \right|. \quad (S29)$$

When $R(t)$ is close to zero, the dynamical network is totally disorder; if $R(t) = 1$, the activities of all nodes are completely synchronized, indicating a complete phase synchronized state. Averaging $R(t)$ over time we obtain a single order parameter to quantify network synchronization, i.e.

$$\langle R \rangle = \frac{1}{t_f - t_0} \sum_{t=t_0}^{t_f} R(t), \quad (S30)$$

where $t_0$ and $t_f$ are the start and the end of the time range respectively. A network usually becomes more synchronized when the coupling strength increases. It is worth noting that, for some network dynamics, such as FHN, the nodes can evolve into two categories, i.e. get trapped in two fixed points (Supplementary Fig. 12e). For this case, the order parameters of the two categories are calculated respectively and the average of them is considered as the order parameter of the whole network.

### B Inferring alternative equations

If the hidden true dynamics is not a combination of elementary functions or if some elementary function of the hidden dynamics is not included in the large libraries $L_F$ and $L_G$, our approach will infer alternative forms of $F$ and $G$ that are still able to characterize the observed system behaviors. To demonstrate such capacity, we deliberately remove some elementary functions that should exist in the true network dynamics. For example, we delete interaction terms $\frac{x_{i,1}}{1 + e^{-x_{j,1} - \beta}}$ and $\frac{1}{1 + e^{-x_{j,1} - \beta}}$ ($\alpha = 10$, $\beta = 1$), yet to infer HR network dynamics with $r = 0.004$ and $x_0 = -1.5$. Even for such deficient libraries our approach infers alternative equations

$$\begin{align*}
\frac{d\hat{x}_{i,1}}{dt} &= 2.903x_{i,1}^2 - 1.006x_{i,1}^3 + 0.981x_{i,2} - 0.985x_{i,3} - 0.043\sin(x_{i,1}) + 3.211 \\
&\quad + 0.119 \sum_{j=1}^{N} A_{ij} \frac{1}{1 + e^{-x_{j,1} - \beta}} + 0.053 \sum_{j=1}^{N} A_{ij} \sin(x_{j,1}), \\
\frac{d\hat{x}_{i,2}}{dt} &= 1.000 - 4.994x_{i,1}^2 - 0.998x_{i,2}, \\
\frac{d\hat{x}_{i,3}}{dt} &= 0.024 + 0.016x_{i,1} - 0.004x_{i,3},
\end{align*} \quad (S31)$$

that still capture the HR network dynamics (see Fig. 4c in the main paper), as well as the alternative equation

$$\begin{align*}
\frac{d\hat{x}_{i,1}}{dt} &= -0.0185 - 0.11\sin x_{i,1} - 0.09x_{i,1}^2 + 0.01 \sum_{j=1}^{N} A_{ij} \frac{x_{j,1} x_{i,1}}{x_{j,1} x_{i,1} + 1} \\
&\quad + 0.012 \sum_{j=1}^{N} A_{ij} \frac{x_{j,1}^5}{x_{j,1}^5 + 1} + 0.017 \sum_{j=1}^{N} A_{ij} \sin x_{j,1}, \\
\frac{d\hat{x}_{i,2}}{dt} &= 0.024 + 0.016x_{i,1} - 0.004x_{i,3}, \quad \frac{d\hat{x}_{i,3}}{dt} = 0.024 + 0.016x_{i,1} - 0.004x_{i,3},
\end{align*} \quad (S32)$$

to capture GR network dynamics (see also Fig. 4c).
**Supplementary Figure 12:** Quantification of synchronization for spiking activities. 

a. Poincaré’s section of HR neuronal dynamics segmented by the blue dotted line. 
b. Order parameter $\langle R \rangle$ of HR network dynamics vs. coupling strength for different network. 
c. Poincaré’s section of FHN neuronal dynamics segmented by the blue dotted line. 
d. Order parameter $\langle R \rangle$ of FHN network dynamics vs. coupling strength. 
e. Two categories of $x(t)$, obtained by clustering. 
f. Inference inaccuracy (sMAPE) increases when the coupling strength becomes larger, indicating that more synchronized is a network, more difficult to infer its hidden interaction mechanisms.

**C Inferrability comparison studies**

In this subsection we compare our approach with SINDy [$^{S23}$] and its variant [$^{S36}$] regarding inferrability, i.e. the capacity of dealing with partially synchronized networked dynamics or dynamical heterogeneity. The results showing in Supplementary Fig. 13 indicate that our method is better at coping with synchronization and dynamical heterogeneity. The network properties and dynamics settings are the same as that in Fig. 4 of the main paper.

**V Details of robustness analyses**

In this work we validated the robustness of our approach against five types of unavoidable uncertainties in observation data, including observational, uncorrelated and correlated dynamical noises in data of nodes’ activities, missing and spurious links in observed topologies, low resolution by experimental techniques. To construct incomplete topologies we randomly add or delete a proportion of entries in the true adjacency matrix $A_{ij}$; To imitate the constraint of low resolution we regularly down-sample the time series of nodes’ activities. In the following we describe in detail the simulations of data with observational, dynamical and correlated noise, and show the failure ratio of inferring the structure of true elementary functions.
Supplementary Figure 13: Inferrability comparison between SINDy, its variant and our approach regarding synchronization (FHN dynamics; a), dynamical heterogeneity (Kuramoto dynamics; b), and deficient libraries (HR dynamics; c,d). c. The alternative equations obtained by SINDy-variant, which selects the smallest number of terms and lowest relative AIC value, resulting in an equation containing 34 terms. d. Comparison of discrepancy between SINDy-variant and our approach. Note that the number of terms in the alternative equation inferred by our approach is 14, much smaller than 34 by SINDy-variant yet similar to 13 in the true equation.

out of 100 independent runs in Supplementary Fig. 14.

A Noisy data

There are usually two types of noises in time series data, i.e. dynamical noise and observational noise[S27]. The latter is raised in the observation or measurement process, while the former indicates the inherent stochasticity of system dynamics. To test the robustness of our approach against these two types of noises, we simulate the data through[S27]

\[
\begin{align*}
\frac{dx_i(t)}{dt} &= F(x_i(t)) + \sum_{j=1}^{n} A_{ij} G(x_i(t), x_j(t)) + \eta A \cdot dw_i(t), \\
X^{obs}(t) &= x_i(t) + a\beta_X(t),
\end{align*}
\]

(S33)

where \(\eta A \cdot dw_i(t)\) is dynamical noise, and \(a\beta_X(t)\) is observational noise. Here \(dw_i(t)\) is the Gaussian white noise, which follows a normal distribution with mean zero and standard deviation \(\sqrt{dt}\), and \(A\) is the average amplitude of the original time series. Here \(\eta\) represents the intensity of dynamical noises relative to the signal amplitude. We simulate Eq. (S33) by using the fourth-order Runge-Kutta method with fixed time step size. The \(\beta_X(t)\) is a Gaussian noise following a normal distribution with zero mean and standard deviation one, hence the intensity of observational noise can be tuned by the value of \(a\). In the paper we generate observational noisy data by awgn in MatLab and characterize the noisy data with specified Signal-to-Noise-Ratio (SNR).

Correlated dynamical noise could exist in complex systems when the noises are correlated across different
nodes (such as in neuronal activities\cite{S28},\cite{S29}). Then the network dynamics become

\[
\frac{dx_i(t)}{dt} = F(x_i(t)) + \sum_{j=1}^{n} A_{ij} G(x_i(t), x_j(t)) + \epsilon_i,
\]

where \(\epsilon = \{\epsilon_i\}\) is correlated noise across all nodes drawn from a multivariate normal distribution with mean 0 and covariance \(\Sigma\), \(i.e.
\]

\[
\epsilon \sim \eta' N(0, \Sigma).
\]

The noise covariance matrix \(\Sigma\) is semidefinite which is constructed by \(\Sigma = AA^\top\), where \(A\) is adjacency matrix, and \(\eta'\) is the strength of correlated noise. In the main paper we also validate the robustness of our approach against such correlated noises (Fig. 5j).

\[
\begin{align*}
\text{Supplementary Figure 14:} & \quad \text{The failure ratio of inferring the structure of true dynamics of 100 independent runs (a-e), and the inference inaccuracies sMAPE of inferring both true elementary functions and their coefficients (f-j).} \\
\text{B Ablation studies} & \quad \text{In the main paper we showed that each key sub-step of our two-phase approach is indispensable for inferring the hidden network dynamics from noisy and incomplete data. Here we further show that, even for clean data these key sub-steps are also necessary. To do so we individually remove each sub-step and test the performance of the corresponding deficient approach. The result of these ablation studies are shown in Supplementary Fig. 15.} \\
& \quad (1) \quad \text{Topological sampling ablated: In Phase II we randomly select sets of single node and its neighbors, instead of the all nodes. Such topological sampling can imitate the incompleteness in observed topology and increase the accuracy of the approach. If this sub-step is removed, the accuracy indeed decreases significantly.} \\
& \quad (2) \quad \text{Using AIC instead of wAIC: As shown in Supplementary Table 5, AIC could rank the relevance of elementary functions incorrectly. Indeed, it brings also several irrelevant terms. In contrast, we find that wAIC can successfully identify the most relevant terms and, importantly, there is a clear threshold separating the relevant and irrelevant terms. Hence, using AIC instead of wAIC causes significantly}
\end{align*}
\]
**Supplementary Figure 15:** The coupled results of ablation studies on clean data, for inferring the hidden equations governing HR network dynamics (a) and coupled Rössler dynamics (b). Five ablation studies were performed and shown in box-whisker plot: (1) removing the sub-step of topological sampling, (2) using original AIC instead of wAIC, (3) removing Phase II, (4) removing Phase I, (5) No normalization in $\Theta_F$ and $\Theta_G$. Statistical significance is obtained by multiple Mann-Whitney test\textsuperscript{[S30],[S31]}. Three or four asterisks indicate $p$-value $10^{-3}$ or $< 10^{-4}$, and n.s. means not significant.

| Ablation Study | Inference Inaccuracy (sMAPE) |
|----------------|------------------------------|
| Full method    | 1.0                          |
| Topological sampling ablated | 0.95                        |
| Use AIC instead of wAIC     | 0.9                          |
| Phase II ablated         | 0.85                         |
| Phase I ablated          | 0.8                          |
| No normalization in $\Theta_F$ and $\Theta_G$ | 0.75                        |

- **Phase II ablated:** Without the local fine-tuning, Phase I alone is not able to obtain a compact form of the hidden equation. Moreover, as shown in Fig. 1d in the main paper, even though the equation inferred by Phase I alone fits the observational data well, it does not have generative power.

- **Phase I ablated:** If Phase I is ablated from the approach, we will lose the global information that captures the coarse yet consistent structure of the hidden dynamics. Hence we find the ablation of Phase I also significantly increase the inference inaccuracy.

- **No normalization in Phase I:** As shown in Supplementary Fig. 3, the value of several elementary functions can span several orders of magnitude, possibly significantly larger than that of others. If the sub-step of normalization is removed the coefficients of these inherently high-value terms, obtained by regression will be very small, which decreases the possibility of the existence of these terms.

### C Robustness comparison studies

In the main paper we show the results of the comparisons between our approach, SINDy, and ARNI. The original ARNI is a model-free inference approach for network structure reconstruction\textsuperscript{[S32]}. ARNI uses the idea of matrix pseudoinverse to connect a node to other most relevant nodes by minimizing the cost between numerical derivatives of observed time series and the sum value of basis functions, enabling the inference of network topology from nodes’ activities. We would like to emphasize that although ARNI originally aimed at inferring network structure, it can also be used to infer network dynamics if network structure is given. For a fair comparison, we transfer ARNI to infer network dynamics as follows. We apply the same idea of matrix pseudoinverse and add most relevant elementary functions one by one, as shown in Supplementary Fig. 16, until the cost is lower than a threshold. To improve ARNI’s ability of coping with the possible incompleteness in observed network structure, we also perform topological samplings in ARNI. The resulting elementary functions and their coefficients are the value averaged for multiple topological samplings. The details of ARNI modification can be seen in our code.

We also perform robustness comparisons between our approach and SINDy variant\textsuperscript{[S26]} against observational noises. The results are shown in Supplementary Fig. 17, where the task is to inferring HR dynamics...
Supplementary Table 5: The example showing the results of using AIC vs. wAIC for automatically choosing the relevant elementary functions. Ten leading elementary functions are ranked by their absolute coefficients. Blue color denotes those which should be selected to compose the true governing equation.

| elementary functions | coefficients | AIC[$\times10^6$] | wAIC[$\times10^7$] |
|---------------------|--------------|-----------------|---------------------|
| $x_{i,2}$           | -1.0         | 0.28            | 0.29                |
| $x_{i,3}$           | -0.8         | -0.16           | -0.02               |
| $\sin(x_{i,3})$     | -0.08        | -0.16           | -0.02               |
| $x_{j,1} - x_{i,1}$ | 0.07         | -0.13           | -0.01               |
| $x_{i,1}$           | -0.03        | -0.15           | -0.05               |
| $x_{i,1}x_{i,3}$    | -0.02        | -0.04           | -0.02               |
| $\frac{1}{1+e^{-10(x_{j,1}-x_{i,1})-1}}$ | 0.01 | -0.16 | -0.3 |
| $\frac{1}{1+e^{-5[(x_{j,1}-x_{i,1})-1]}}$ | 0.00 | -0.16 | -1.4 |
| $e^{x_{j,1}-x_{i,1}}$ | 0.00 | -0.16 | -166 |
| $e^{x_{j,1}}$       | 0.00         | -0.20           | -1750               |

Supplementary Figure 16: Inference of elementary functions by ARNI. a. The cost curve when most relevant elementary functions are added one by one. The IDs of the true elementary functions are highlighted by red. b. The accuracy of the inferred elementary functions is evaluated by AUC score.

from simulated activity data with 40dB observational noise. The simulation time $T = 500$ with step size $\delta t = 0.01$. The network is generated with directed BA model with size $n = 100$ and average degree $\langle k \rangle = 5$. 


Supplementary Figure 17: Inferring HR dynamics from simulated activity data with 40dB observational noise by SINDy variant \cite{S36} and our approach respectively. **a.** Number of terms in inferred model by SINDy variant vs. the relative AIC. SINDy variant selects the optimal model with the lowest relative AIC value and the smallest number of terms (highlighted by purple). **b.** Inference inaccuracies comparison between SINDy variant and our approach.
VI Details of empirical system inference

In the main text we have demonstrated the applicability of our approach to inferring dynamical equations from empirical data. Specifically we used the global spreading data of H1N1 disease reported daily from April 24th to July 6th in year 2009. As we focused on the dynamics of early spreads before governments introduce quarantine policies, only the first 45 daily data were used for the inference. For instance, if the first case in a country was reported on May 1st, we employed the data from May 1st to June 14th. Note that, although the starting time of spread for different countries are different, in the coupling dynamics $G(x_i, x_j)$ the time $t$ is the same date for all nodes. In other words, in Fig. 6 and Supplementary Figs. 18-20 the initial time $t$ [/day] being shifted to $t = 1$ for all nodes are just for the convenience of visualization. The inference approach indeed used the original dates and data. The comparisons between the inferred and the empirical activities are also displayed in Supplementary Fig. 18 for H1N1, in Supplementary Fig. 19 for SARS and in Supplementary Fig. 20 for COVID-19.
Supplementary Figure 18: Comparison between empirical number of H1N1 cases, the infected numbers generated by the inferred Eq. (5) in the main text with identical parameters $a$ and $b$ for all nodes (i.e. effective), and the infected numbers generated by the form of Eq. (5) yet with heterogeneous parameters $a_i$ and $b_i$ for each node $i$ (i.e. heterogeneous). The parameter for effective dynamics are $a = 0.074$ and $b = 7.13$, and those $a_i$ and $b_i$ for heterogeneous dynamics are shown in the sub-plots respectively.
Supplementary Figure 19: Comparison between empirical number of SARS cases, the infected numbers generated by the inferred Eq. (5) in the main text with identical parameters \(a\) and \(b\) for all nodes (i.e. effective), and the infected numbers generated by the form of Eq. (5) yet with heterogeneous parameters \(a_i\) and \(b_i\) for each node \(i\) (i.e. heterogeneous). The parameter for effective dynamics are \(a = 0.046\) and \(b = 3.083\), and those \(a_i\) and \(b_i\) for heterogeneous dynamics are shown in the sub-plots respectively. Note that Eq. (5) was inferred based only on H1N1 data, hence the results in this figure demonstrate the generalizability of the inferred dynamical equation by our approach.
Supplementary Figure 20: Comparison between empirical number of COVID-19 cases, the infected numbers generated by the inferred Eq. (5) in the main text with identical parameters $a$ and $b$ for all nodes (i.e. effective), and the infected numbers generated by the form of Eq. (5) yet with heterogeneous parameters $a_i$ and $b_i$ for each node $i$ (i.e. heterogeneous). The parameter for effective dynamics are $a = 0.040$ and $b = 105.160$, and those $a_i$ and $b_i$ for heterogeneous dynamics are shown in the sub-plots respectively. Note that Eq. (5) was inferred based only on H1N1 data, hence the results in this figure show that the inferred equation is also able to capture the early spreads of COVID-19.
References

[S1] Pedregosa, F. et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

[S2] Erdős, P. & Rényi, A. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci* **5**, 17–60 (1960).

[S3] Barabási, A.-L. & Bonabeau, E. Scale-free networks. *SciAm.* **288**, 60–69 (2003).

[S4] Goh, K., Kahng, B. & Kim, D. Universal behavior of load distribution in scale-free networks. *Phys. Rev. Lett.* **87**, 278701 (2001).

[S5] Hagberg, A. A., Schult, D. A. & Swart, P. J. Exploring network structure, dynamics, and function using networkx. In Varoquaux, G., Vaught, T. & Millman, J. (eds.) *Proceedings of the 7th Python in Science Conference*, 11 – 15 (Pasadena, CA USA, 2008).

[S6] White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **314**, 1–340 (1986).

[S7] Varshney, L. R., Chen, B. L., Paniagua, E., Hall, D. H. & Chklovskii, D. B. Structural properties of the *Caenorhabditis elegans* neuronal network. *PLoS Comput. Biol.* **7**, e1001066 (2011).

[S8] Yan, G. et al. Network control principles predict neuron function in the *Caenorhabditis elegans* connectome. *Nature* **550**, 519–523 (2017).

[S9] Scheffer, L. K. et al. A connectome and analysis of the adult drosophila central brain. *Elife* **9**, e57443 (2020).

[S10] Menck, P. J., Heitzig, J., Kurths, J. & Schellnhuber, H. J. How dead ends undermine power grid stability. *Nature Commun.* **5**, 1–8 (2014).

[S11] Kunegis, J. Konect: the koblenz network collection. In *Proceedings of the 22nd International Conference on World Wide Web*, 1343–1350 (2013).

[S12] Rossi, R. A. & Ahmed, N. K. The network data repository with interactive graph analytics and visualization. In *AAAI*. http://networkrepository.com (2015).

[S13] Rabinovich, M. I., Varona, P., Selverston, A. I. & Abarbanel, H. D. Dynamical principles in neuroscience. *Rev. Mod. Phys* **78**, 1213 (2006).

[S14] Borges, F. et al. Inference of topology and the nature of synapses, and the flow of information in neuronal networks. *Phys. Rev. E* **97**, 022303 (2018).

[S15] FitzHugh, R. Impulses and physiological states in theoretical models of nerve membrane. *Biophys. J.* **1**, 445–466 (1961).

[S16] Mazur, J., Ritter, D., Reinelt, G. & Kaderali, L. Reconstructing nonlinear dynamic models of gene regulation using stochastic sampling. *BMC Bioinform.* **10**, 1–12 (2009).

[S17] Schröder, M., Timme, M. & Witthaut, D. A universal order parameter for synchrony in networks of limit cycle oscillators. *Chaos* **27**, 073119 (2017).

[S18] Pietras, B. & Daffertshofer, A. Network dynamics of coupled oscillators and phase reduction techniques. *Phys. Rep.* **819**, 1–105 (2019).
[S19] Arenas, A., Díaz-Guilera, A., Kurths, J., Moreno, Y. & Zhou, C. Synchronization in complex networks. *Phys. Rep.* **469**, 93–153 (2008).

[S20] Li, X.-W. & Zheng, Z.-G. Phase synchronization of coupled rössler oscillators: amplitude effect. *Commun. Theor. Phys.* **47**, 265 (2007).

[S21] Minati, L. et al. Connectivity influences on nonlinear dynamics in weakly-synchronized networks: Insights from rössler systems, electronic chaotic oscillators, model and biological neurons. *IEEE Access* **7**, 174793–174821 (2019).

[S22] Marvel, S. A., Kleinberg, J., Kleinberg, R. D. & Strogatz, S. H. Continuous-time model of structural balance. *Proc. Natl Acad. Sci. USA* **108**, 1771–1776 (2011).

[S23] Boaretto, B., Budzinski, R., Prado, T. & Lopes, S. Mechanism for explosive synchronization of neural networks. *Phys. Rev. E* **100**, 052301 (2019).

[S24] Checco, P., Righero, M., Biey, M. & Kocarev, L. Synchronization in networks of hindmarsh–rose neurons. *IEEE Trans. Circuits Syst. II Express Briefs* **55**, 1274–1278 (2008).

[S25] Brunton, S. L., Proctor, J. L. & Kutz, J. N. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc. Natl. Acad. Sci. USA* **113**, 3932–3937 (2016).

[S26] Mangan, N. M., Kutz, J. N., Brunton, S. L. & Proctor, J. L. Model selection for dynamical systems via sparse regression and information criteria. *Proc. Math. Phys. Eng. Sci.* **473**, 20170009 (2017).

[S27] Sase, T., Ramírez, J. P., Kitajo, K., Aihara, K. & Hirata, Y. Estimating the level of dynamical noise in time series by using fractal dimensions. *Phys. Lett. A* **380**, 1151–1163 (2016).

[S28] Eyherabide, H. G. & Samengo, I. When and why noise correlations are important in neural decoding. *J. Neurosci.* **33**, 17921–17936 (2013).

[S29] Van Bergen, R. & Jehee, J. F. Modeling correlated noise is necessary to decode uncertainty. *Neuroimage* **180**, 78–87 (2018).

[S30] Cheung, Y. K. & Klotz, J. H. The Mann Whitney Wilcoxon distribution using linked lists. *Stat. Sin.* **805–813** (1997).

[S31] Klotz, J. H. A computational approach to statistics. *Department of Statistics, University of Wisconsin, Madison* 15–25 (2006).

[S32] Casadiego, J., Nitzan, M., Hallerberg, S. & Timme, M. Model-free inference of direct network interactions from nonlinear collective dynamics. *Nature Commun.* **8**, 1–10 (2017).