RESEARCH

Enhanced network inference from sparse incomplete time series through automatically adapted L_1 regularization

Zhongqi Cai[1,2*], Enrico Gerding[1] and Markus Brede[1]

*Correspondence:
Zhongqi.cai@soton.ac.uk

[1] School of Electronics
and Computer Science,
University of Southampton,
Southampton, UK
[2] School of Computer
and Information Engineering,
Xiamen University of Technology,
Xiamen, China

Abstract

Reconstructing dynamics of complex systems from sparse, incomplete time series data is a challenging problem with applications in various domains. Here, we develop an iterative heuristic method to infer the underlying network structure and parameters governed by Ising dynamics from incomplete spin configurations based on sparse and small-sized samples. Our method iterates between imputing missing spin states given current coupling strengths and re-estimating couplings from completed spin state data. Central to our approach is the novel application of adaptive l_1 regularization on updating coupling strengths, which features an automatic adjustment of the regularization strength throughout the iterative inference process. By doing so, we aim at preventing over-fitting and enforcing the sparsity of couplings without access to ground truth parameters. We demonstrate that this approach accurately recovers parameters and imputes missing spins even with substantial missing data and short time series, providing improvements in the inference of Ising model parameters even for relatively small sample sizes.

Keywords: Network inference, Inverse ising model, Complex networks

Introduction

Parameter inference from observed time-series data of the dynamics of complex systems is an important research topic in the statistical mechanics community (Zdeborová and Krzakala 2016). A widely recognized and well-established framework for statistical inference of complex-networked systems is the inverse kinetic Ising model (Roudi and Hertz 2011b). This model aims at reconstructing pairwise interactions between spins in the Ising model from time-series observations, such as spin states. With its versatility, the inverse kinetic Ising model has found applications in various domains, including neuroscience (Zeng et al. 2011; Tyrcha et al. 2013; Roudi et al. 2015; Donner and Opper 2017), computational biology (Nguyen et al. 2017; Cresswell-Clay and Periwal 2021), economics (Hoang et al. 2019a; Lee et al. 2021), and social sciences (Bisconti et al. 2015).

Due to the growing need to analyze high-throughput data using statistical models, significant efforts have been devoted to devising efficient approximate inference techniques for the inverse kinetic Ising model. These efforts have led to the development of

utilizing mean-field approximations (Roudi and Hertz 2011b; Mézard and Sakellariou 2011), the Thouless-Anderson-Palmer (TAP) method (Roudi and Hertz 2011a), cavity analysis (Zhang 2012), and the expectation-reflection approach (Hoang et al. 2019b). However, these studies are based on the assumption that the system dynamics are fully observed. In contrast, the complexity of real-world systems and limitations in data sampling techniques often make it difficult to obtain complete observations. For example, observing every neuron in the brain and monitoring their complete spiking activities is almost impossible (Soudry et al. 2015). Furthermore, considering the effects of hidden units in the statistical model can help improve the inference quality (Pearl 2000). Therefore, recent research has primarily focused on the inference with incomplete data (Dunn and Roudi 2013; Tyrcha and Hertz 2014; Bachschmid-Romano and Opper 2014; Battistin et al. 2015; Dunn and Battistin 2017; Hoang et al. 2019a; Campajola et al. 2019; Lee et al. 2021; Gemao et al. 2021).

One commonly used setting for inference with incomplete data is the assumption of the existence of hidden spins, whose states can never be observed (Dunn and Roudi 2013; Tyrcha and Hertz 2014; Bachschmid-Romano and Opper 2014; Battistin et al. 2015; Dunn and Battistin 2017; Hoang et al. 2019a; Gemao et al. 2021). Of the studies above, Dunn and Battistin (2017) and Gemao et al. (2021) emphasize the impacts of hidden spins on dynamics of observed spins. However, they only provide estimates for coupling strengths between observed spins. In contrast, Hoang et al. (2019a) tackle a broader aspect of this challenge by inferring not just the interaction strengths but also the states and the number of hidden spins. Similarly, other research efforts, including those by Dunn and Roudi (2013), Tyrcha and Hertz (2014), Bachschmid-Romano and Opper (2014), Battistin et al. (2015), employ an iterative estimation procedure to infer the coupling strengths between hidden spins. In more detail, their procedures alternate between two steps: estimating the states of hidden variables given current parameters, and optimizing model parameters given current hidden variable estimates.

On top of assuming the existence of hidden units, a more practical setting in the context of incomplete data is to consider that even visible spins may not be consistently observable throughout the dynamics. This consideration holds significant relevance in domains such as neuroscience (Soudry et al. 2015), finance (Mazzarisi et al. 2020), and the social sciences (Zipkin et al. 2016), where partially observable units are prevalent. To the best of our knowledge, there are only two previous studies investigating the inverse kinetic Ising model with partially observable spins using the framework of iterative estimations . Specifically, the work of Campajola et al. (2019) tackles this issue via employing the mean-field and TAP approximations. However, the mean-field method and TAP approach both assume a priori that interactions are weak and dense. As sparse networks, with only a fraction of nodes significantly connected while the majority have limited or no connections, are widely observed in real-world systems (Han et al. 2015; Mirshahvalad et al. 2012; Singh and Humphries 2015), it is important to devise an alternative method that does not impose the limitation of dense connectivity. The other study that considers network inference in the existence of partially observable spins is Lee et al. (2021). In this work, the authors iterate between the step of generating estimations for hidden spin states, and the step of optimizing couplings by utilizing linear regression. However, through our experiments in Sect. 3.1, we observed that this algorithm converges most effectively with intermediate sample sizes. For small

sizes, a specific threshold for the stopping criterion must be designed to ensure the convergence of their algorithm. Nevertheless, in many applications, it is common to encounter situations where experimental data may not be sufficiently large to reconstruct the interaction network of a given system (Hoang et al. 2019b). Hence, more research is needed to improve methods for the inference of networks from short incomplete time series.

Motivated by the above-mentioned challenges of network sparsity and limited data availability, we propose an iterative heuristic method to reconstruct the couplings and hidden spin states of the Ising model from incomplete time series data in the context of sparse networks and small sample sizes. Our work focuses on the setting where we know all the spins, but some of the visible spins can have unobserved states, similar to the scenarios described in Campajola et al. (2019), Lee et al. (2021). From the dataset we can identify the set of unobserved spin states and its size. Drawing inspiration from Lee et al. (2021), our algorithm consists of two main components. The first part employs iterative estimations and optimizations to deduce missing spin states and determine coupling strengths, gradually refining these parameters towards their true values. However, due to the small size of our datasets, this process is susceptible to overfitting. To address this issue and impose sparsity on the network's structure, the second part of our algorithm incorporates l_1 regularization. While existing methods often rely on manually tuned hyper-parameters by using cross-validation or a priori knowledge about network sparsity levels, our approach automates the adjustment of the l_1 regularization strength based on the convergence of the iteration process. This automation is a key methodological novelty in our approach, as it eliminates the dependency on known network structures to manually tune the hyper-parameters, offering a robust solution when such ground truths are unavailable.

Our contributions are thus as follows: (1) We propose an iterative heuristic that can reconstruct sparse Ising networks from incomplete time series data from small sample sizes. (2) We introduce an automatic regularization technique to prevent over-fitting during the inference process when the ground truth is unavailable. (3) We demonstrate through simulations that our method outperforms the state-of-the-art method in inferring network structure especially in the setting of small samples, via an automated regularization approach without requiring ground truth for hyper-parameter tuning.

The remainder of this paper is organized as follows. In Sect. 2, we provide a formal description for the framework of network inference with missing data and the algorithm to solve this problem. In Sect. 3, we provide numerical results demonstrating the validity of our proposed approach on simulated sparse network data across varying settings such as sample sizes, missing data percentages, network connectivity, and network types. We demonstrate the effectiveness of the automatic l_1 regularization in preventing over-fitting and enhancing inference accuracy compared to prior methods. In Sect. 4, we summarise the main findings and contributions and discuss ideas for future work.

Model description and methods

Consider the stochastic dynamics of an Ising system with N binary spins. Each spin has two possible states at time t, denoted as $\sigma_i(t) = \pm 1$ for $i = 1, \ldots, N$. Following the discrete-time and synchronously updated Glauber algorithm (Glauber 1963), the system evolves according to the conditional probability at time steps $t = 1, \ldots, T$:

$$P(\sigma_i(t+1) \mid \{\sigma_j(t)\}_{j=1}^N) = \frac{e^{\sigma_i(t+1)H_i(t)}}{2\cosh(H_i(t))}. \tag{1}$$

Here, the local field $H_i(t) = \sum_j w_{ij}\sigma_j(t)$ represents the weighted sum of interactions between spin i and each of its connected spins j at time t, where w_{ij} is the coupling strength from spin j to spin i.

In this paper, instead of assuming all spin states are always observable during the whole observation period T, we consider a more realistic scenario of partial observability, in which some of the spin states are unobserved (or missing). More specifically, we use the set $\mathcal{U}$ to represent these unobserved data points. To distinguish between unobserved and observable spin states, we introduce notations $\sigma_i^u(t) \in \mathcal{U}$ for unobserved states and $\sigma_i^o(t) \notin \mathcal{U}$ for observed states. When the context is clear, we drop the superscripts u and o for brevity. Given the partially observed time series of spin states, we aim at inferring the model parameters $\theta = \{w_{ij}\}_{i,j=1}^N$, i.e. the coupling network between the spins, which we assume to be constant over time.

In the presence of latent variables (i.e., $\sigma_i^u(t)$), the iterative estimation procedure is a commonly applied technique for finding maximum likelihood estimates of parameters in statistical models (Hastie et al. 2009). Here, following Lee et al. (2021), we extend the use of their iterative solution to compute coupling strengths from incomplete spin configurations for small samples and sparse network structures. Our approach contains four key components:

(i)   Estimation (E) step: Estimate missing spin states given observed spin states and the current estimate of couplings;

(ii) Maximization (M) step: Recalibrate couplings based on completed spin state configurations after applying the E-step;

(iii) Automatic l_1 regularization: Encourage sparsity and prevent over-fitting in the M-step with the regularization hyper-parameter automatically determined;

(iv) Stopping criterion: Halt iterations between the E and M steps to further prevent over-fitting.

In the above, we provided a high-level overview of the four key components. Next, we detail the specific procedures for the E-step, M-step, regularization, and stopping criterion, followed by a summary of the overall algorithm.

E-step

In the E-step, the missing spin states $\sigma_i^u(t) \in \mathcal{U}$ are estimated by calculating the likelihoods of spins taking the states ± 1. More specifically, the values of $\sigma_i^u(t)$ are stochastically updated to 1 with probability $\mathcal{L}_{i,t}^{+1}/\left(\mathcal{L}_{i,t}^{+1} + \mathcal{L}_{i,t}^{-1}\right)$, and to -1 with probability $\mathcal{L}_{i,t}^{-1}/\left(\mathcal{L}_{i,t}^{+1} + \mathcal{L}_{i,t}^{-1}\right)$. Here, $\mathcal{L}_{i,t}^{\pm 1}$ stands for the likelihood of $\sigma_i^u(t)$ taking ± 1, computed as:

$$\begin{aligned}
\mathcal{L}_{i,t}^{\pm 1} = &P\left[\sigma_i^u(t) = \pm 1 \mid \{\sigma_j(t-1)\}_{j=1}^N, \theta\right] \\
&\times \prod_{j=1}^N P\left[\sigma_j(t+1) \mid \{\sigma_1(t), \ldots, \sigma_i^u(t) = \pm 1, \ldots, \sigma_N(t)\}, \theta\right].
\end{aligned} \tag{2}$$

The first term of the right-hand side of Eq. (2) represents the one-step-backward likelihood of $\sigma_i^u(t)$ taking the values ± 1. The second term stands for the one-step-forward likelihood. For the special case of $t = 1$ or $t = T$, we only include the forward or backward likelihood.

M-step with expectation reflection

In the M-step, we integrate the observed spin states with the imputed missing states from the E-step to compile complete spin configurations, represented as $\mathcal{S} = \{\sigma_j(t) \mid 1 \leq j \leq N, 1 \leq t \leq T\}$. At this stage, the distinction between observed and unobserved data points becomes unnecessary for our analysis, as we operate with these full configurations to update the coupling strengths. Consequently, we simplify our notation by omitting the superscripts u and o.

The expectation-reflection algorithm then consists of three main components: i) optimization of coupling strengths w_{ij}, ii) updating of local fields $H_i(t)$, and iii) a stopping criterion to determine when to terminate the iterations between optimizing w_{ij} and updating $H_i(t)$. In this study, we adopt the expectation-reflection algorithm for inferring coupling strengths for two main reasons. First, its computational efficiency surpasses traditional methods like maximum likelihood estimation, primarily due to its use of multiplicative weights. Second, the expectation-reflection algorithm exhibits enhanced performance through its stopping criterion. This criterion, by assessing mismatches between observed spins and model expectations, curtails unnecessary iterations. Such a mechanism can not only be applied to improve the expectation-reflection algorithm's performance but also has the potential to augment maximum likelihood estimation, especially in data-limited scenarios as shown in Hoang et al. (2019b). Overall, our selection of the expectation-reflection algorithm is driven by its efficiency and its effectiveness in addressing the challenges associated with data scarcity, as outlined in our research.

To optimize the parameters $\{w_{ij}\}_{i,j=1}^N$ across all times t, we first sum $H_i(t)$ over time t and multiply this by spin fluctuations $\delta\sigma_k(t) = \sigma_k(t) - \langle\sigma_k\rangle$, with $\langle f \rangle = 1/T \sum_{t=1}^T f(t)$ denoting the time average of function f:

$$\sum_t H_i(t)\delta\sigma_k(t) = \sum_t \sum_j w_{ij}\sigma_j(t)\delta\sigma_k(t) \tag{3}$$

By replacing $H_i(t)$ with its fluctuation $\delta H_i(t)$ and mean $\langle H_i\rangle$ (i.e., $H_i(t) = \delta H_i(t) + \langle H_i\rangle$) in Eq. (3), we obtain

$$\sum_t \delta H_i(t)\delta\sigma_k(t) = \sum_t \sum_j w_{ij}\delta\sigma_j(t)\delta\sigma_k(t). \tag{4}$$

This leads to a simplified expression:

$$\langle\delta H_i\delta\sigma_k\rangle = \sum_j w_{ij}\langle\delta\sigma_j\delta\sigma_k\rangle, \tag{5}$$

where $\langle\delta\sigma_j\delta\sigma_k\rangle$ is the covariance matrix. Note that, in the absence of l_1 regularization, the coupling strengths w_{ij} can be obtained via:

$$w_{ij} = \sum_k \langle \delta H_i \delta \sigma_k \rangle \left[\delta \sigma_k \delta \sigma_j \right]^{-1}. \tag{6}$$

As mentioned above, to prevent over-fitting and encourage sparsity given the small sample sizes, we incorporate l_1 regularization (Liu and Ihler [2011]). Therefore, instead of optimizing Eq. (5) to find the best $\{w_{ij}\}_{i,j=1}^N$, we optimize a regularized loss function:

$$\sum_k \left(\langle \delta H_i \delta \sigma_k \rangle - \sum_j w_{ij} \langle \delta \sigma_j \delta \sigma_k \rangle \right)^2 + \lambda \sum_j |w_{ij}| \tag{7}$$

where λ controls the regularization strength. A larger value of λ will lead to stronger regularization, which tends to drive more of the coupling strengths towards zero, promoting sparsity in the model. Conversely, a smaller λ will result in weaker regularization, allowing the model to have larger coupling strengths. The process of updating the coupling strengths under l_1 regularization is carried out through the numerical optimization of the regularized loss function of Eq. (7). This optimization is specifically achieved using the coordinate descent method (Wu and Lange [2008]). Note that, the above requires a choice of the parameter λ. We will explain later how λ can be automatically adapted based on a criterion related to our algorithm's convergence in Sect. 2.3.

After obtaining the coupling strengths, the expectation-reflection employs a two-step process to update the local field $H_i(t)$. A detailed deduction for the updating of $H_i(t)$ can be found in Appendix A. Initially, $H_i(t)$ is estimated using:

$$H_i(t) \leftarrow \sum_j w_{ij} \sigma_j(t). \tag{8}$$

Subsequently, $H_i(t)$ is refined to better align with the next time step's state $\sigma_i(t+1)$ through:

$$H_i(t) \leftarrow \frac{\sigma_i(t+1)}{E[\sigma_i(t+1) \mid H_i(t)]} H_i(t) = \sigma_i(t+1) \frac{H_i(t)}{\tanh H_i(t)}, \tag{9}$$

where $E[\sigma_i(t+1) \mid H_i(t)]$ represents the expected value of $\sigma_i(t+1)$ given $H_i(t)$. This step helps correct the sign of $H_i(t)$ and pull the expectation of $\sigma_i(t+1)$ closer to ± 1.

As overfitting poses a significant challenge in the context of inference with small sample sizes, it is important to determine an appropriate stopping criterion for optimizing w_{ij} and updating $H_i(t)$. In this regard, our focus shifts to the overall discrepancy between the observed values, denoted by $\sigma_i(t+1)$, and the model's predictions, $E[\sigma_i(t+1) \mid H_i(t)]$. This discrepancy is quantified as follows:

$$D_i = \frac{1}{T} \sum_t (\sigma_i(t+1) - E[\sigma_i(t+1) \mid H_i(t)])^2. \tag{10}$$

Following the heuristic proposed by Hoang et al. ([2019a]), we repeat optimizing Eq. (7) and updating $H_i(t)$ until D_i starts to increase. Note that the parameter updates given by Eqs. (7), (8) and (9) are completely independent of the computation of D_i. This crucial aspect allows the inference process to avoid overfitting when dealing with small sample sizes, as the minimization of D_i serves solely as a stopping criterion. This approach

is notably different from maximum likelihood methods, which directly minimize a cost function and terminate at the minimal cost configuration, even if it represents a local minimum rather than the global optimum. A more systematic comparison between the method of minimization of data and predictions mismatch and the maximum likelihood methods can be found in Hoang et al. (2019b).

Stopping criterion and automatic l_1 regularization

By executing E and M steps iteratively, the model parameter θ is expected to converge to a local maximum. However, iterating too often can worsen inference accuracy due to over-fitting, especially with small samples (Lee et al. 2021). Therefore, an appropriate stopping criterion is critical. To address this issue, we consider prediction consistency between observed and unobserved data, where the model should have similar performance in predicting the observed data and the unobserved data. Specifically, the model-data discrepancy for the observed data D_{obs} and the unobserved data D_{un} is determined by

$$D_{obs} = \frac{1}{NT - |\mathcal{U}|} \sum_{\{t+1,i|\sigma_i^o(t+1)\notin\mathcal{U}\}} (\sigma_i(t+1) - E[\sigma_i(t+1) \mid H_i(t)])^2 \tag{11}$$

and

$$D_{un} = \frac{1}{|\mathcal{U}|} \sum_{\{t+1,i|\sigma_i^u(t+1)\in\mathcal{U}\}} (\sigma_i(t+1) - E[\sigma_i(t+1) \mid H_i(t)])^2 \tag{12}$$

where, $|\mathcal{U}|$ stands for the number of unobserved data points. Typically, the iterative estimation procedure halts when $|D_{obs} - D_{un}| < \varepsilon$, where ε is a threshold that needs to be properly chosen to stop the estimation procedure at an appropriate time (Lee et al. 2021). However, as there is no ground truth for most real-world inference problems, it is challenging to determine the best choice of ε. To address this issue, together with the problem of determining the hyper-parameter of the l_1 regularization λ, we propose a heuristic that gets around the problem of finding the threshold ε. Instead, we propose an approach, which is based on the observation that best results from the iterated approach tend to be found at parameter values just before convergence is lost. Hence, we directly set $D_{obs} = D_{un}$ as the stopping criterion and, incrementally increasing λ from $\lambda = 0$, choose λ to be right at the threshold at which the algorithm changes from convergence to non-convergence. The effectiveness of this heuristic is systematically verified in Sect. 3.

In summary, the overall procedure for determining the model parameters θ involves these main steps:

  (i)   Start by setting the regularization parameter to $\lambda = 0$ and set the maximum number of iterations to $Q_{max} = 100$. (A detailed discussion for the choice of Q_{max} can be found in Sect. 3.2.)

  (ii)   Randomly initialize every unobserved spin state $\sigma_i^u(t) \in \mathcal{U}$ to to either $+1$ or -1, each with a 50% probability. Then, given the observed and randomly assigned spin states, we obtain a complete collection of binary time series: $\mathcal{S} = \{\sigma_j(t) \mid 1 \leq j \leq N, 1 \leq t \leq T\}$.

(iii) M-step: We then infer the model parameters θ via the above-mentioned expectation-reflection method.

    a)   We first initialize the local field $H_i(t)$ as $\sigma_i(t)$.

    b)   Then the coupling strength w_{ij} can be obtained via the regularized loss function of Eq. (7).

    c)   Update $H_i(t)$ according to Eqs. (8) and (9).

    d)   Repeat steps b) and c) iteratively until the discrepancy between true and expected values of the spin states (see Eq. (10)) is minimized.

(iv) E-step: Stochastically update the values of $\sigma_i^u(t)$ to 1 with the probability of $\mathcal{L}_{i,t}^{+1}/\left(\mathcal{L}_{i,t}^{+1} + \mathcal{L}_{i,t}^{-1}\right)$, and to -1 with probability $\mathcal{L}_{i,t}^{-1}/\left(\mathcal{L}_{i,t}^{+1} + \mathcal{L}_{i,t}^{-1}\right)$ according to Eq. (2).

(v) Repeat steps (iii) and (iv) until $D_{obs} = D_{un}$ (see Eqs. (11) and (12)) or the number of iterations exceed the maximum number of iterations Q_{max}.

(vi) The procedure is terminated if the number of iterations exceeds the maximum number of iterations Q_{max}. Otherwise, increase λ by a small increment of $\Delta\lambda$ and repeat steps (i) to (vi). This process continues, with λ incrementally increasing, until the iterative procedure fails to converge. The λ value used in the final converged iteration prior to non-convergence is then considered as the optimal λ for the l_1 regularization.

Furthermore, to evaluate our algorithm, we use the mean square error (MSE) between the inferred weights and the true weights:

$$MSE = N^{-1} \sum_{i,j} (w_{ij} - w_{ij}^*)^2 \tag{13}$$

where w_{ij} refers to the inferred linking weights from spin j to spin i and the w_{ij}^* represent the corresponding true values. We choose MSE because it is a standard metric, and is used widely for evaluating the accuracy of inferred network connections and weights (Hoang et al. 2019a; Lee et al. 2021). By benchmarking with MSE, we can directly compare the performance of our proposed approach against existing methods.

Results

In this section, we validate the applicability of our proposed algorithm for network inference on sparse networks for small sample sizes. In Sect. 3.1, we start by analyzing the performance of network inference without l_1 regularization. Results shown there provide a benchmark for later comparisons. In this section, we also present a counterintuitive observation, where diluting the quality of available information by adding empty rows to the time series is found to result in improved inference accuracy. We show that this is explained by over-fitting to noise in the case of limited data. These experiments set the stage for the necessity of l_1 regularization. Next, in Sect. 3.2, we present an initial exploration of the iterative heuristic including the procedure of automatically determining the regularization parameter. We then conduct comprehensive tests on the algorithm's

performance across various data availability scenarios, including sample sizes, amounts of missing data, average degrees $\langle k \rangle$ of the network, average connectivity strength, and network types in Sect. 3.3.

In our study, we employ a structured approach to generate simulated time series for network inference. This process begins with the construction of sparse networks using commonly studied network topologies. Specifically we use the Barabási-Albert (BA) model (Barabási and Albert 1999), the Erd?s-Rényi (ER) model (Newman et al. 2001), and the small-world (SW) model (Watts and Strogatz 1998) to build sparse networks of size $N = 100$. Once the network structure is established, we assign link weights to the existing connections. More specifically, link weights are drawn from a normal distribution with a mean of 0 and a standard deviation of $g/\sqrt{N}$. Here, the parameter g influences the connectivity strength.

Subsequent to the network and weight setup, we simulate the Ising dynamics on these networks using Glauber dynamics. For the initial conditions of the spin states, we employ a random assignment where 50% of the states are set to -1 and the remaining 50% to $+1$, ensuring a balanced start for each simulation. Crucially, the recording of the time series begins from time 0, capturing the entire evolution of the system from the very start.

Network inference without l_1 regularization for small sample sizes

We first explore network inference based on small sample sizes without applying the automatically determined l_1 regularization. In this case, our approach reduces to the algorithm introduced by Lee et al. (2021), which provides state-of-the-art performance in reconstructing coupling strengths from limited observations with missing data. A major drawback of their algorithm is the need to manually tune the threshold ε of $|D_{obs} - D_{un}| < \varepsilon$ to ensure the algorithm stops at the right time. In their setting with observation lengths $T \geq 2500$, they choose a specific value of $\varepsilon = 0.01$ as the stopping threshold. However, the appropriate threshold is highly dependent on the amount of data available. In Fig. 1a, we demonstrate how the optimal threshold ε^* is influenced by varying observation lengths T in contexts with different percentages of missing data. The optimal threshold ε^* is determined as the specific value of $|D_{obs} - D_{un}|$ at which the MSE is minimized during the algorithm's iterations. In Fig. 1a, with extremely limited data $T \leq 1500$, we observe higher optimal thresholds ($\varepsilon^* > 0.01$) in order to halt the iterations earlier.

Correspondingly, in Fig. 1b, we compare the MSE when stopping at the optimal threshold ε^* (as determined above) versus at $|D_{obs} - D_{un}| < 0.01$ for varying observation lengths T. The relative MSE on the y-axis shows the percentage increase in MSE from using 0.01 as the threshold compared to the optimal threshold. This comparison reveals a key insight: Whereas performance losses for longer time series are not significant, for smaller values of observation lengths T, adhering to a fixed threshold of $|D_{obs} - D_{un}| < 0.01$ results in excessive iterations and decidedly suboptimal performance. However, it is unrealistic to design a specific ε value for each T, especially without knowledge of the ground truth. Therefore, for small lengths of the time series T, a new stopping criterion that eliminates manually assigning ε is required.
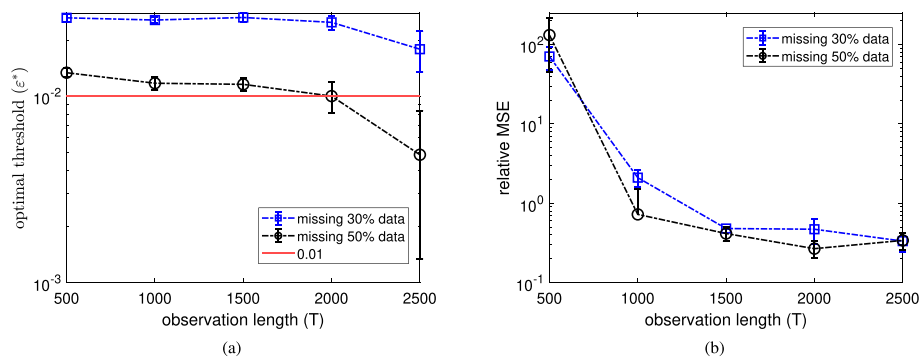
Fig. 1 **a** Dependence of optimal threshold ε^* on varying observation lengths T. The optimal threshold ε^* is determined as the specific value of $|D_{obs} - D_{un}|$ at which the MSE is minimized during the algorithm's iterations. **b** Dependence of relative MSE on varying observation lengths T. The relative MSE on the y-axis is calculated as the difference between the minimal MSE at $|D_{obs} - D_{un}| < \varepsilon^*$ and the MSE at $|D_{obs} - D_{un}| < 0.01$, divided by the minimal MSE. Here, we use a setting of BA networks with network size $N = 100$, average degree $\langle k \rangle = 10$, and connectivity strength coefficient $g = 1$. The blue squares represent the scenarios of 30% randomly missing data, while the black circles are for 50% randomly missing data. The results are based on 10 repetitions of the experiments, and error bars indicate 95% confidence intervals

To further demonstrate the over-fitting challenges with limited samples, we explore the impact of supplementing the dataset with additional empty rows that do not contain any spin state information. Specifically, these empty rows are added to the end of the original dataset, and they represent time steps where the spin states for all spins in the network are unobserved or missing. Somewhat surprisingly, we note that adding empty rows significantly improves the inference accuracy, as shown in Fig. 2a. More specifically, Fig. 2a shows the dependence of MSE on the number of additional empty rows. The inference is carried out on a dataset containing two parts: a 90% visible data part with length $T_1 = 500$ plus a number of empty rows as shown on the x-axis. The unobserved data points in T_1 are chosen randomly. Here, padding with empty rows leads to monotonic improvement despite the fact that the appended rows do not provide any extra information. This highlights the tendency of the unregularized algorithm of Lee et al. (2021) to capitalize on spurious noise patterns, fitting to spurious patterns in limited data rather than real underlying correlations.

To directly illustrate this over-fitting behavior, Fig. 2b, c compare true weights versus those inferred by the algorithm of Lee et al. (2021) without and with 500 added empty rows. Without extra rows in Fig. 2b, the lack of data causes inferred weights to be drastically overestimated in early iterations. However, with empty padding in Fig. 2c, weights initially underestimate true values as the algorithm attempts to latch onto non-existent signals in the meaningless rows. These experiments conclusively demonstrate the need for safeguards such as $l1$ regularization to constrain over-fitting.

Network inference with l_1 regularization

As discussed above, network inference without l_1 regularization can lead to over-fitting, especially for small data sets. To address this, we now examine the performance when including automatic l_1 regularization through the heuristic approach proposed in Sect. 2. Notice that, l_2 can also be used to avoid overfitting. However, it typically results in models where most coefficients are non-zero but small. This attribute makes l_2 regularization
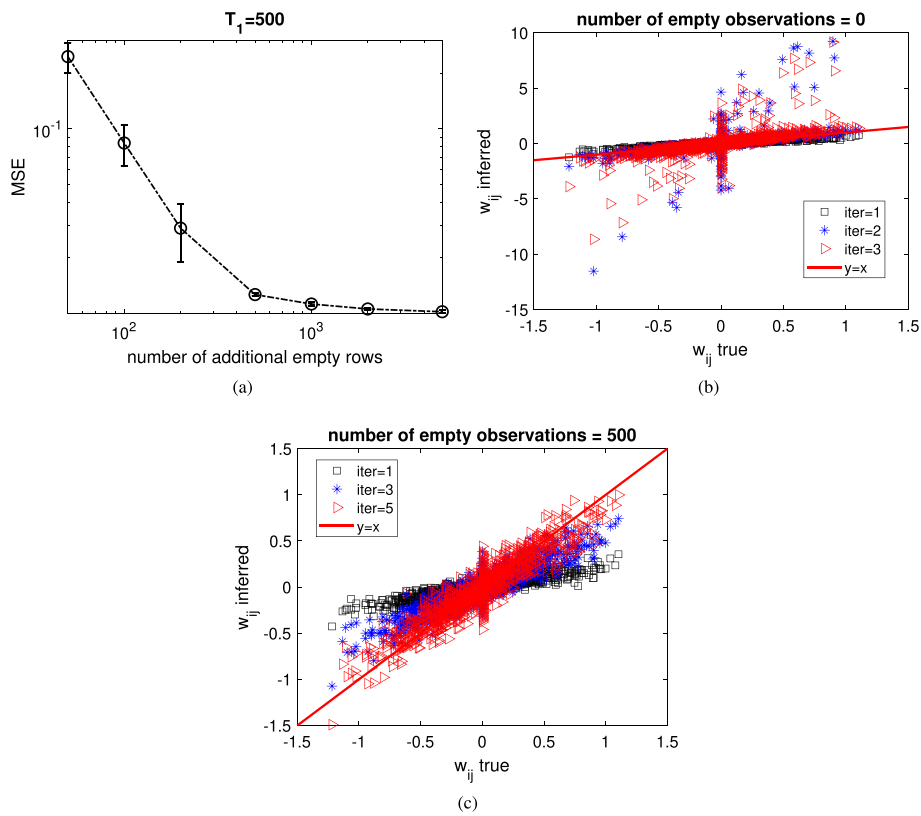
Fig. 2 **a** Dependence of MSE on numbers of additionally empty rows padded at the end of the original dataset. More specifically, the inference is carried out on a dataset containing two parts: a 90% visible data part with length $T_1 = 500$, and an empty dataset with lengths shown in the x-axis. The unobserved data points in T_1 are chosen randomly. **b** Comparison between the true and inferred w_{ij} at different iterations: 1 (black circles), 2 (blue stars), and 3 (red triangles). The inference is conducted in a dataset with length $T = 500$, and 10% data points are randomly missed. After the first iteration, w_{ij} is overestimated. **c** Comparison between the true and inferred w_{ij} at different iterations: 1 (black circles), 3 (blue stars), and 5 (red triangles). The inference is conducted in a dataset with two parts: a 90% visible data part with length $T_1 = 500$, and an empty dataset part with length $T_2 = 500$. At early iterations, w_{ij} is underestimated. Here, we use a setting of BA networks with network size $N = 100$, average degree $\langle k \rangle = 10$, and connectivity strength coefficient $g = 1$

less suited for our purpose. A detailed comparison between l_1 and l_2 regularization can be found in Appendix C.

We first present a sweep across different values of the regularization parameter λ and examine the resulting inference errors. As shown in Fig. 3a, we note that the MSE initially decreases as λ increases, reaching a minimum and then starts increasing again. This aligns with expectations – a small amount of regularization typically constrains over-fitting, but too much regularization might degrade performance. Moreover, we also see that the optimal regularization parameter and the benefits of regularization depend on the lengths of the time series, with the largest benefits observed for $T = 500$.

Additionally, in Fig. 3b, we show the corresponding dependence of the number of iterations until convergence is reached on λ, again for varying observation lengths T. Here, we set a maximum of 100 iterations, $Q_{max} = 100$. Iterations exceeding $Q_{max} = 100$ are considered non-convergent. As seen in Fig. 3b, the λ value corresponding to the minimum MSE aligns closely with the transition point between convergence and
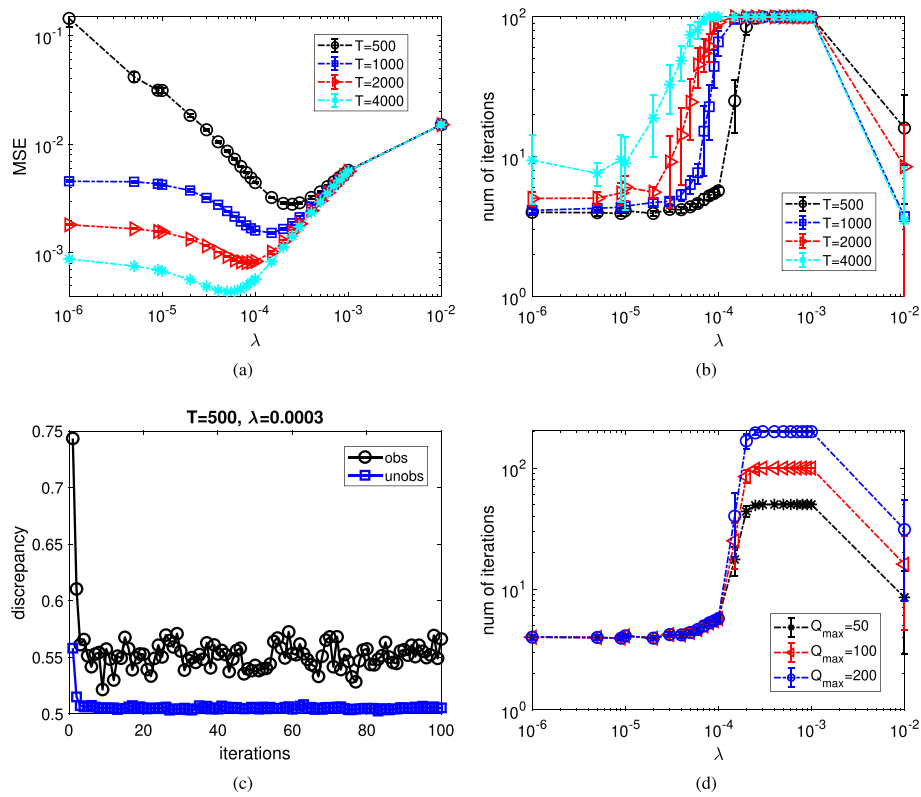
Fig. 3 **a** Dependence of MSE on the l_1 regularization parameter λ for varying observation lengths T. **b** The number of iterations until convergence or reaching maximum iterations $Q_{max} = 100$ as a function of λ. **c** Example evolution of the discrepancies D_{obs} and D_{un} for non-convergent iterations at $T = 500$ and $\lambda = 0.0003$. **d** Dependence of convergence iterations on λ for varying values of maximum iterations Q_{max}. Here, we use a setting of BA networks with network size $N = 100$, average degree $\langle k \rangle = 10$, connectivity strength coefficient $g = 1$, and with 50% missing data points. The results are based on 10 repetitions of the experiments, and error bars indicate 95% confidence intervals

non-convergence, validating the proposed heuristic. Initially, increasing λ causes more iterations until eventual non-convergence is reached at $Q_{max} = 100$. With further λ increases, the strong regularization pushes coupling strengths near zero, abruptly re-establishing convergence (note the last data points in panels (b) and (d)).

To illustrate the non-convergent behavior, Fig. 3c shows the evolution of discrepancies D_{obs} and D_{un} at $T = 500$, and relatively large $\lambda = 0.0003$, with 50% missing data. We see that in spite of fluctuations D_{obs} remains above D_{un}, so the iterations continue beyond $Q_{max} = 100$, as the convergence criterion cannot be met. Finally, Fig. 3d verifies that the overall shape and turning points are robust to the choice of Q_{max}. For large enough Q_{max}, the curves exhibit two distinct behaviors: quick convergence or sustained non-convergence, and results for the first point of non-convergence become independent of the exact choice of Q_{max}. A more systematic exploration for the non-convergent behavior under varying numbers of missing data can be found in Appendix .

Dependence of performance on data availability

Next, we carry out a comprehensive set of experiments on our proposed algorithm from the perspective of missing data selection, the structure of the weights matrix, and network

types to verify the effectiveness of the automatically determined l_1 regularization. To be specific, we test the performance of our algorithm depending on the following model specifications: (i) Percentages of unobserved data p, (ii) Average degrees $\langle k \rangle$ for reflecting levels of network sparsity, (iii) connectivity strengths g, (iv) Network models, such as the BA model, the ER model, and the SW model. To maintain conciseness in the main discussion, we have placed the detailed examination of network size (N) in Appendix B.

To evaluate the robustness of our approach to missing observations, we test the inference performance under varying percentages of unobserved data p in the range from 10% to 70%. This range was selected to encompass a realistic spectrum of missing data scenarios encountered in practical applications, starting from minimal data loss (10%) to a substantial but manageable level (70%). At the lower end of the range, a 10% level of missing data represents scenarios with minimal data loss, which can occur in well-connected online social networks or transportation networks (Li et al. 2014; Kossinets 2006) during periods of normal operation. On the other hand, a 70% level of missing data represents a substantial but manageable level of unobserved nodes, as often seen in biological networks where experimental techniques may not capture the complete set of biomolecules (Liao et al. 2018). The decision to cap the range at 70% was informed by the consideration that beyond this threshold, the volume of missing data likely becomes too extensive for reliable inference. Additionally, experiments are conducted for several observation lengths $T = 500, 1000, 2000$, and $T = 4000$ time steps with a fixed connectivity strength $g = 4$.

In our analysis, we determine the quality of the network reconstructions by computing the MSE between the true and inferred networks, using three distinct approaches: (i) Optimal Inference MSE (min MSE$_{\text{true}}$): This represents the MSE obtained at the λ value that yields the minimum MSE, reflecting the best possible inference outcome when the ground truth is known. This optimal MSE serves as a benchmark for the highest achievable performance in network inference. (ii) Algorithmic Inference (min MSE$_{\text{approx}}$): This MSE is calculated at the λ value determined by our algorithm, providing a measure of how well our proposed method performs in approximating the true network. (iii) Baseline MSE (MSE$_{\lambda=0}$): This metric calculates the MSE without applying l_1 regularization, which corresponds to the method described by Lee et al. (2021). To be specific, MSE$_{\lambda=0}$ acts as a reference for the state-of-the-art performance in network inference with missing data.

To summarize, here, we introduce two reference classes including an ideal benchmark (i.e., min MSE$_{\text{true}}$) and the state-of-the-art (i.e., MSE$_{\lambda=0}$) for evaluating MSE values. The above enables us to more meaningful comparisons across different scales and parameter settings, facilitating a better understanding of the effectiveness of our approach in achieving accurate network reconstruction.

The results of these experiments are plotted in the panels of Fig. 4. We make the following observations. First, inspecting panels (a)-(d), we see that results for the truly best regularization (min MSE$_{\text{true}}$) and approximated best regularization (min MSE$_{\text{approx}}$) are very close across the range of missing data p. Second, performances without regularization (MSE$_{\lambda=0}$) tend to be significantly worse, with decreasing differences as the lengths of the time series increase. The closeness of the optimally and heuristically regularized results demonstrates our algorithm's ability to automatically find a suitable regularization

Cai *et al. Applied Network Science* (2024) 9:13

Page 14 of 22

strength approaching the optimal MSE. For larger p, min $\mathrm{MSE_{true}}$ and min $\mathrm{MSE_{approx}}$ begin to diverge, indicating greater difficulty in identifying the optimal λ as the amount of missing data increases. However, even for 70% missing data, our approach maintains strong gains over the no regularization scenario. Correspondingly, in Fig. 4e, we show the difference between the optimal regularization strength, labelled as λ_{true} that minimizes MSE, versus the automatically determined $\lambda_{\mathrm{approx}}$ based on the convergence transition point. We find that, our algorithm manages to find a very close approximation for the optimal regularization strength, especially when not many data points are missing.
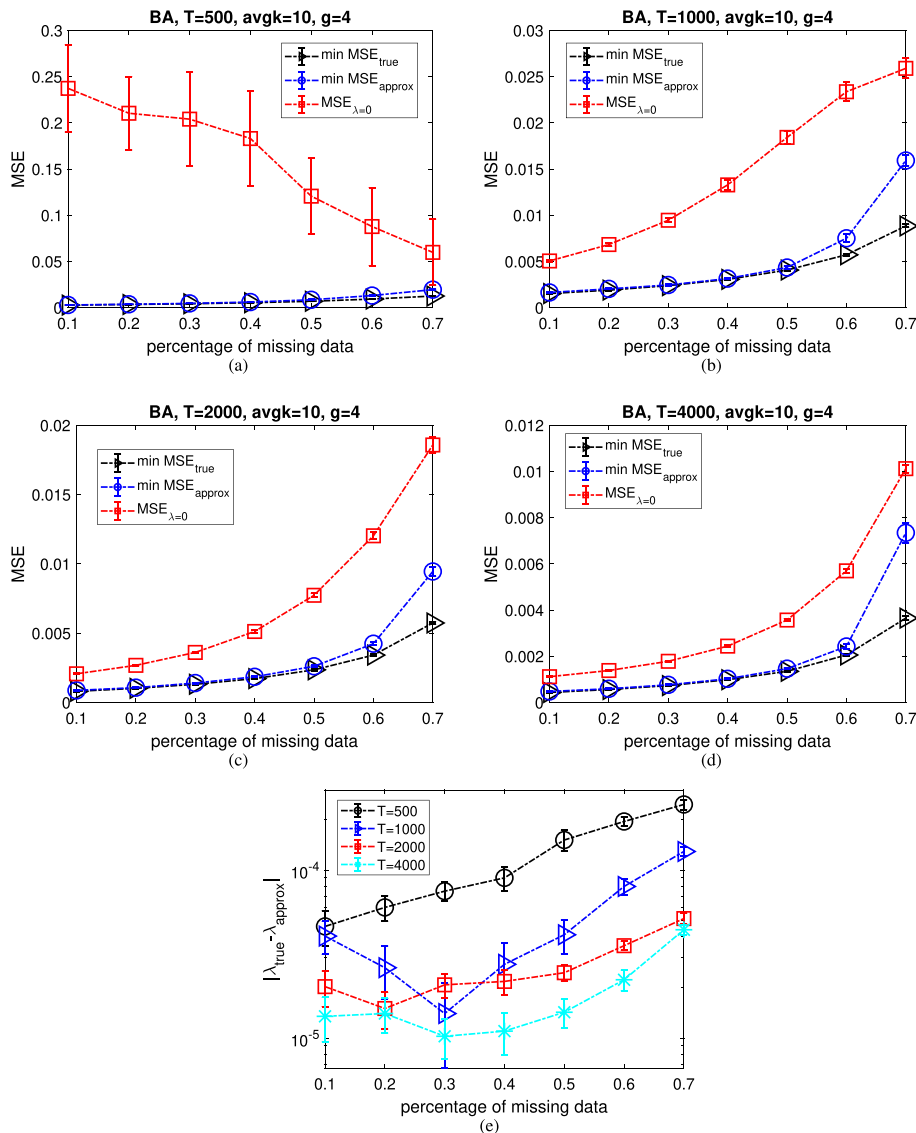
Fig. 4 MSE between inferred and true networks versus percentage of missing data p for observation lengths T of **a** 500, **b** 1000, **c** 2000, and **d** 4000 time steps. Three MSEs are compared: min $\mathrm{MSE_{true}}$ at optimal λ, min $\mathrm{MSE_{approx}}$ at automatically determined λ by our algorithm, and $\mathrm{MSE_{\lambda=0}}$ without regularization. **e** Comparison of the optimal regularization strength λ_{true} that minimizes MSE versus the automatically determined λ_{approx} based on the convergence transition point. Here, we use a setting of BA networks with network size $N = 100$, average degree $\langle k \rangle = 10$, and connectivity strength coefficient $g = 4$. The results are based on 10 repetitions of the experiments, and error bars indicate 95% confidence intervals

Next, we continue with an exploration of the effects of varying network connectivity strengths g on the accuracy of network inference. Specifically, we examine four cases: $g = 1$ in Fig. 5a, $g = 2$ in Fig. 5b, $g = 4$ in Fig. 4a, and $g = 10$ in Fig. 5c. Throughout these evaluations, we observe a consistent and noteworthy alignment between the minimum mean square error of the true model min MSE_{true} and its approximation min MSE_{approx}. This finding reinforces the robustness of our proposed algorithm, particularly in its ability to significantly surpass the performance metrics of the baseline model $MSE_{\lambda=0}$, across varied network connectivity paradigms.

Following the evaluation of connectivity strengths, we examine the role of network sparsity by considering different average degrees $\langle k \rangle$ of the networks. A higher average degree corresponds to a denser network with more connections, while a lower average degree represents a sparser network with fewer connections. Specifically, we examine scenarios with $\langle k \rangle = 5$ in Fig. 6a, $\langle k \rangle = 10$ in Fig. 4a, $\langle k \rangle = 20$ in Fig. 6b, $\langle k \rangle = 30$ in Fig. 6c, and $\langle k \rangle = 40$ in Fig. 6d. These average degrees cover a range of network densities, from highly sparse networks with $\langle k \rangle = 5$ to denser networks with $\langle k \rangle = 40$. Across all the examined average degree scenarios, we observe that our proposed algorithm maintains a high level of accuracy and significantly outperforms the baseline model $MSE_{\lambda=0}$ without regularization. This is evident from the close correspondence
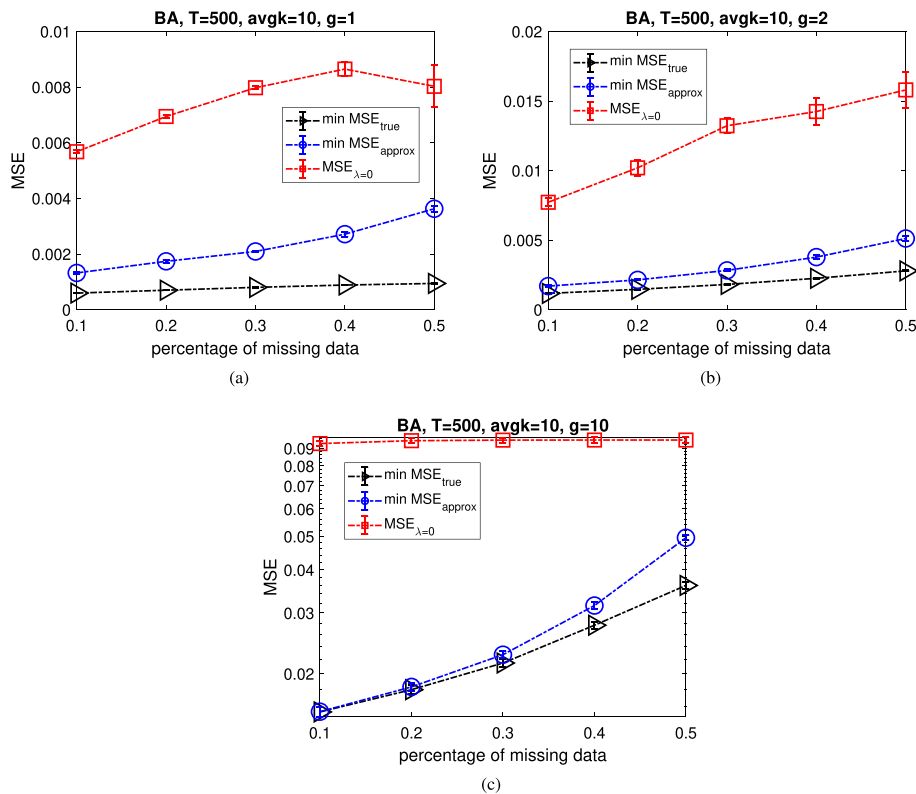
Fig. 5 MSE between inferred and true networks on varying connectivity strengths $g = 1$ (**a**), $g = 2$ (**b**), and $g = 10$ (**c**) for observation length $T = 500$. Three MSEs are compared: min MSE_{true} at optimal λ, min MSE_{approx} at automatically determined λ, and $MSE_{\lambda=0}$ without regularization. Here, we use a setting of BA networks with network size $N = 100$, average degree $\langle k \rangle = 10$. The results are based on 10 repetitions of the experiments, and error bars indicate 95% confidence intervals
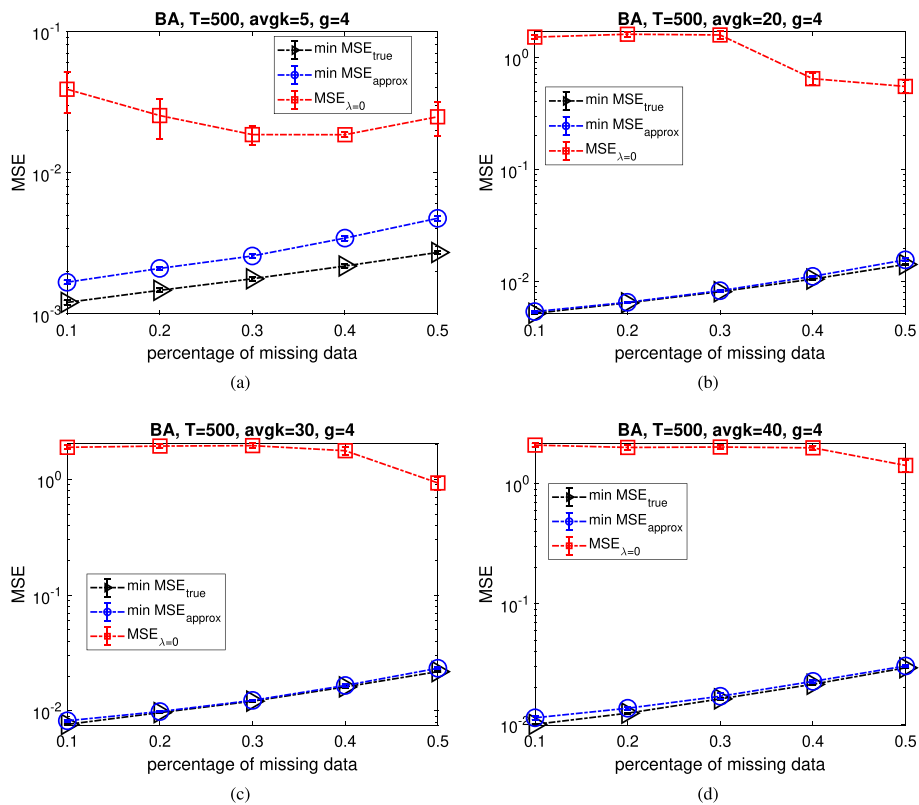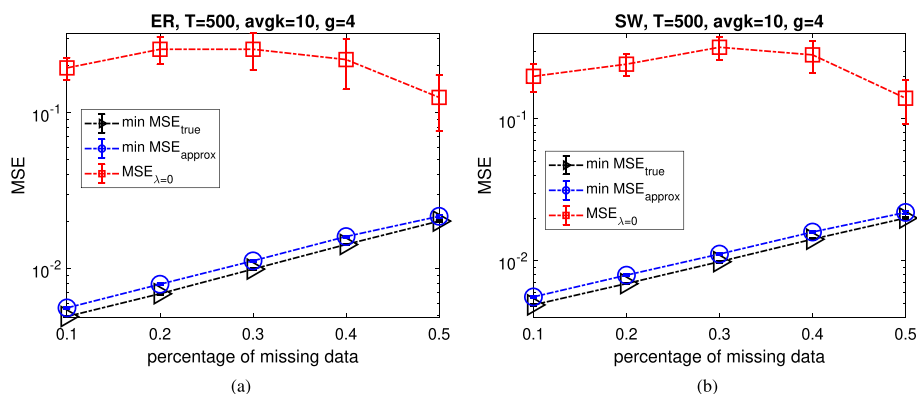
Fig. 6 MSE between inferred and true networks on varying average degrees $\langle k \rangle = 5$ (**a**), $\langle k \rangle = 20$ (**b**), $\langle k \rangle = 30$ (**c**), and $\langle k \rangle = 40$ (**d**) for observation length $T = 500$. Three MSE are compared: min MSE_{true} at optimal λ, min $\text{MSE}_{\text{approx}}$ at automatically determined λ, and $\text{MSE}_{\lambda=0}$ without regularization. Here, we use a setting of BA networks with network size $N = 100$, and connectivity strength coefficient $g = 4$. The results are based on 10 repetitions of the experiments, and error bars indicate 95% confidence intervals

between the minimum mean square error of the true model min MSE_{true} and the approximate model min $\text{MSE}_{\text{approx}}$ obtained with our method. Moreover, we find the relative improvement compared to the unregularized model generally becomes more substantial for denser graphs with higher $\langle k \rangle$. For instance, at $\langle k \rangle = 5$ in Fig. 6a, regularized MSE (min $\text{MSE}_{\text{approx}}$) is 2–25% of baseline $\text{MSE}_{\lambda=0}$. However, at $\langle k \rangle = 40$ in Fig. 6d, regularized MSE decreases to only $0.3 - 2\%$ of the baseline, representing a larger accuracy gain. This reveals greater over-fitting risks in dense networks, further motivating the regularization to constrain complexity. This observation suggests that our method becomes more effective and beneficial as the network connectivity increases and the networks become denser. The greater improvement in accuracy for dense networks can be attributed to the higher risk of overfitting in such scenarios, further motivating the need for regularization to constrain the complexity of the models and prevent overfitting.

So far our experiments have explored the performance of our approach on networks based primarily on the BA model. To demonstrate wider applicability beyond the BA model, we additionally evaluate performance using two other commonly studied network models: SW networks and ER networks. Figure 7 shows the MSE between inferred and true networks on different network types: ER networks (a), and

Fig. 7 MSE between inferred and true networks on different network types: ER networks (**a**), SW networks (**b**) for observation length $T = 500$. Three MSEs are compared: min MSE$_{true}$ at optimal λ, min MSE$_{approx}$ at automatically determined λ, and MSE$_{\lambda=0}$ without regularization. Here, we use a setting of network size $N = 100$, average degree $\langle k \rangle = 10$, and connectivity strength coefficient $g = 4$. The results are based on 10 repetitions of the experiments, and error bars indicate 95% confidence intervals

SW networks (b) for observation length $T = 500$. As seen in Fig. 7a, b, our method continues achieving significant accuracy gains over the baseline across both network models. We observe that on both ER and SW networks, the MSE is improved by over 70% relative to the unregularized inference. Importantly, these generalized gains are achieved without needing to tune the regularization approach to the specific network structure. The consistent benefits verify the broad applicability of our proposed framework beyond reliance on particular topological constraints or connectivity patterns. By automatically adapting the complexity penalty during inference, the benefits persist whether there are highly structured motifs like triangular clustering or entirely random edge formation as in ER graphs. This flexibility highlights the potential to extend the regularization approach to diverse network inference tasks.

Conclusion

In this work, we have proposed and validated an enhanced approach for stochastic network inference from limited time-series observations. Our method introduces automatic l_1 regularization to constrain over-fitting, which is particularly problematic when data are scarce. The algorithm heuristically determines a suitable regularization strength by identifying the transition point between convergence and non-convergence of the heuristic iterations. This provides a data-driven technique for regularization without requiring manual parameter tuning.

Experiments on simulated sparse networks demonstrated significant performance gains over the state-of-the-art method of Lee et al. (2021), especially for small sample sizes. The l_1 regularization is found to result in reduced inference error across varying conditions including observation length, missing data percentages, and network connectivity. Moreover, our approach eliminates the need to manually select a stopping threshold ε, which was previously required to halt iterations before over-fitting. The automatic regularization inherently prevents excess iterations by driving the model toward sparsity. The ability to infer networks from limited, incomplete data with minimal parameter tuning will help enable adoption in real-world settings. Potential applications include

reconstructing gene regulatory networks from scarce biological data and learning neural connectivity from partial observations.

Future work can focus on extending the approach to additional network models beyond the binary Ising spin glass tested here. Testing on real-world inference tasks and integrating side information are also worthwhile directions. Overall, this work takes a significant step toward practical and automated inference of complex networks from scarce, noisy data.

Updating of local fields $H_i(t)$

The local field, denoted as $H_i(t)$, determines the impact of the current state, $\{\sigma_k(t)\}_{k=1}^N$, on the predicted future state, $\sigma_i(t+1)$. Here, we examine a scenario where the local field is defined by $H_i(t) = \sum_j w_{ij}\sigma_j(t)$, focusing on identifying the elements of the weight matrix w_{ij}.

The tendency of the state $\sigma_i(t+1)$ to align with its local field $H_i(t)$ forms the basis of our model. This alignment is quantitatively expressed as

$$E[\sigma_i(t+1) \mid H_i(t)] = \tanh[H_i(t)],$$

where the model expectation is a function of the probabilities of $\sigma_i(t+1)$ assuming values of ± 1 given the current state $\{\sigma_k(t)\}_{k=1}^N$. We observe that the magnitude of the model expectation relative to the actual state, $\sigma_i(t+1)$, is constrained by

$$\left| \frac{E[\sigma_i(t+1) \mid H_i(t)]}{\sigma_i(t+1)} \right| \leqslant 1.$$

To refine the predictive accuracy, we carry out an update to the local field $H_i(t)$ that enhances the alignment of the model expectation with the actual state

$$H_i(t) \leftarrow \frac{\sigma_i(t+1)}{E[\sigma_i(t+1) \mid H_i(t)]} H_i(\sigma(t)),$$

resulting in an improved predictive capability as evidenced by the increased magnitude of the model expectation. This update, importantly, not only improves the prediction but can also correct the direction (sign) of the local field $H_i(t)$ if initially incorrect.

Dependence of performance on network sizes N

In Fig. 8, we investigate the impact of different network sizes N on the accuracy of our network inference method. This exploration specifically covers two scenarios: a network size of $N = 20$, illustrated in Fig. 8a, and a network size of $N = 50$, shown in Fig. 8b. In these analyses, we consistently find that the MSE of our approximation, denoted as min $\text{MSE}_{\text{approx}}$ outperforms the baseline model's MSE ($\text{MSE}_{\lambda=0}$) across a spectrum of network size settings. This outcome underscores the superior robustness and reliability of our proposed algorithm in various network configurations.
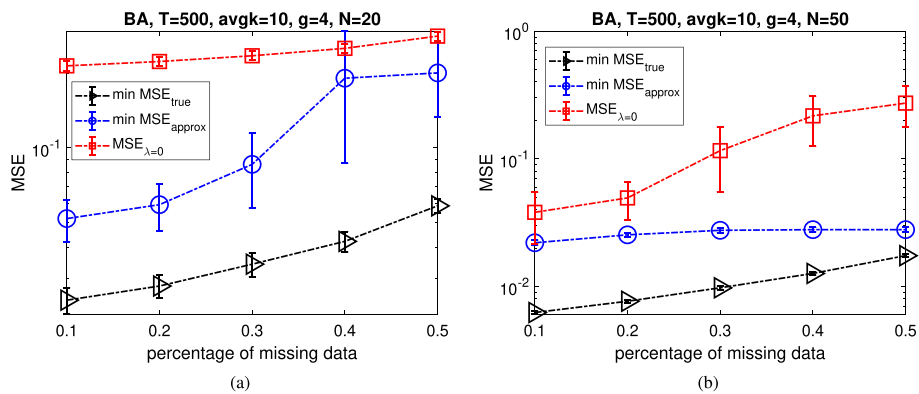
Fig. 8 MSE between inferred and true networks on varying network sizes N, **a** $N = 20$ and **b** $N = 50$ for observation length $T = 500$. Three MSE are compared: min MSE_{true} at optimal λ, min $\text{MSE}_{\text{approx}}$ at automatically determined λ, and $\text{MSE}_{\lambda=0}$ without regularization. Here, we use a setting of BA networks with connectivity strength coefficient $g = 4$. The results are based on 10 repetitions of the experiments, and error bars indicate 95% confidence intervals
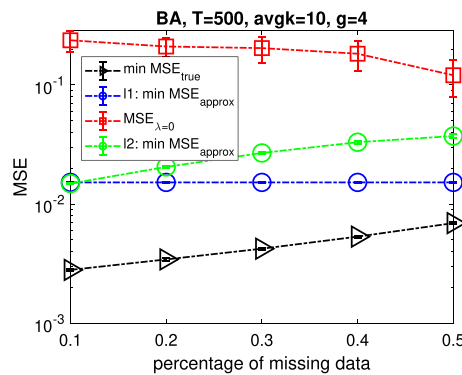
Fig. 9 MSE between inferred and true networks for observation length $T = 500$. Four MSE are compared: min MSE_{true} at optimal λ, l1: min $\text{MSE}_{\text{approx}}$ at automatically determined λ calculated by the l_1 regularization, $\text{MSE}_{\lambda=0}$ without regularization, and l2: min $\text{MSE}_{\text{approx}}$ at automatically determined λ calculated by the l_2 regularization. Here, we use a setting of BA networks with network size $N = 100$, and connectivity strength coefficient $g = 4$. The results are based on 10 repetitions of the experiments, and error bars indicate 95% confidence intervals

Comparison between l_1 and l_2 regularization

In Fig. 9, we evaluate the performance of l_1 and l_2 regularization techniques in inferring network structures, particularly within BA networks characterized by a size of $N = 100$ and a connectivity strength coefficient $g = 4$. This evaluation is conducted over an observation length of $T = 500$, with the goal of comparing four distinct MSE metrics: "min MSE_{true}" at optimal λ, "l1: min $\text{MSE}_{\text{approx}}$" at automatically determined λ calculated by the l_1 regularization, "$\text{MSE}_{\lambda=0}$" without regularization, and "l2: min $\text{MSE}_{\text{approx}}$" at automatically determined λ calculated by the l_2 regularization. The experiments reveal that l_1 regularization can improve the inference accuracy of network structures in most cases compared to the l_2 regularization, especially in the scenarios when a large amount of data is missing. The superiority of l_1 regularization in inferring sparse networks can be attributed to its ability to encourage many coefficients to shrink to zero.
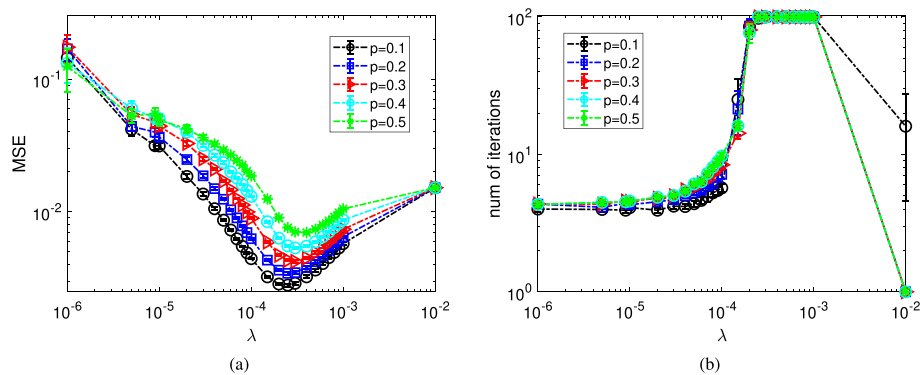
Fig. 10 a Dependence of MSE on the l_1 regularization parameter λ for varying percentages of unobserved data p for observation length $T = 500$. **b** The number of iterations until convergence or reaching maximum iterations $Q_{max} = 100$ as a function of λ for observation length $T = 500$. Here, we use a setting of BA networks with network size $N = 100$, average degree $\langle k \rangle = 10$, and connectivity strength coefficient $g = 4$. The results are based on 10 repetitions of the experiments, and error bars indicate 95% confidence intervals

On the other hand, l_2 regularization, which penalizes the square of the coefficients, does not inherently promote sparsity in the same way l_1 does. While l_2 regularization can help in reducing overfitting by discouraging large coefficients and thus limiting the model's complexity, it typically results in models where most coefficients are non-zero but small. This attribute makes l_2 regularization less suited for our purpose.

Non-convergent behavior under different percentage of missing data

Consistent with Fig. 3a, which showed the impact of varying temperature T, in Fig. 10a, we present a sweep across different values of the regularization parameter λ and examine the resulting inference errors for a fixed observation length T = 500 and varying percentages of missing data $p = 0.1, 0.2, 0.3, 0.4, 0.5$. As shown in Fig. 10a, we note that the MSE initially decreases as λ increases, reaching a minimum, and then starts increasing again. This behavior is observed for all percentages of missing data. Additionally, in Fig. 10b, we show the corresponding dependence of the number of iterations required for convergence on λ, again for varying percentages of missing data p. Here, we set a maximum of 100 iterations, $Q_{max} = 100$, and iterations exceeding $Q_{max} = 100$ are considered non-convergent. As seen in Fig. 10b, the λ value corresponding to the minimum MSE in Fig. 10a aligns closely with the transition point between convergence and non-convergence for each percentage of missing data. This observation further validates our proposed heuristic.

Author Contributions
Conceptualization, ZC, MB and EG; Methodology, ZC, MB and EG; Software, ZC; Validation, MB and EG; Formal analysis, ZC; Investigation, ZC, MB and EG; Resources, ZC, MB and EG; Data curation, ZC; Writing-original draft preparation, ZC; Writing-review and editing, ZC, MB and EG; Visualization, ZC; Supervision, MB and EG; Project administration, MB and EG.

References

Bachschmid-Romano L, Opper M (2014) Inferring hidden states in a random kinetic ising model: replica analysis. J Stat Mech: Theory Exp 2014:P06013

Barabási AL, Albert R (1999) Emergence of scaling in random networks. Science 286:509–512

Battistin C, Hertz J, Tyrcha J, Roudi Y (2015) Belief propagation and replicas for inference and learning in a kinetic ising model with hidden spins. J Stat Mech: Theory Exp 2015:P05021

Bisconti C, Corallo A, Fortunato L, Gentile AA, Massafra A, Pellè P (2015) Reconstruction of a real world social network using the potts model and loopy belief propagation. Front Psychol 6:1698

Campajola C, Lillo F, Tantari D (2019) Inference of the kinetic ising model with heterogeneous missing data. Phys Rev E 99:062138

Cresswell-Clay E, Periwal V (2021) Genome-wide covariation in sars-cov-2. Math Biosci 341:108678. https://doi.org/10.1016/j.mbs.2021.108678

Donner C, Opper M (2017) Inverse ising problem in continuous time: a latent variable approach. Phys Rev E 96:062104. https://doi.org/10.1103/PhysRevE.96.062104

Dunn B, Battistin C (2017) The appropriateness of ignorance in the inverse kinetic ising model. J Phys A: Math Theor 50:124002

Dunn B, Roudi Y (2013) Learning and inference in a nonequilibrium ising model with hidden nodes. Phys Rev E 87:022127

Gemao B, Lai PY et al (2021) Effects of hidden nodes on noisy network dynamics. Phys Rev E 103:062302

Glauber RJ (1963) Time-dependent statistics of the ising model. J Math Phys 4:294–307

Han X, Shen Z, Wang WX, Di Z (2015) Robust reconstruction of complex networks from sparse data. Phys Rev Lett 114:028701

Hastie T, Tibshirani R, Friedman JH, Friedman JH (2009) The elements of statistical learning: data mining, inference, and prediction, vol 2. Springer

Hoang DT, Jo J, Periwal V (2019a) Data-driven inference of hidden nodes in networks. Phys Rev E 99:042114. https://doi.org/10.1103/PhysRevE.99.042114

Hoang DT, Song J, Periwal V, Jo J (2019b) Network inference in stochastic systems from neurons to currencies: improved performance at small sample size. Phys Rev E 99:023311. https://doi.org/10.1103/PhysRevE.99.023311

Kossinets G (2006) Effects of missing data in social networks. Soc Netw 28:247–268

Lee S, Periwal V, Jo J (2021) Inference of stochastic time series with missing data. Phys Rev E 104:024119. https://doi.org/10.1103/PhysRevE.104.024119

Li Y, Li Z, Li L (2014) Missing traffic data: comparison of imputation methods. IET Intel Transport Syst 8:51–57

Liao L, Li K, Li K, Yang C, Tian Q (2018) A multiple kernel density clustering algorithm for incomplete datasets in bioinformatics. BMC Syst Biol 12:99–116

Liu Q, Ihler A (2011) Learning scale free networks by reweighted l_1 regularization. In: Proceedings of the fourteenth international conference on artificial intelligence and statistics, JMLR Workshop and Conference Proceedings, pp 40–48

Mazzarisi P, Zaoli S, Campajola C, Lillo F (2020) Tail granger causalities and where to find them: extreme risk spillovers vs spurious linkages. J Econ Dyn Control 121:104022

Mézard M, Sakellariou J (2011) Exact mean-field inference in asymmetric kinetic ising systems. J Stat Mech: Theory Exp 2011:L07001

Mirshahvalad A, Lindholm J, Derlén M, Rosvall M (2012) Significant communities in large sparse networks. PLoS ONE 7:e33721

Newman ME, Strogatz SH, Watts DJ (2001) Random graphs with arbitrary degree distributions and their applications. Phys Rev E 64:026118

Nguyen HC, Zecchina R, Berg J (2017) Inverse statistical problems: from the inverse ising problem to data science. Adv Phys 66:197–261. https://doi.org/10.1080/00018732.2017.1341604

Pearl J (2000) Causal inference without counterfactuals: comment. J Am Stat Assoc 95:428–431

Roudi Y, Dunn B, Hertz J (2015) Multi-neuronal activity and functional connectivity in cell assemblies. Curr Opin Neurobiol 32:38–44. https://doi.org/10.1016/j.conb.2014.10.011

Roudi Y, Hertz J (2011a) Dynamical tap equations for non-equilibrium ising spin glasses. J Stat Mech: Theory Exp 2011:P03031

Roudi Y, Hertz J (2011b) Mean field theory for nonequilibrium network reconstruction. Phys Rev Lett 106:048702. https://doi.org/10.1103/PhysRevLett.106.048702

Singh A, Humphries MD (2015) Finding communities in sparse networks. Sci Rep 5:8828

Soudry D, Keshri S, Stinson P, Oh MH, Iyengar G, Paninski L (2015) Efficient "shotgun" inference of neural connectivity from highly sub-sampled activity data. PLoS Comput Biol 11:e1004464

Tyrcha J, Hertz J (2014) Network inference with hidden units. Math Biosci Eng 11:149–165

Tyrcha J, Roudi Y, Marsili M, Hertz J (2013) The effect of nonstationarity on models inferred from neural data. J Stat Mech: Theory Exp 2013:P03005. https://doi.org/10.1088/1742-5468/2013/03/P03005

Watts DJ, Strogatz SH (1998) Collective dynamics of "small-world" networks. Nature 393:440–442

Wu TT, Lange K (2008) Coordinate descent algorithms for lasso penalized regression. Ann Appl Stat

Zdeborová L, Krzakala F (2016) Statistical physics of inference: thresholds and algorithms. Adv Phys 65:453–552. https://doi.org/10.1080/00018732.2016.1211393

Zeng HL, Aurell E, Alava M, Mahmoudi H (2011) Network inference using asynchronously updated kinetic ising model. Phys Rev E 83:041135. https://doi.org/10.1103/PhysRevE.83.041135

Zhang P (2012) Inference of kinetic ising model on sparse graphs. J Stat Phys 148:502–512

Zipkin JR, Schoenberg FP, Coronges K, Bertozzi AL (2016) Point-process models of social network interactions: parameter estimation and missing data recovery. Eur J Appl Math 27:502–529

Publisher's Note