

# SCIENTIFIC REPORTS

**OPEN**

## Inferring sparse networks for noisy transient processes

Hoang M. Tran<sup>1,2</sup> & Satish T.S. Bukkapatnam<sup>1</sup>

Received: 08 October 2015  
Accepted: 03 February 2016  
Published: 26 February 2016

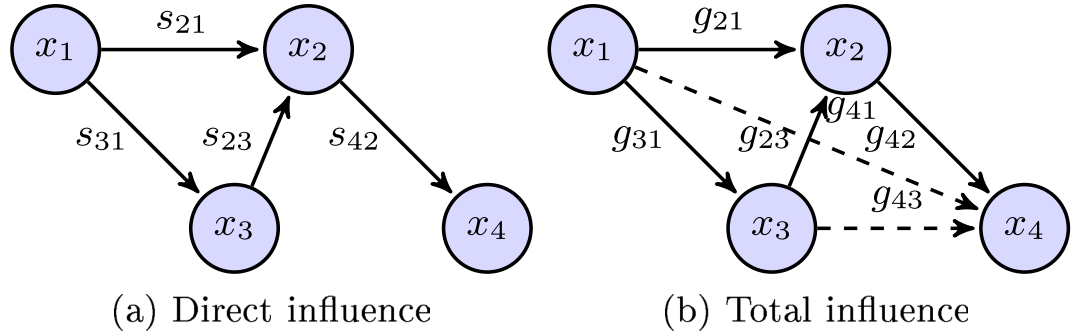
Inferring causal structures of real world complex networks from measured time series signals remains an open issue. The current approaches are inadequate to discern between direct versus indirect influences (i.e., the presence or absence of a directed arc connecting two nodes) in the presence of noise, sparse interactions, as well as nonlinear and transient dynamics of real world processes. We report a sparse regression (referred to as the  $\ell_1$ -min) approach with theoretical bounds on the constraints on the allowable perturbation to recover the network structure that guarantees sparsity and robustness to noise. We also introduce averaging and perturbation procedures to further enhance prediction scores (i.e., reduce inference errors), and the numerical stability of  $\ell_1$ -min approach. Extensive investigations have been conducted with multiple benchmark simulated genetic regulatory network and Michaelis-Menten dynamics, as well as real world data sets from DREAM5 challenge. These investigations suggest that our approach can significantly improve, oftentimes by 5 orders of magnitude over the methods reported previously for inferring the structure of dynamic networks, such as Bayesian network, network deconvolution, silencing and modular response analysis methods based on optimizing for sparsity, transients, noise and high dimensionality issues.

Many real world processes including biological<sup>1,2</sup>, socio-economics<sup>3,4</sup>, and engineering systems<sup>5</sup>, can be represented as large scale dynamic networks<sup>6</sup>. The multitude of state variables of the process represent the network nodes and the arcs represent the dynamic coupling between pairs of state variables. Inferring the structure of these networks is critical for multiple purposes such as identifying key causal relationship, clustering, partitioning or reducing the system state space; thereby facilitating effective prediction, control and/or interventions of its underlying processes. For example, inferring the signaling pathways of the gene p53 was noted to be crucial towards advancing cancer treatment<sup>7</sup>.

Real world processes exhibit nonlinear dynamics and they almost always occur in transient conditions. Identifying the structure, especially the existence or absence of a direct dynamic coupling between the variables of such systems has been noted to be a standing challenge of modern science<sup>8</sup>, and the underlying causal mechanisms remain largely undiscovered. Most often, only noisy measurements of the network outputs in the form of a small ensemble of time series data are available for network inference<sup>8–13</sup>. The use of conventional system identification approaches can produce many spurious links due to the transitivity of influences among the nodes. Several methods for network inference notably based on Bayesian update<sup>14–19</sup>, Granger causality and multivariate autoregressive<sup>20–24</sup>, partial correlation<sup>25</sup>, network deconvolution (ND)<sup>26</sup>, network silencing<sup>27</sup> and conditional causal relation<sup>28–31</sup> have been investigated to filter the effect of indirect influences. When the time series gathered under transient conditions were available, a Modular Response Analysis (MRA)<sup>32–34</sup> method was proposed to infer the network structure at each time point. However, these methods suffer from serious drawbacks such as they mostly assume the system to exhibit linear and time-invariant dynamics<sup>26</sup>, determinism (noise-free)<sup>33–35</sup>, and/or the existence of a point attractor under steady state<sup>27</sup>. While MRA method can be employed to reconstruct dynamics under transient conditions<sup>33</sup>, its performance deteriorates sharply in the presence of noise and the method encounters severe numerical stability issues, especially when the underlying dynamics is highly nonlinear. This tends to severely restrict its applicability to real world processes. Notably, the earlier methods essentially focus on dealing with each of the following scenarios including transient time series<sup>33</sup>, noisy measurements<sup>14–19</sup>, and indirect influence removal<sup>14–25,33</sup> separately. The realistic scenario combining all these scenarios has not been considered. All available methods literally break down when presented with this scenario.

Towards addressing this gap, we introduce an approach based on modifying ND, silencing and MRA methods to account for sparsity, transients, noise and high dimensionality issues. Specifically, we have investigated a sparse

<sup>1</sup>Department of Industrial & Systems Engineering, Texas A&M University, College Station, TX 77840, USA. <sup>2</sup>School of Applied Mathematics & Informatics, Hanoi University of Science & Technology, Hanoi, Vietnam. Correspondence and requests for materials should be addressed to S.T.S.B. (email: [satish@tamu.edu](mailto:satish@tamu.edu))



**Figure 1. Illustration of direct and total influence.** The total influences in (b) are the accumulation of the influences transited through all paths in (a). For example, the total influence  $1 \rightarrow 4$  in (b) is the accumulation of the influence transited through the paths  $1 \rightarrow 3 \rightarrow 2 \rightarrow 4$  and  $1 \rightarrow 2 \rightarrow 4$ .

regression (henceforth referred to as the  $\ell_1$ -min) formulation to recover the structure of dynamic networks from noisy data gathered under transient conditions. Our main contribution is in providing a theoretical bound on the constraints of the  $\ell_1$ -min formulation and providing stable numerical procedures that overcome effects of nonlinear couplings in large interconnected processes, availability of only a small sample of short time series ensembles, and inaccuracies in estimating noise levels. These bounds mitigate tedious trial and error procedures employed customarily as part of  $\ell_1$ -min implementations<sup>1,34–36</sup>. The theoretical results and subsequent experimental studies suggest that the present  $\ell_1$ -min approach is more robust to noise compared to the contemporary dynamic Bayesian network<sup>14–19</sup> as well as NDs<sup>26,27,32</sup>. It is shown that up to 5 orders of magnitude reduction in the inference error are possible from the present approach, leading to a more accurate inference of the network structure for complex real world networks.

### Methods

Towards a more formal treatment, we define a real world system as high dimensional coupled differential equation of the form

$$\frac{dx}{dt} = f(x, p), \tag{1}$$

or

$$x(t) = \Phi(t, x(\tau)), \tag{2}$$

where  $x \in M \subset \mathbb{R}^n$  is a state vector,  $p$  is the parameter vector,  $x(\tau)$  is an initial condition. As noted in the foregoing, such dynamics can also be represented in form of a network<sup>37</sup> shown in Fig. 1, where the node  $i$  represents the state variable  $x_i$  and a directed arc represents the existence and the strength of the coupling (direct influence)  $s_{ij}$  between node  $i$  and node  $j$ . In this context, the direct influence  $s_{ij}(t)$  of node  $j$  on node  $i$  around a certain point  $x$  in the state space defined in Eq. (1) can be expressed as

$$s_{ij}(t) \triangleq \frac{\partial x_i}{\partial x_j(\tau)} = \frac{\partial \Phi_i(t, x(\tau))}{\partial x_j(\tau)}. \tag{3}$$

It may be noted that, a node  $j$  is connected to a node  $i$  at time  $t$  if  $s_{ij}(t) \neq 0$ . Hence,  $S(t) = (s_{ij}(t))$  captures the physical structure of the dynamical system (1) at time  $t$ . In practice,  $S(t)$  needs to be inferred from the measurements of the total influence  $g_{ij}(t)$  between every pair of nodes<sup>26,27</sup> or estimated from time series outputs of the dynamic system gathered under transient conditions<sup>33</sup>. The total influence  $g_{ij}(t)$  is the sum of the direct influence of node  $j$  on node  $i$  and all indirect influences from node  $j$  to node  $i$  through other nodes connecting to both of them (see Fig. 1b). For example, total influence from  $1 \rightarrow 4$ ,  $g_{41}(t)$  is the sum of indirect influences along the paths  $1 \rightarrow 3 \rightarrow 2 \rightarrow 4$  and  $1 \rightarrow 2 \rightarrow 4$ , or  $g_{41}(t) = s_{31}(t)s_{23}(t)s_{42}(t) + s_{21}(t)s_{42}(t)$ . In other words, the total influence that node  $j$  has on node  $i$  around a certain point  $x$  on the state space defined in Eq. (1) is defined recursively as

$$g_{ij}(t) \triangleq \frac{dx_i(t)}{dx_j(\tau)} = \frac{d\Phi_i(t, x(\tau))}{dx_j(\tau)} \tag{4}$$

$$= \frac{\partial \Phi_i(t, x(\tau))}{\partial x_j(\tau)} + \sum_{k \neq j} \frac{\partial \Phi_i(t, x(\tau))}{\partial x_k(\tau)} \frac{dx_k(t)}{dx_j(\tau)} \tag{5}$$

$$=s_{ij}(t) + \sum_{k \neq j} s_{ik}(t) g_{kj}(t), \quad (6)$$

which is similar to the expression noted in in Barzel and Barabási<sup>27</sup>. Conventionally, under stationarity assumptions,  $g_{ij}(t)$  can be approximated using similarity measures, such as correlation and mutual information<sup>8</sup> estimated from raw samples of time series. The direct and total influence matrices are related at every time  $t$  by the following equation:

$$S(t)B(t) - C(t) = 0, \quad (7)$$

where  $B(t)$  and  $C(t)$  are functions (defined depending on the context) of  $S(t)$  and  $G(t)$ , respectively. Pertinently, when the underlying dynamical system is linear and time-invariant,  $S(t)$  and  $G(t)$  do not depend on time. Eq. (7) generalizes previous network deconvolution formulations as follows: for Feizi *et al.*<sup>26</sup>,  $B(t) = (I + G)$ ,  $C(t) = G$ , for Barzel and Barabási<sup>27</sup>  $B(t) = G$ ,  $C(t) = G - I + \mathcal{D}(SG)$ , and for Sontag *et al.*<sup>33</sup>,  $B(t) = R(t)$ ,  $C(t) = \Gamma(t)$ , where  $R_{ij}(t) = \partial x_i(t, \mathbf{p}) / \partial p_j$ ;  $\Gamma_{ij}(t) = \partial R_{ij}(t) / \partial t$ . For simplicity of expressions, we use henceforth  $S$ ,  $B$  and  $C$  instead of  $S(t)$ ,  $B(t)$  and  $C(t)$  in this subsection. The “true” network structure  $S^0$  can be estimated by solving the following  $\ell_1$ -min formulation:

$$S^* = \arg \min_S \|S\|_1 \text{ s.t. } \|SB - C\|_F \leq \mathcal{E}, \quad (8)$$

where  $\|S\|_1 = \sum_{i,j} |s_{ij}|$ , and  $\mathcal{E}$  is the allowable perturbation that captures the effects of noise in the measured data. We note that in the absence of noise, this formulation is equivalent to ND and MRA. In the following sections we present two alternative  $\ell_1$ -min formulations for direct influence inference. The first formulation presented in Eqs (9, 10) addresses the estimation of  $s_{ij}$  for real world scenarios when the total influence  $g_{ij}$  is directly measurable (e.g., based on the strengths of co-excitations), and the second formulation Eqs (21, 22) addresses the inference of the network structure (i.e., determine all node pairs where  $s_{ij}(t) = 0 \forall t$ ) under one of the most generic scenarios of using multiple ensembles of time series realizations of the state variables, collected under noisy and transient conditions with different parameter settings. It may be noted that inferring the network structure under such generic conditions has not been investigated to date.

**Network inference when total influence matrix is available.** For the case where the measurements of total influence matrix  $G$  are provided<sup>26</sup>, the relaxed  $\ell_1$ -min formulation can be written as

$$\min \|S\|_1 \text{ s.t. } \|S(G + I) - G\|_F \leq \mathcal{E}, \quad (9)$$

or in vector form as

$$\min_{s_i} \|s_i\|_1 \text{ s.t. } \|(G + I)^T s_i - \mathbf{g}_i\|_2 \leq \varepsilon_i, \quad \forall i, \quad (10)$$

where  $\mathbf{g}_i$  is the  $i^{\text{th}}$  column of  $G$ . In order to solve for an accurate estimate of  $S^0$  from Eqs (9) or (10) using standard solvers<sup>38,39</sup>, estimation of  $\mathcal{E}$  and  $\varepsilon_i$  are crucial. Specifically, when noisy measurements of the total influence matrix differ from the “true” total influence as  $G = G^0 + \Delta G$ , the estimated direct influence matrix differs from the true direct influence matrix as  $S = S^0 + \Delta S$ , and

$$(S^0 + \Delta S)(G + I) = G \quad (11)$$

$$\Rightarrow S^0(G + I) - G = -\Delta SG - \Delta S. \quad (12)$$

The quantity  $\|\Delta SG + \Delta S\|_F$  is called total perturbation. In vector form,  $(\|\Delta G s_i^0\|_2 + \|\mathbf{g}_i - \mathbf{g}_i^0\|_2)$  can represent the total perturbation for computing row  $i$  of  $S^0$ . The bounds on  $\mathcal{E}$  and  $\varepsilon_i$  are as follows (See Theorem 1 in Supplementary Information):

$$\mathcal{E} \approx \mathcal{E}^{(1)} = (1 + \|G\|_F) \gamma, \quad (13)$$

$$\mathcal{E} \leq \mathcal{E}^{(2)} = (1 + \|G\|_F) \frac{\|\Delta G\|_F}{(1 - \|G\|_F - \|\Delta G\|_F)(1 - \|G\|_F)}, \quad (14)$$

$$\varepsilon_i \leq 2 \left( \|\mathbf{g}_i - \mathbf{g}_i^0\|_2^2 + \|\Delta G\|_F \frac{1}{\sqrt{1 - \delta_K}} (\|\mathbf{g}_i\|_2 + \|\mathbf{g}_i - \mathbf{g}_i^0\|_2)^2 \right), \quad (15)$$

where  $\gamma$  is the largest eigenvalue of  $\Delta G$ ,  $\delta_K$  is the restricted isometry constant<sup>40</sup> and  $\|\cdot\|_F$  is the Frobenius norm of a matrix. By employing these bounds, we can set the values of  $\mathcal{E}$  and  $\varepsilon_i$  for effective network inference. As subsequent numerical investigations indicate, the performance of the method does not degrade significantly due to the presence of noise, and this is the major advantage of the present approach. It may be noted that our method is

designed to provide the sparsest network structure that replicates the measured total influence  $G$  within a bound (specified in terms of the allowable total perturbation). This is very important because only a small set of noisy observations are available, for most real world applications. For example, in the case of genetic regulatory networks, only a subset of dynamic regimes (i.e. marked by the active degrees of freedom) of the underlying process are captured. Therefore, identification of true network structure would never be guaranteed by any approach, and among the network structures that can replicate the observed total influence within a specified bound, the sparsest network would be of the most interest. Although sparser than the network derived by ND,  $\ell_1$ -min derived structure might be adequate to uncover the total dynamic couplings of the process captured in the observed data.

In real world scenarios,  $\Delta G$  is not always known. Overestimation of  $\Delta G$  can lead to network structures that are sparser than the original. However, we show that the effects of under-estimation of noise can be alleviated to a great extent. When noise level is unknown but multiple realizations of the noisy measurements of  $G$  are available, it is possible to further reduce the inference error by combining the estimates with different realizations of  $G$  as  $\bar{S}^{(N)} = \frac{1}{N} \sum_{r=1}^N \hat{S}^{(r)}$  (See the Proposition 1 in Supplementary Information), where  $\hat{S}^{(r)}$ 's are direct influence matrices computed from  $\hat{G}^{(r)}$  and  $\hat{G}^{(1)}, \dots, \hat{G}^{(N)}$  are  $N$  different measurements or estimates of the total influence matrix  $G^0$ . This result assumes that  $Var(\hat{S}^{(r)})$  is bounded. However, it may be noted that even if  $Var(\hat{S}^{(r)})$  is arbitrarily large we find that  $\bar{S}^{(N)}$  is at least as good as  $\hat{S}^{(r)}$ . This averaging procedure allows us to improve the network inference accuracy when multiple measurements of the total influence matrix are available. For example, when the network structure does not change significantly as the system approaches a steady state, the total influence matrices can be measured multiple times, each corresponds to one time window.

**Network inference when the time series under transient conditions are available (total influence matrix not given).**

In practice,  $g_{ij}$  are often estimated using convenient similarity measures such as correlation or mutual information between the time series  $x_i(t)$  and  $x_j(t)$  of the nodes  $i$  and  $j$  as stated in the foregoing section. These estimations have a very low accuracy due to nonstationaries (transient), low sampling rates and sample size limitation; and can not capture the total influence in the system. Also, in most real world applications, only finite samples of time series  $x(t)$  are available, and the present NDs can not be employed in these scenarios. To overcome these drawbacks, we have adapted an approach to estimate the direct influence based on multiple time series ensembles obtained by perturbing parameters of the dynamical system Eq. (1)<sup>33</sup>. We first modify the perturbation procedure proposed by Sontag *et al.*<sup>33</sup> to make it more robust to numerical error then further improve the accuracy of network inference by introducing a sparse regression formulation and the averaging scheme.

*A robust perturbation procedure.* According to Sontag *et al.*<sup>33</sup>,  $s_{ij}(t) = \left[ \frac{\partial f_i(\mathbf{x}, \mathbf{p})}{\partial x_j} \right]_{i,j=1..n}$  can be derived from the following equation:

$$\Gamma(t) = S(t)R(t), \tag{16}$$

where

$$R_{ij}(t) = \frac{\partial x_i}{\partial p_j} \approx \frac{x_i(t, p_j + \Delta p_j) - x_i(t, p_j)}{\Delta p_j}, \Gamma_{ij}(t) = \frac{\partial \dot{x}_i}{\partial p_j} \approx \frac{R_{ij}(t + \Delta t) - R_{ij}(t)}{\Delta t}, \tag{17}$$

and

$$i = 1..n, p_k \in P_i = \{p_k \in \mathbf{p} : \partial f_i / \partial p_k(\mathbf{x}, \mathbf{p}) = 0\}. \tag{18}$$

Note that  $\Gamma$  plays the role of the total influence matrix  $G$  in the previous section. To compute the row  $i$  of the matrix  $S$ , the parameters  $p_j$  to be perturbed are chosen such that  $p_j \in P_i$ <sup>33</sup>. As a consequence, changes in  $p_j$  indirectly affect  $x_i$ , and  $\frac{dx_i}{dp_j}$  are much smaller than  $\frac{dx_k}{dp_j}$ , for  $k \neq i$ . As a result, the  $i^{th}$  column  $(\frac{dx_i}{dp_j})_{n \times 1}$  in the matrix  $(\frac{dx_k}{dp_j})_{n \times n}$  is much smaller (2 orders of magnitude smaller as in the Table 1 for the network studied in case study 1) compared to other columns when  $p_j \in P_i$ . A numerical issue this poses can be understood based on the following linear system of equations

$$A\mathbf{u} = \mathbf{b}.$$

Here, the sensitivity of solution  $\mathbf{u}$  to the change in  $A$  can be quantified as follows<sup>41</sup>

$$\frac{\partial u_i}{\partial a_{jk}} = -c_{ij} \sum_l c_{kl} b_l,$$

where  $C = A^{-1}$ . Whenever  $A$  contains a  $j$  column such that  $\|a_j\| \ll \|a_k\|, \forall k \neq j$ ,  $C$  contains a row  $i$  such that  $\|c_i\| \gg \|c_r\|, \forall r \neq i$ . As a consequence,  $\frac{\partial u_i}{\partial a_{jk}}$  becomes several magnitudes larger than other rows. Therefore, the perturbation procedure proposed by Sontag *et al.*<sup>33</sup> is very unrobust to noise or numerical error in  $x_i$ 's.

	$r_{\cdot 1}$	$r_{\cdot 2}$	$r_{\cdot 2}$	$r_{\cdot 2}$	$r_{\cdot 2}$	$r_{\cdot 2}$	$r_{\cdot 2}$	$\max  r_{\cdot j} $
$r_{1\cdot}$	-2.868e-4	-7.284e-5	-3.106e-4	-1.578e-4	2.443-e4	-8.315e-5	-4.896e-4	<b>0.0005</b>
$r_{2\cdot}$	1.160e-4	0.1050	-4.261e-4	-1.803e-4	6.490e-4	2.261e-4	-2.379e-4	<b>0.1050</b>
$r_{3\cdot}$	-1.136e-4	-2.658e-4	0.1179	-2.766e-4	1.370e-4	-2.524e-4	2.776e-4	<b>0.1179</b>
$r_{4\cdot}$	-4.431e-4	-4.543e-4	-4.138e-4	0.0961	6.824e-4	6.710e-5	1.609e-4	<b>0.0961</b>
$r_{5\cdot}$	-2.397e-4	-4.439e-4	-1.225e-4	-7.024e-4	0.1100	3.256e-4	2.069e-4	<b>0.1100</b>
$r_{6\cdot}$	4.053e-4	-3.773e-4	-2.577e-4	-5.065e-5	0.0012	0.1195	4.481e-4	<b>0.1195</b>
$r_{7\cdot}$	-1.030e-4	3.312e-5	-2.900e-4	-5.258e-5	0.0100	-1.651e-4	0.0820	<b>0.0820</b>

**Table 1.** The matrix  $R$  for computing the first row of  $S$  is estimated using Sontag *et al.*<sup>33</sup>'s perturbation procedure. The first row/column of  $R$  is two orders of magnitude smaller than others, which presents major numerical issues for inferring structures of large networks.

The following modification to the perturbation procedure addresses the aforementioned issue. Consider the case when  $\dot{x}_i$  depends linearly on  $x_i$  as in the following system<sup>42</sup>:

$$\dot{x}_i = p_i x_i + \sum_{j=1, j \neq i}^n s_{ij}^0 \frac{x_j}{1 + x_j}.$$

This system describes popular biochemical reactions when the activity of a chemical species is inhibited by its own concentration<sup>43,44</sup>. To compute the  $i^{\text{th}}$  row of the Jacobian, the parameters  $p_i$  is also perturbed. Note that

$$\frac{\partial \dot{x}_i}{\partial p_i} = x_i + p_i \frac{\partial x_i}{\partial p_i} + \sum_{j=1, j \neq i}^n \frac{\partial f_i(\mathbf{x}, \mathbf{p})}{\partial x_j} \frac{\partial x_j}{\partial p_i}, \quad (19)$$

or

$$\frac{\partial \dot{x}_i}{\partial p_i} - x_i = \sum_{j=1, j \neq i}^n \frac{\partial x_j}{\partial p_j} \frac{\partial f_i(\mathbf{x}, \mathbf{p}_i)}{\partial x_j}, \quad \frac{\partial f_i(\mathbf{x}, \mathbf{p}_i)}{\partial x_i} = p_i. \quad (20)$$

The remaining parameters are perturbed as in Eqs (17, 18). Therefore, to compute  $\frac{\partial f_i(\mathbf{x}, \mathbf{p}_i)}{\partial x_k}$ , we can solve the system of equations (16) with

$$P_{ii}(t) = \frac{\partial \dot{x}_i}{\partial p_i} - x_i, \quad R_{ii}(t) = \frac{\partial x_i}{\partial p_i}$$

and other  $P_{ij}$ s,  $R_{ij}$ s are defined as in (17, 18).

*A robust network identification approach.* In addition to the perturbation procedure proposed in Eqs (17–19), we present a method to solve Eq. (16) that is more robust to the presence of noise. In the present context, the  $\ell_1$ -min formulation of Eq. (16) takes the following form:

$$\min \|S\|_1 \text{ s.t. } \|\Gamma - SR\|_F \leq \mathcal{E}, \quad (21)$$

or

$$\min \|s_i\|_1 \text{ s.t. } \left\| \Gamma_{ik}(t) - \sum_{l=1}^n R_{lk}(t) s_{il}(t) \right\| \leq \varepsilon_i, \quad \forall i, \forall k: p_k \in P_i. \quad (22)$$

As noted in the foregoing section, estimation of  $\mathcal{E}$  and  $\varepsilon_i$  based on the noise levels when measuring  $\mathbf{x}(t)$  is essential to ensure that the solution to Eq. (21) serves as a viable estimator of the “true” direct influence  $S^0$ . The following bounds and approximation allow the specification of  $\mathcal{E}$  and  $\varepsilon_i$  (Theorems 4 and 5 in Supplementary Information)

$$\mathcal{E} \leq (\|\Gamma\|_F + \|\Delta\Gamma\|_F) \frac{\|R^{-1}\Delta R\|_F}{1 - \|R^{-1}\Delta R\|_F} + \|\Delta\Gamma\|_F, \quad (23)$$

$$\varepsilon_i \leq \frac{\|R^{-1}\Delta R\|_F}{1 - \|R^{-1}\Delta R\|_F} \left\| [(\Gamma - \Delta\Gamma)' ]_i \right\| + \|(\Delta\Gamma')_i\|, \quad (24)$$

$$\varepsilon_i \approx \|R^{-1}\Delta R\| \left\| [(\Gamma - \Delta\Gamma)' ]_i \right\| + \|(\Delta\Gamma')_i\|, \quad (25)$$

where

Formula	Mean
$\mathcal{E}^{(0)} = \ (G + I)^T S^0 - G\ _F$	$9.79 \times 10^{-3}$
$\mathcal{E}^{(1)}$ (Eq. 13)	$8.89 \times 10^{-3}$
$\mathcal{E}^{(2)}$ (Eq. 14)	$8.61 \times 10^{-2}$

**Table 2.** Comparison of bounds on total perturbation obtained using Eqs (13) and (14) suggests that Eq. (13) provides a good approximation and Eq. (14) serves as an upper bound of  $\mathcal{E}^{(0)}$ .

$$(\Delta R)_{ik}(t) = (e_{ik}^{(1)}(t) - e_{ik}^{(2)}(t)) / \Delta p_k,$$

$$(\Delta \Gamma)_{ik}(t) = \frac{[(e_{ik}^{(1)}(t + \Delta t) - e_{ik}^{(2)}(t + \Delta t)) - (e_{ik}^{(1)}(t) - e_{ik}^{(2)}(t))]}{\Delta t \Delta p_k},$$

and  $e_{ik}^{(1)}(t)$ ,  $e_{ik}^{(2)}(t)$  are the errors incurred when measuring  $x_i^0(t, p_k)$ ,  $x_i^0(t, p_k + \Delta p_k)$ , respectively. As stated in the foregoing, noise level is not known *a priori* in most real world systems. In this situation, the network structure is deduced based on the entries in the estimated  $S^0(t)$  that are equal to zero for all  $t$  and can be estimated by the entries in  $\bar{S}^{(N)} = \frac{1}{N} \sum_{r=1}^N \hat{S}(t_r)$  that converge to zero, where  $\hat{S}(t_r)$  is the direct influence matrix computed from  $\hat{\Gamma}(t_r)$ , and  $\hat{\Gamma}(t_r)$ , ( $r = 1.. N$ ) are measurements or approximations of the total influence matrix  $\Gamma^0(t)$  at time  $t_r$  (see Proposition 2 in Supplementary Information). This averaging procedure allows us to improve the accuracy to predict the pair of nodes that are not connected when the measurement noise level is not available. As a result, our method ensures low false positive rates on the “arcs”. As noted in the context of Proposition 1, network inference with  $\bar{S}^{(N)}$  tends to be at least as good as with  $\hat{S}(t_r)$  even when  $Var(\hat{S}(t_r))$  is arbitrarily large.

## Results

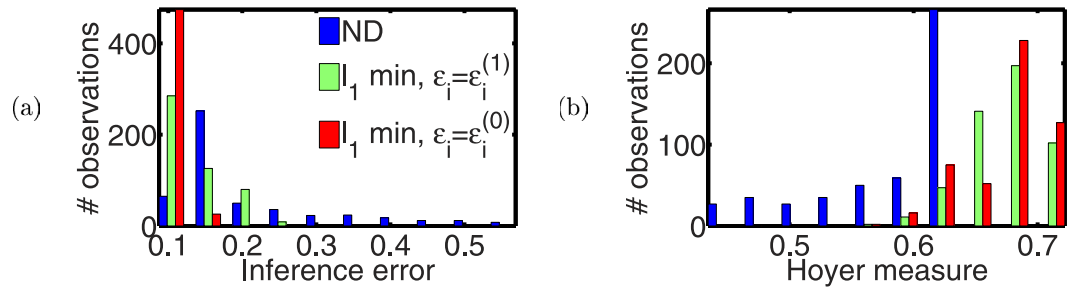
We have considered two case studies to validate the theoretical results and evaluate the performance of the  $\ell_1$ -min approach. The first case study contains two simulation scenarios. The first scenario simulates a scale-free network whose structure resembles that of the genetic regulation process of *E. Coli* species<sup>45</sup>. Here, the challenge is to estimate the true network structure, i.e., the direct influence matrix  $S^0$  from a noisy total influence matrix  $G$ . This scenario is optimal for assessing the closeness of the bounds stated in Eqs (14,15) relative to the true bounds on the constraints  $\mathcal{E}^{(0)} = \|(G + I)^T S^0 - G\|_F$ , and comparing the performance of the  $\ell_1$ -min formulation relative to the recent ND methods in terms of inference error and sparsity. The next scenario simulates a system of Hill-type differential equations modeling a gene interaction network. Here, the challenge is to estimate the true network structure from noisy and transient time series data. The second case study is an application of our method to infer genetic regulatory networks (GRNs) from empirical data in the context of DREAM5 challenge<sup>46</sup>. This challenge is a standard framework for evaluating GRN inference methods.

**Case I: simulation studies.** *Inferring direct influence networks from total influence network.* First, we adapted the procedure specified by Muchnik<sup>47</sup> to generate 500 random realizations of scale-free networks consisting of  $n = 100$  nodes, with a degree exponent of 2.2. In each realization, the weights of the true direct influence network,  $s_{ij}^0$  follow the distribution  $\mathcal{N}(\mu_{s^0}, \sigma_{s^0}^2)$  with  $\mu_{s^0} \sim \mathcal{N}(0, 0.04)$ , and  $\sigma_{s^0} \sim \mathcal{N}(0, 0.04)$ . The true total influence matrix  $G^0$  was obtained as  $G^0 = S^0(I - S^0)^{-1}$ . The noisy total influence matrix was generated as  $G = G^0 + \Delta G$ , where the contaminated noise  $\Delta G$  was considered in two cases: (1) proportional, i.e.,  $(\Delta G)_{ij} = \alpha \mathcal{N}(\mu_{s^0}, \sigma_{s^0}^2)$  and (2) independent, i.e.,  $(\Delta G)_{ij} = \mathcal{N}(0, \sigma_{s^0}^2)$ . We considered cases where the measurement noise level  $\|\Delta G\|_F$  is known as well as those where there is uncertainty in estimating the measurement noise level.

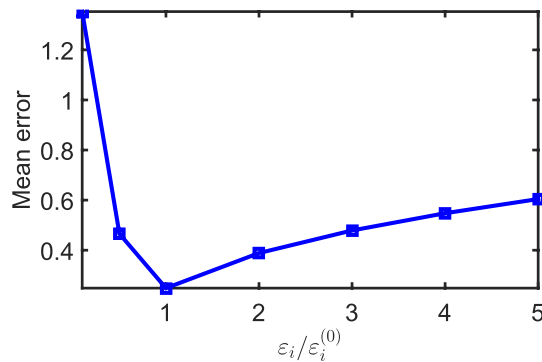
We first compare the “true” bound  $\mathcal{E}^{(0)}$  (computed using  $S^0$ ) and the bounds for  $\mathcal{E}$  estimated based on Eqs (13,14). In the presence of noise, the bounds appear to be in the same order of magnitude for all simulated networks (Table 2). The results also suggest that the bound specified in Eq. (13) closely matches the “true” bound and can be used to approximate the feasible region when  $\mathcal{E}^{(0)}$  is unknown with high accuracy. Although the bound in Eq. (14) tends to be loose, it can be used as an upper bound for  $\mathcal{E}^{(0)}$ .

We next compared the performance of ND and  $\ell_1$ -min approaches (using our bounds Eqs (13) and (14)) in terms of inference error defined as  $\rho = \frac{\|\hat{S} - S^0\|_F}{\|S^0\|_F}$ , where  $\hat{S}$  is computed using the different methods being compared. The  $\ell_1$ -min approach with “true” constraint bound  $\varepsilon_i = \varepsilon_i^{(0)}$  significantly improves the ND (the mean and the variance of the estimated  $\rho$  were reduced by 45% and 99%, respectively) (Fig. 2). Employing  $\varepsilon_i = \varepsilon_i^{(1)} / \sqrt{n}$  (based on Eq. (13)), the  $\ell_1$ -min approach performs much better than ND (the mean and variance of  $\rho$  are reduced by 33.5% and 87.5%, respectively). More importantly, the inference error of  $\ell_1$ -min approaches were concentrated around of 0.15 within  $\pm 0.05$ , while those of ND were spread over a larger range, from 0.3 to 0.6. This suggests that  $\ell_1$ -min approach using our bound in Eq. (13) is more robust than ND to noise and approximation error incurred when measuring the total influence matrix.

We also compared the sparsity of the recovered networks measured in terms of Hoyer sparsity measure<sup>48</sup> defined as follows



**Figure 2.** Histograms summarizing the relative performance of ND and  $\ell_1$ -min approaches for the benchmark numerical case in terms of (a) inference error that quantifies the accuracy and (b) Hoyer measure that quantifies the sparsity of the solution. The solution from the  $\ell_1$ -min approach is more precise and sparser than ND: compared to NDs, the mean and the variance of the inference error are reduced by 45% and 99%, respectively, when using  $\ell_1$ -min with  $\varepsilon_i = \varepsilon_i^{(0)}$ ; 33.5% and 87.5%, respectively when using  $\ell_1$ -min with  $\varepsilon_i = \varepsilon_i^{(1)}$ ; the mean of Hoyer measure is increased by 16.38% and variance reduced by 69% when using the  $\ell_1$ -min with  $\varepsilon_i = \varepsilon_i^{(0)}$ , and is increased by 15.90% in mean, reduced by 75.69% in variance when using  $\varepsilon_i = \varepsilon_i^{(1)}$ .



**Figure 3.** Variation of inference error with total perturbation bound  $\varepsilon_i$ . The inference error attains a minimum near the true bound  $\varepsilon_i^{(0)}$ , and it trends almost linearly with  $\varepsilon_i$  as it is increased beyond  $\varepsilon_i^{(0)}$ . As  $\varepsilon_i \rightarrow 0$ , the inference error increases exponentially, which is an evidence of over fitting.

$$Hoyer(S) = \frac{n - (\sum_{i,j=1}^n |s_{ij}|) / \sqrt{\sum_{i,j=1}^n s_{ij}^2}}{n - 1}$$

Note that  $Hoyer(S) \in [0, 1]$ . The closer it is to 1, the sparser  $S$  is. In terms of this measure, the solution of the  $\ell_1$ -min approach is much sparser (mean is 16.38% larger, variance is 69% smaller when using the true bound  $\varepsilon_i = \varepsilon_i^{(0)}$ , and mean is 15.90% larger, variance is 75.69% smaller when using the approximated bound  $\varepsilon_i = \varepsilon_i^{(1)}$ ) than solution of ND (Fig. 2b). Also, the Hoyer measure of the  $\ell_1$ -min approach is concentrated more around a much higher value (sparse matrices) than that of ND indicating that the  $\ell_1$ -min approach using our bound gives a significantly sparser solution than ND. As a result, this gives a more interpretable connection structure without the loss of performance.

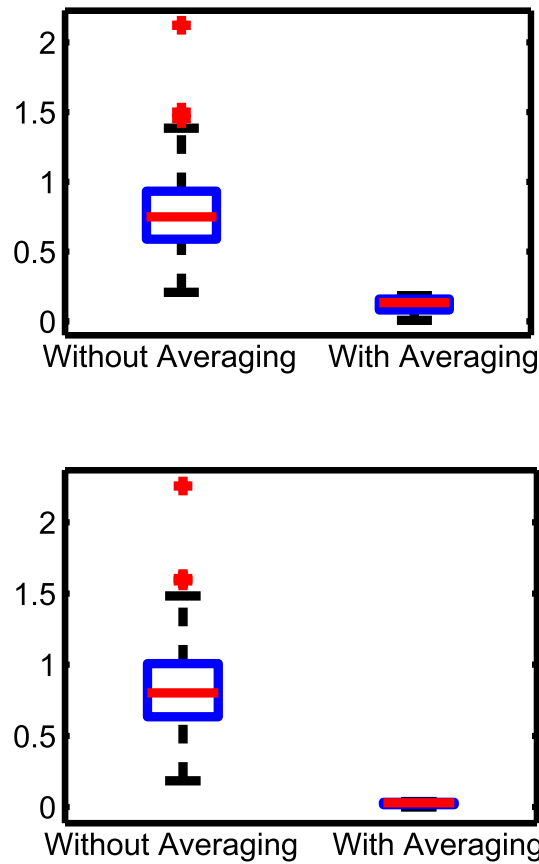
We also studied the effects of the bounds of  $\ell_1$ -min formulation on inference error to verify Eq. (40) numerically. When  $\varepsilon_i / \varepsilon_i^{(0)} > 1$ , the inference error trends almost linearly with  $\varepsilon_i$  (see Fig. 3). This confirms the conclusion of Theorem 3. Also, when  $\varepsilon_i / \varepsilon_i^{(0)} < 1$  and tends toward 0, the inference error increases. This shows an evidence of over-fitting.

Subsequently, we studied the effect of averaging (Proposition 1) in the context of the  $\ell_1$ -min and ND methods. We conducted  $N = 40$  simulations, in each of which,  $S^0$ ,  $G^0$  and  $\Delta G$  were generated as stated in the foregoing. We used the inference error without  $\rho^{(N)}$  and with averaging  $\bar{\rho}$  as measures for comparison from each simulation defined as follows:

$$\rho^{(N)} = \frac{1}{N} \sum_{k=1}^N \left\| \hat{S}^{(k)} - S^0 \right\|_F / \|S^0\|_F, \tag{26}$$

$$\bar{\rho} = \|\bar{S}^{(N)} - S^0\|_F / \|S^0\|_F, \tag{27}$$





**Figure 4.** Box plots summarizing the effects of averaging on (a) ND and (b)  $\ell_1$ -min with  $\varepsilon_i = \varepsilon^{(1)}$ . The inference errors were almost unchanged with  $\ell_1$ -min compared to ND. Averaging (light/red) reduced inference error further by about 8 times compared to without averaging (dark/blue). The  $\bar{p}$  values were 0.1196 with ND and 0.0259 with  $\ell_1$ -min (p-values of the paired t-tests between the inference error without and with averaging were  $\leq 10^{-5}$  in all cases).

where  $\hat{S}^{(k)} (k = 1, \dots, N)$  is the  $k^{th}$  realization of  $S^{(k)}$  and  $\bar{S}^{(N)}$  is estimated as stated in Proposition 1. The results suggest that averaging reduces the inference error of both methods by about 8 times in all cases, thus supporting the validity of Proposition 1 (Fig. 4). The inference errors were almost the same between ND and  $\ell_1$ -min with  $\varepsilon_i = \varepsilon^{(1)}$ .

*Inferring direct influence network structure from multiple time series under transient conditions.* In this section we represent the performance of  $\ell_1$ -min approach in inferring network structure from transient time series with an unknown noise level. In this study we used Michaelis-Menten dynamic system given by<sup>27</sup>:

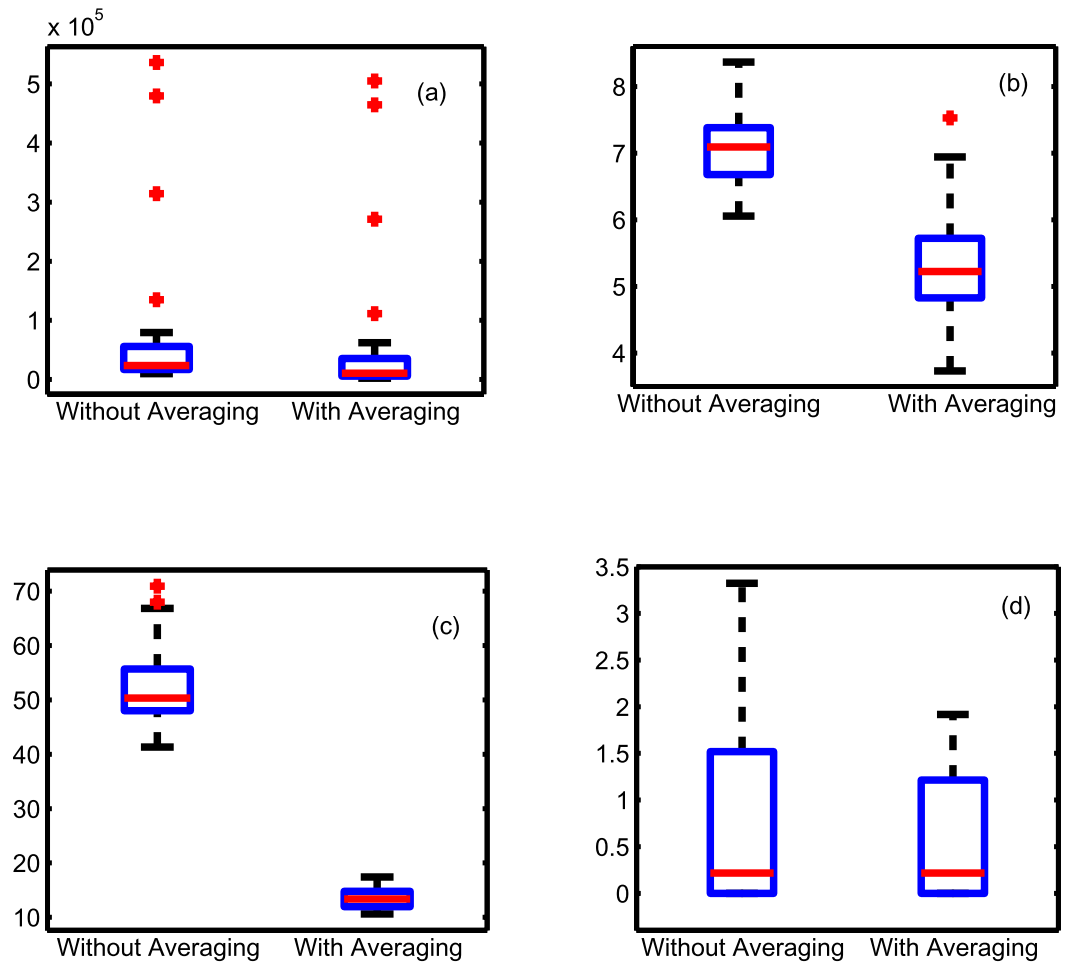
$$\dot{x}_i = p_i x_i + \sum_{j=1, j \neq i}^n s_{ij}^0 \frac{x_j}{1 + x_j}, \tag{28}$$

where the “true” network defined by  $(s_{ij}^0)$  is a scale-free network<sup>45</sup> generated randomly with degree exponent  $\gamma = 2.2$  consisting of  $n = 40$  nodes with about 70 edges, whose weights  $s_{ij}^0$  follow the distribution  $\mathcal{N}(5, 0.25)$ .

We obtained 30 different variants of this network. For each of these invariants (trials), a perturbed network was obtained by changing (perturbing) the parameters according to Eqs (18–20). Every solution  $\mathbf{x}(t), t \in [0, 1]$ , obtained from an initial condition  $\mathbf{x}(0)$  was contaminated with noise of the form  $\mathcal{N}(0, \sigma^2)$  to simulate a noisy measurement  $\hat{\mathbf{x}}(t)$ . Here  $\sigma^2$  was chosen to be  $10^{-4}$ . The direct influence matrix  $\hat{S}(t_k)$  were estimated using Sontag *et al.*'s<sup>33</sup> method, as well as  $\ell_1$ -min formulations, with different values of bounds. Next,  $\bar{S}^{(N)}$  was estimated as in Proposition 2 by averaging over 30 time samples  $t_k \in [0, 1], k = 1..30$  chosen randomly. For performance evaluation, we used the inference error without  $\rho^{(N)}$  and with averaging  $\bar{\rho}$ , given by

$$\rho^{(N)} = \sqrt{\frac{1}{N} \sum_{k=1}^N \sum_{i,j} (1 - \mathcal{H}(|s_{ij}^0|)) (\hat{s}_{ij}(t_k))^2}, \tag{29}$$





**Figure 5. Box plots summarizing the inference errors without and with averaging for.** (a) Sontag *et al.*'s<sup>33</sup> method ( $\rho = 7.58 \times 10^4, \bar{\rho} = 5.87 \times 10^4$ ); (b)  $\ell_1$ -min with noise magnitude given ( $\rho = 7.11, \bar{\rho} = 5.32$ ), (c)  $\ell_1$ -min with noise magnitude underestimated as 10% the actual ( $\rho = 52.50, \bar{\rho} = 13.40$ ), and (d)  $\ell_1$ -min with noise magnitude overestimated as 10 times the actual ( $\rho = 0.80, \bar{\rho} = 0.60$ ). The inference error was reduced by  $10^5$  times when using the  $\ell_1$ -min approach (21, 22), compared to Sontag *et al.*'s<sup>33</sup> method. Averaging further reduced inference error by at least 30% in all cases (p-values of the paired t-tests consistently were below 0.0282).

$$\bar{\rho} = \sqrt{\sum_{ij} (1 - \mathcal{H}(|s_{ij}^0|)) (\bar{s}_{ij}^{(N)})^2}, \tag{30}$$

where  $\mathcal{H}(\cdot)$  is Heaviside function. These error measures quantify the number of absent links ( $s_{ij}^0 = 0$ ) that are correctly identified.

As summarized in Fig. 5, the  $\ell_1$ -min approach performs better than Sontag *et al.*'s<sup>33</sup> method in all cases tested. In fact,  $\rho, \bar{\rho}$  were reduced by  $10^5$  times. The poor performance of Sontag *et al.*'s<sup>33</sup> method is attributed to the numerical issues noted in the earlier section. A further 30% reduction in inference error resulted from averaging for both cases. Next, the cases (c) and (d) were designed to simulate the real situations where the noise magnitude is unknown. We considered cases where the noise levels are under or overestimated by 1 order of magnitude. While Sontag *et al.*'s<sup>33</sup> method would not be applicable in such cases,  $\ell_1$ -min without averaging was found to lead to suboptimal inference. Under underestimation  $\varepsilon_i < \varepsilon_i^{(0)}/10$ , averaging was found to further reduce the inference error by about 70%, and the inference error  $\bar{\rho}$ s were of the same level as one would obtain when the noise level is known. This result is consistent with and is a clear verification of Proposition 2. When the noise level is overestimated, the resulting network tends to be highly sparse, offering excellent specificity in identifying the absence of direct coupling. The inference errors are therefore low even without averaging by default. In this case averaging reduces the inference errors by 5%. The p-values of the paired t-tests between the inference error with and without averaging were below 0.0282 in all cases suggesting that averaging helps improve network inference.

**Case II: Application to empirical genetic regulatory network inference.** Next, we applied our method to infer real world GRNs and compare its performance with other methods including ND<sup>26</sup>, Bayesian network inference, Pearson and Spearman correlation networks<sup>8</sup> using the framework presented in DREAM5 challenge. Here, the Pearson and Spearman correlations were considered as they are the most widely used methods for network inference and can provide a reasonable estimation of the total influence matrix<sup>26,27</sup>. In addition, ND has been most effective in inferring network topology when the total influence matrix  $G$  is estimated using Pearson and Spearman correlations. Therefore, these serve as the challenging test cases to evaluate the performance of  $\ell_1$ -min where ND is already effective. The DREAM5 challenge contains gene-expression microarray data of three species including an *in silico* benchmark, a prokaryotic model organism (*E. coli*) and a eukaryotic model organism (*S. cerevisiae*). Beside  $\rho$  and Hoyer metrics, we employed the following score, which was used in earlier works<sup>8</sup> to assess the performance of a network inference method for recovering the structure underlying these data sets:

$$\xi = -\frac{\log(p_{ROC}) + \log(p_{PR})}{2}, \quad (31)$$

where  $p_{ROC}$  and  $p_{PR}$  are p-values computed from AUROC (area under receiver operating characteristic curve) and AUPR (area under precision-recall curve).

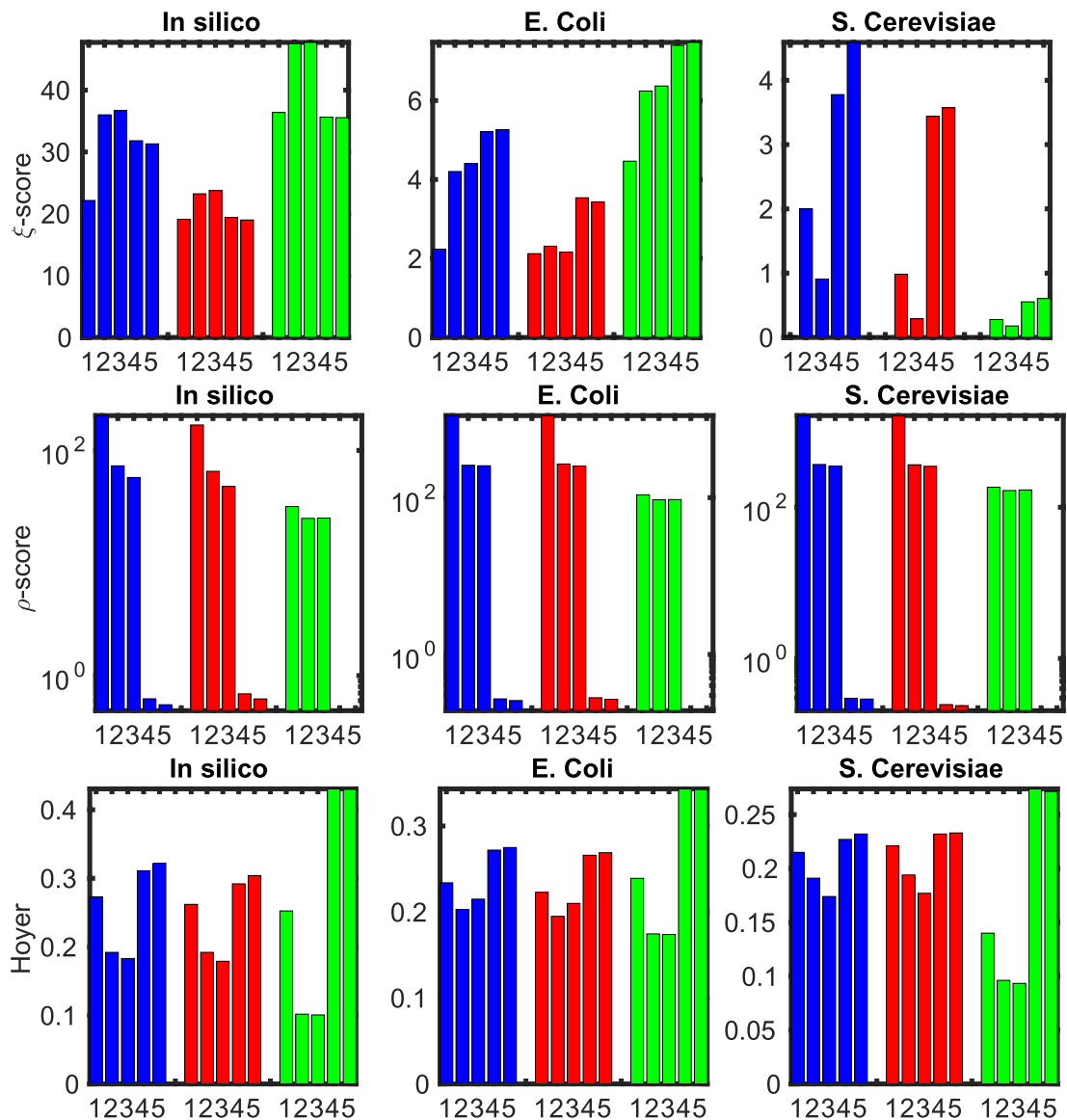
The results of the performance evaluation are summarized in Fig. 6. We note that for computing the performance metrics we first generated 30 different  $G$  matrices with Pearson correlation, 30 others with Spearman correlation and another 30 with Mutual Information for each data set. The  $G$  matrix in each case was estimated using samples of size 75% of the data set. The averaging procedure considers the  $S$  matrices estimated from these  $G$  matrices using different methods. In terms of  $\xi$ -score (Eq. (31)), which quantifies how well—in terms of having low false negative rates (FNR, related to sensitivity), and low false positive rates (FRN, related to specificity), the true positive rate (TPR) and true negative rate (TNR)—the estimated  $\hat{S}$  captures  $S^0$ ,  $\ell_1$ -min approach yields  $\hat{S}$  with at least 18.53% higher than with ND in all cases tested except the *in silico* case (see Fig. 6). Both ND and  $\ell_1$ -min performed better than Bayesian network approach whose  $\xi$ -scores were 14.891, 0.029, 0.0001, respectively, for the three data sets<sup>8</sup>. In terms of  $\rho$ -score (Eq. (29)), which quantifies the false positive rates (i. e., the specificity),  $\ell_1$ -min approach reduces  $\rho$  by 2-3 orders compared to ND in all cases. These results provide a strong evidence for the relevance of the  $\ell_1$ -min approach for network structure inference. In terms of sparsity,  $\ell_1$ -min approach increased the Hoyer measure by about 20% in most cases, and were much closer to the Hoyer measures of the gold-standard network, compared to ND.

As noted earlier for *in silico* data, although the  $\rho$ -score with  $\ell_1$ -min was at least 1160% lower (i.e., higher specificity) and Hoyer was 33% higher (i.e., higher sparsity), the  $\xi$ -score was slightly (10%) lower than with ND. The lower  $\xi$ -score for  $\ell_1$ -min is perhaps a consequence of the method being susceptible to over-specification of the noise level. In this context, it must be noted that the solutions from both ND and  $\ell_1$ -min can replicate the observed total influence  $G$  within a specified bound (as total perturbation). However, the solutions from  $\ell_1$ -min tend to be much sparser and have lower false positive rate. Given that there were only 805 sample measurements to reconstruct  $G$  matrices for 1643 nodes in the *in silico* network, it is highly likely that several dynamic modes (degree of freedom) are not observable from the data. Therefore,  $\ell_1$ -min generated a much sparser network which, by formulation, is guaranteed to be adequate to capture the observed modes of the dynamics within the specified total perturbation limits. The ND derived networks for *in silico* and other cases that have higher  $\xi$ -score, intriguingly, were consistently found to have much lower Hoyer score (hence sparsity) even compared to the specified total influence matrix. Thus,  $\ell_1$ -min-generated solutions provide significant improvement in specificity, although the sensitivity at times were found to be slightly lower than with ND.

Averaging improves the  $\xi$ -scores (Eq. (31)) with all methods by at most 10%. This is perhaps due to the near-stationarity of the total influence matrix  $G$ , when computed using data over long time windows that smooths out various higher order transient effects. Also, one may note that the averaging makes the network inferred from ND less sparse than without averaging. This is because under noise, transients and data sparsity, ND yields vastly different network topologies depending on the samples employed. Averaging over these vastly different networks causes a reduction in sparsity. These results, taken together suggest that the  $\ell_1$ -min approach is perhaps the best known means to provide specificity for network inference from transient and noisy data. The utility of the approach would be to provide a minimal set of arcs (dynamic couplings or direct influences) to be considered for further network dynamics reconstruction applications.

**Discussion and Concluding remarks.** In this paper, we have investigated a method to robustly infer the structure of a network representing a sparse dynamical system from noisy, transient time series data. When the noise level is known, the  $\ell_1$ -min formulation employing our theoretical formula for the bound on total perturbation improves the recently reported NDs in terms of both accuracy and sparsity. When the noise level is unknown, we have shown that by averaging the networks inferred from different time points or conditions, the inference of network structure of real world processes becomes highly plausible.

Pertinently, for most real world processes, the total influence is not known a priori; only the time series ensembles gathered under transient conditions are available (e.g., gene expression microarray data<sup>8,49</sup>, protein-protein interaction data<sup>50</sup> as in the case of Michaelis-Menten dynamics). It has been noted that most of the earlier approaches present severe accuracy, noise sensitivity and/or numerical stability issues for such realistic scenarios. To overcome these limitations, we have investigated the  $\ell_1$ -min approach with a novel perturbation procedure for time series based network inference. Averaging over the solutions estimated at different time windows has



**Figure 6. Performance comparison of (1) original  $G$  matrix, (2) ND, (3) ND with averaging, (4)  $\ell_1$ -min and (5)  $\ell_1$ -min with averaging for the DREAM5 challenge datasets.** The total influence  $G$  matrix is estimated by Pearson correlation (blue/dark), Spearman correlation (red/light) and Mutual Information (green/light). Compared to ND, the prediction scores with  $\ell_1$ -min are increased by 23.94% (for  $G$  from Pearson correlation), 53.03% (for  $G$  from Spearman correlation) & 18.53% (for  $G$  from Mutual Information) for *E. Coli*, 89.09%, 249.7% & 116.74% for *S. cerevisiae*, respectively; the inference errors  $\rho$  (29) are reduced by 2 to 3 orders of magnitude in all cases; Hoyer measures are increased by 34%, 36.41% & 322.91% for *E. Coli*, 18.85%, 19.59% & 96.65% for *S. cerevisiae*, respectively. For *in silico* data, ND gives a solution with 11% higher prediction score but 33% less sparse than  $\ell_1$ -min approach. Averaging slightly improves the performance of all methods (<10%).

been shown to allow inference of the structure for complex real world networks, especially when the noise levels are unknown or cannot be accurately estimated.

Next, we have applied our method to three benchmark systems: a sparse scale-free network<sup>51</sup> with a specified noise level and the total influence between any two nodes given, a genetic regulatory network model formulated in terms of a system of Hill-type differential equations<sup>27</sup>, and GRNs of DREAM5 challenge<sup>46</sup>. These analyses suggest that our proposed bounds on the constraints for the  $\ell_1$ -min formulation, extracted from a few time series samples acquired under transient conditions, are of the same order (i.e., they closely envelop) with the constraints estimated based on the full knowledge of the noise level. The  $\ell_1$ -min formulation reduces the inference errors defined in (31) and (29) by 18.53% and 2 to 3 orders of magnitude, respectively, and improves the sparsity of the solution (measured in terms of Hoyer sparsity measure) by 15.9%, in comparison with conventional approaches including various versions of dynamic Bayesian approaches for network inference as well as ND. If instead of the total influence, only the time series gathered under transient conditions is provided, such as in the case of Michaelis-Menten dynamics,  $\ell_1$ -min approach achieves a 4 order reduction in inference error compared to MRA.

These theoretical and numerical studies suggest that our proposed method can be employed to effectively infer the presence of dynamic coupling (i.e., arc set or the direct influence in a dynamic network) based on sparse samples.

As with any network reconstruction approach, the method assumes that the time series realizations taken together can adequately mirror the salient dynamic regimes of the underlying process<sup>52</sup>, and as noted earlier, the approach is restricted to ensuring high levels of specificity and not sensitivity in identifying the direct influences. Additionally, while the approach is fairly robust to the presence of noise, the estimates  $\hat{s}_{ij}$  from the averaging procedure for the arcs with  $s_{ij}^0 = 0$  is guaranteed to converge to zero only in the presence of additive noise. More specifically, one of the following conditions need to hold for the approach to be applicable: (1) the governing equation of the process dynamics is specified, so that  $G(t)$  or  $R(t)$  can be constructed; (2) one or more realizations of  $G(t)$  (based on ND or silencing method) or  $R(t)$  (based on MRA) are given. In our experience, 30 realizations ensured the convergence of the averaging method; (3) one realization of a  $n$ -dimensional time series is available for estimating  $G(t)$  using various alternative methods outlined in Feizi *et al.*'s<sup>26</sup> or  $n^2$  time series realizations with the same initial condition are available for estimating  $R(t)$  using Eq. (17). Note that Scenario 1 is useful only for applications such as to investigate if there exists a more compact (sparser) network representation to capture the specified process dynamics. In Scenarios 2 and 3, we assume that the noise level or its lower limit is known, and adequate number of realizations are available to ensure convergence of the averaging method. In scenario 3, Eq. (17) yields a finite space-time approximation of the partial derivatives  $\frac{\partial x_i}{\partial p_j}, \frac{\partial \dot{x}_i}{\partial p_j}$ . They are estimated by perturbing the parameters  $p_j$  and keeping the initial condition the same for two time series signals. The length of the time series in this case can be really small, or it can just be samples taken over multiple (roughly 30), short (can be even 2 samples) time windows. However, the time steps (or sampling interval) in each time window must be small enough to ensure that  $R(t)$  values locally converge. Sensitivity of the network inference performance to time step size, however, needs further investigation.

Efforts are underway to address some of the  $\ell_1$ -min aforementioned limitations. We are investigating a two-stage approach to recover local nonlinear dynamics from sparse time series data. For future research, we will consider a more realistic scenario where not all state variables can be measured. In GRN inference, for example, only the outputs/activations of only those genes that have been discovered are measured. However, unknown genes might have significant influence on the network structure. Removing the effects of unmeasured variables, when combined with the method proposed in this paper, will lead to a more advanced network inference method.

## References

- Chen, T., He, H. L. & Church, G. M. Modeling gene expression with differential equations. In *Pacific Symposium on Biocomputing* vol. 4, 4 (1999).
- Hecker, M., Lambeck, S., Toepfer, S., Van Someren, E. & Guthke, R. Gene regulatory network inference: data integration in dynamic models - a review. *Biosystems* **96**, 86–103 (2009).
- Schweitzer, F. *et al.* Economic networks: The new challenges. *Science* **325**, 422 (2009).
- Carrington, P. J., Scott, J. & Wasserman, S. *Models and Methods in Social Network Analysis*, vol. 28 (Cambridge University Press, 2005).
- Guimera, R., Mossa, S., Turtschi, A. & Amaral, L. N. The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proceedings of the National Academy of Sciences* **102**, 7794–7799 (2005).
- Newman, M. E. The structure and function of complex networks. *SIAM Review* **45**, 167–256 (2003).
- Vogelstein, B., Lane, D. & Levine, A. J. Surfing the p53 network. *Nature* **408**, 307–310 (2000).
- Marbach, D. *et al.* Wisdom of crowds for robust gene network inference. *Nature Methods* **9**, 796–804 (2012).
- De Smet, R. & Marchal, K. Advantages and limitations of current network inference methods. *Nature Reviews Microbiology* **8**, 717–729 (2010).
- Marbach, D. *et al.* Revealing strengths and weaknesses of methods for gene network inference. *Proceedings of the National Academy of Sciences* **107**, 6286–6291 (2010).
- Faisal, F. E. & Milenković, T. Dynamic networks reveal key players in aging. *Bioinformatics* **30**, 1721–1729 (2014).
- Žitnik, M. & Zupan, B. Gene network inference by probabilistic scoring of relationships from a factorized model of interactions. *Bioinformatics* **30**, i246–i254 (2014).
- Tang, Q., Sun, S. & Xu, J. Learning scale-free networks by dynamic node specific degree prior. In *Proceedings of The 32nd International Conference on Machine Learning*, 2247–2255 (2015).
- Chiuso, A. & Pillonetto, G. A bayesian approach to sparse dynamic network identification. *Automatica* **48**, 1553–1565 (2012).
- Friedman, N., Linial, M., Nachman, I. & Pe'er, D. Using bayesian networks to analyze expression data. *Journal of Computational Biology* **7**, 601–620 (2000).
- Friedman, N. Inferring cellular networks using probabilistic graphical models. *Science* **303**, 799–805 (2004).
- Zou, M. & Conzen, S. D. A new dynamic bayesian network (dbn) approach for identifying gene regulatory networks from time course microarray data. *Bioinformatics* **21**, 71–79 (2005).
- Young, W. C., Raftery, A. E. & Yeung, K. Y. Fast bayesian inference for gene regulatory networks using scanbma. *BMC Systems Biology* **8**, 47 (2014).
- Hill, S. M. *et al.* Bayesian inference of signaling network topology in a cancer cell line. *Bioinformatics* **28**, 2804–2810 (2012).
- Seth, A. K. A matlab toolbox for granger causal connectivity analysis. *Journal of Neuroscience Methods* **186**, 262–273 (2010).
- Basu, S., Shojaie, A. & Michailidis, G. Network granger causality with inherent grouping structure. *Journal of Machine Learning Research* **16**, 417–453 (2015).
- Bolstad, A., Van Veen, B. D. & Nowak, R. Causal network inference via group sparse regularization. *IEEE Transactions on Signal Processing* **59**, 2628–2641 (2011).
- Haufe, S., Nolte, G., Müller, K.-R. & Krämer, N. Sparse causal discovery in multivariate time series. *JMLR W&CP* **6**, 97–106 (2010).
- Lozano, A. C., Abe, N., Liu, Y. & Rosset, S. Grouped graphical granger modeling for gene expression regulatory networks discovery. *Bioinformatics* **25**, i110–i118 (2009).
- De La Fuente, A., Bing, N., Hoeschele, I. & Mendes, P. Discovery of meaningful associations in genomic data using partial correlation coefficients. *Bioinformatics* **20**, 3565–3574 (2004).
- Feizi, S., Marbach, D., Médard, M. & Kellis, M. Network deconvolution as a general method to distinguish direct dependencies in networks. *Nature Biotechnology* **31**, 726–733 (2013).

27. Barzel, B. & Barabási, A.-L. Network link prediction by global silencing of indirect correlations. *Nature Biotechnology* **31**, 720–725 (2013).
28. Ebert-Uphoff, I. & Deng, Y. Causal discovery for climate research using graphical models. *Journal of Climate* **25**, 5648–5665 (2012).
29. Runge, J., Heitzig, J., Petoukhov, V. & Kurths, J. Escaping the curse of dimensionality in estimating multivariate transfer entropy. *Physical Review Letters* **108**, 258701 (2012).
30. Runge, J. *et al.* Identifying causal gateways and mediators in complex spatio-temporal systems. *Nature Communications* **6** (2015).
31. Runge, J., Petoukhov, V. & Kurths, J. Quantifying the strength and delay of climatic interactions: the ambiguities of cross correlation and a novel measure based on graphical models. *Journal of Climate* **27**, 720–739 (2014).
32. Kholodenko, B. N. *et al.* Untangling the wires: a strategy to trace functional interactions in signaling and gene networks. *Proceedings of the National Academy of Sciences* **99**, 12841–12846 (2002).
33. Sontag, E., Kiyatkin, A. & Kholodenko, B. N. Inferring dynamic architecture of cellular networks using time series of gene expression, protein and metabolite data. *Bioinformatics* **20**, 1877–1886 (2004).
34. Wang, W. X., Yang, R., Lai, Y. C., Kovanis, V. & Grebogi, C. Predicting catastrophes in nonlinear dynamical systems by compressive sensing. *Physical Review Letters* **106**, 154101 (2011).
35. Napolitano, D. & Sauer, T. D. Reconstructing the topology of sparsely connected dynamical networks. *Physical Review E* **77**, 026103 (2008).
36. Wang, W.-X., Yang, R., Lai, Y.-C., Kovanis, V. & Harrison, M. A. F. Time-series-based prediction of complex oscillator networks via compressive sensing. *EPL (Europhysics Letters)* **94**, 48006 (2011).
37. Boccalletti, S., Latora, V., Moreno, Y., Chavez, M. & Hwang, D.-U. Complex networks: Structure and dynamics. *Physics Reports* **424**, 175–308 (2006).
38. Candes, E. & Romberg, J.  $l_1$ -magic: Recovery of sparse signals via convex programming (2005), (Date of access: 03/05/2014). Available at: <http://users.ece.gatech.edu/justin/l1magic/>.
39. Gurobi Optimization, I. Gurobi optimizer reference manual (2014), (Date of access: 02/03/2014). Available at: <http://www.gurobi.com>.
40. Herman, M. A. & Strohmer, T. General deviants: An analysis of perturbations in compressed sensing. *IEEE Journal of Selected Topics in Signal Processing* **4**, 342–349 (2010).
41. Horn, R. A. & Johnson, C. R. *Matrix Analysis* (Cambridge University Press, 1985).
42. Barzel, B. & Barabási, A.-L. Network link prediction by global silencing of indirect correlations. *Nature Biotechnology* **31**, 720–725 (2013).
43. Karlebach, G. & Shamir, R. Modelling and analysis of gene regulatory networks. *Nature Reviews Molecular Cell Biology* **9**, 770–780 (2008).
44. Alon, U. *An Introduction to Systems Biology: Design Principles of Biological Circuits* (CRC press, 2006).
45. Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabási, A.-L. The large-scale organization of metabolic networks. *Nature* **407**, 651–654 (2000).
46. Stolovitzky, G., Monroe, D. & Califano, A. Dialogue on reverse-engineering assessment and methods. *Annals of the New York Academy of Sciences* **1115**, 1–22 (2007).
47. Muchnik, L. Complex networks package for matlab (version 1.6) (2013), (Date of access: 12/08/2014). Available at: <http://www.levmuchnik.net/Content/Networks/ComplexNetworksPackage.html>.
48. Hoyer, P. O. Non-negative matrix factorization with sparseness constraints. *The Journal of Machine Learning Research* **5**, 1457–1469 (2004).
49. Arbeitman, M. N. *et al.* Gene expression during the life cycle of drosophila melanogaster. *Science* **297**, 2270–2275 (2002).
50. Pagel, P. *et al.* The mips mammalian protein-protein interaction database. *Bioinformatics* **21**, 832–834 (2005).
51. Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
52. Cheng, D. *et al.*'s Time series forecasting for nonlinear and non-stationary processes: a review and comparative study. *IIE Transactions* **47**, 1053–1071 (2015).

## Acknowledgements

The authors thank the anonymous reviewers for their constructive comments that have helped improve the manuscript. They also acknowledge the National Science Foundation CMMI division (Grants 1437139 and 1432914) for the generous support of this research. The open access publishing fees for this article have been covered by the Texas A&M University Online Access to Knowledge (OAK) Fund, supported by the University Libraries and the Office of the Vice President for Research.

## Author Contributions

H.M.T. and S.T.S.B. designed and performed the research, analyzed the results and wrote the paper.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Tran, H. M. and Bukkapatnam, S. T.S. Inferring sparse networks for noisy transient processes. *Sci. Rep.* **6**, 21963; doi: 10.1038/srep21963 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>