

Sparse Bayesian Learning for Sequential Inference of Network Connectivity from Small Data

Jinming Wan, Jun Kataoka, Jayanth Sivakumar, Eric Peña, Yiming Che, Hiroki Sayama, and Changqing Cheng

Abstract—While significant efforts have been attempted in the design, control, and optimization of complex networks, most existing works assume the network structure is known or readily available. However, the network topology can be radically recast after an adversarial attack and may remain unknown for subsequent analysis. In this work, we propose a novel Bayesian sequential learning approach to reconstruct network connectivity adaptively: A sparse Spike and Slab prior is placed on connectivity for all edges, and the connectivity learned from reconstructed nodes will be used to select the next node and update the prior knowledge. Central to our approach is that most realistic networks are sparse, in that the connectivity degree of each node is much smaller compared to the number of nodes in the network. Sequential selection of the most informative nodes is realized via the between-node expected improvement. We corroborate this sequential Bayesian approach in connectivity recovery for a synthetic ultimatum game network and the IEEE-118 power grid system. Results indicate that only a fraction (~50%) of the nodes need to be interrogated to reveal the network topology.

Index Terms— Network Reconstruction; Inverse Problem; Network Inference; Spike and Slab; Sequential Node Selection

I. INTRODUCTION

THE past decades have witnessed the expanding complexity of interconnected engineering systems to accomplish sophisticated design functions. For instance, manufacturing systems are becoming more complicated in the context of globalization and the infiltration of renewable energy sources has compounded the control of power grid systems. As those systems are increasingly interconnected, the interdependency in conjunction with the bewildering complexity has brought network science into the spotlight. Yet, the vast majority of network research is focused on the forward problem: given the network topology and interaction between constituent components, what emergent behaviors will the system exhibit [1], and what is the shortest path to traverse the network? While significant effort has been attempted to date on optimization and control (e.g., resilience design [2] and optimal control [3]) of networked systems for desired dynamics or functionalities, most works hinge on the assumption that the network topology is readily available or can be accurately

estimated efficiently [4], [5]. However, direct access to network structures remains elusive, leaving us with only a restricted set of observable data [6]. For example, in contingencies of malicious attack, natural disaster, or human misoperation, connectivity of a significant portion of the power grid or communication networks remains unknown at the onset of such unexpected events, which could substantially crimp rescue efforts. Moreover, the next-generation power grid system is poised to become more sophisticated, such that the effective transmission-line parameters can be actively controlled with alternating-current transmission devices [7], which has essentially made the whole system susceptible to misoperation. Thus, network connectivity could be seriously altered in such extreme events. Such circumstances underscore the imperative for reconstruction approaches to unveiling the intricate network structures with the wide availability of big data. The task of network reconstruction is inherently formidable. This complexity arises from the fact that structural information is obscured within the measurable data in an enigmatic fashion. Moreover, the solution space encompassing all conceivable structural configurations is characterized by an exceedingly high dimensionality, as in most inverse problem settings.

We seek to infer network topology and connectivity from a paucity of sensing data in a timely and efficient fashion. This inverse problem plays a quintessential role in anomaly detection, root cause diagnosis, and timely deployment of corrective actions. It is noted that most realistic networks possess sparse connectivity, in that each node is only connected to a small subset of the nodes. This sparsity in the connectivity reconstruction translates into a sparse representation problem. Accordingly, sparse learning or compressive sensing techniques developed in the machine learning community have been attempted in network reconstruction using small samples of measurement data [7], [8], [9]. In those works, nodal connectivity is reconstructed iteratively via sparse learning until all nodes are looped through. However, the exhaustive reconstruction strategy does not consider the redundant and potentially conflicting information in the iterative process. Specifically, the connectivity between nodes i and j is unveiled from both nodes i and j separately and independently. This inefficiency has stymied the widespread application to large-

Jinming Wan, Jun Kataoka, Eric Peña, Hiroki Sayama, and Changqing Cheng are with the Systems Science and Industrial Engineering Department, Binghamton University, 4400 Vestal Parkway East, Binghamton, NY 13902, USA (email: jwan8@binghamton.edu; jkataok1@binghamton.edu; epena4@binghamton.edu; sayama@binghamton.edu; ccheng@binghamton.edu). (Corresponding author: Changqing Cheng).

Jayanth Sivakumar is with Citibank, N.A. Company, 6460 Las Colinas Blvd, Irving, TX 75039, USA (email: jsivakul@binghamton.edu).

Yiming Che is with School of Computing and Augmented Intelligence, Arizona State University, Tempe, AZ 85281, USA (email: yche14@ase.asu.edu).

scale networks.

We developed a novel sequential Bayesian analysis framework to reconstruct network connectivity from short time series data recorded at the node level. The connectivity recovered from one node is treated as the prior information of the nodes that are not yet investigated, which will be updated via a Bayesian framework. This resolves the issue of conflicting information, particularly for 0-1 connectivity in unweighted networks [8]. Further, leveraging the redundant information, a sequential sampling scheme based on expected improvement (EI) is adopted to adaptively select the most informative node for recovery and maximally reduce the uncertainty in network inference. This allows efficient reconstruction with only a subset of pivotal nodes and has the potential to scale to massive networks. The proposed methodology is corroborated in the unweighted ultimatum game (UG) network and the weighted IEEE-118 power grid network in this study. This approach could potentially advance research in connectivity inference for complex networks with small sets of sparse data and pave the way for effective network design, optimization, and control.

II. RELATED WORK

Mathematically, network topology inference is a high-dimensional inverse problem. Most attempts assume prior knowledge of network structure and emphasize concerns regarding directed or undirected interactions, the existence of a link, and the strength of interactions along with their temporal and spatial scales. For instance, Timme and Casadiego [9] revealed the interaction topology of a network from the collective dynamics of its constituent nodes driven by a given external force. Yu et al. [10] developed a network copy whose topology can be continuously updated to mimic the dynamical behaviors of the target network. With a priori functional form to describe the dynamics of spatially extended networks, topology estimation is achieved via a control signal for steady-state stabilization [11]. Similarly, Ren et al. [12] found that noise in dynamical correlation between nodal dynamics leads to more accurate topology estimation, as the interaction between nodes vanishes during synchronization, and the noise brings an additional dimension to study the network topology. Alternatively, the leap forward in smart sensing and Internet-of-Things technologies has brought the proliferation of “big data” to retrieve network topology. Donges et al. [13] combined mutual information of nonlinear time series data and the betweenness of the network systems to uncover global energy and information flow in the climate network.

To handle the nonstationary time series data, transient Granger causal interaction was explored to reveal the time-varying connectivity of the stimulus-activated neural sources [14]. Regression-based methods for connectivity identification were reported in [15]. However, this requires a large amount of data in the least-squares fitting. Variations of Granger causality have also been developed to recover network connectivity from time series measurements for each node, subject to stochastic perturbation [16], [17].

A thorough review of data-driven reconstruction of complex

networks and dynamical processes is provided in [18]. Nevertheless, a “small data” issue is also omnipresent in the scenario of rare or extreme events [19], [20]. For instance, the power grid network connectivity is vulnerable to perturbations owing to cascading failures. To rapidly roll out the rescue measures and dispatch frontline staff, only a short duration of phasor data can be collected [21]. This has posed a tremendous challenge to the accuracy of network topology inference, particularly considering the uncertainty and ambient noise associated with data recording. Indeed, standard measures (e.g., information theory and correlation) may not decipher the network interdependency with extremely short time series.

A recent line of research exploits compressive sensing on the premise that realistic complex networks are sparse: the connectivity degree of each node is considerably low compared to the total number of nodes in the network. Thus, the connectivity vector to be reconstructed is sparse with only a few non-zero entities. Correspondingly, the observational or data collection requirements can be relaxed, and compressive sensing is powerful to reconstruct a sparse signal from small data. For instance, capitalizing on compressive sensing theory, Shen et al. [22] reconstructed the epidemic spreading networks with highly stochastic dynamics from binary time series (e.g., infected or not).

A sparse learning framework for inference of network connectivity was introduced in [8]. The construction process entails the iterative estimation of the sparse adjacency matrix $\mathbf{X}$, and the connectivity vector $\mathbf{X}_i = [x_{i1}, \dots, x_{iN}]^T$ is successively obtained for each node $i = 1, \dots, N$, which is in essence an inverse problem: given nodal observation $\mathbf{Y}$ and a measurement matrix $\boldsymbol{\phi}$, the network connectivity matrix $\mathbf{X}$ is sought after. More specifically, for node i , $\mathbf{Y}_i = \boldsymbol{\phi}_i \mathbf{X}_i$. Here, $\mathbf{Y}_i = [y_i(t_1), \dots, y_i(t_M)]^T$ is the nodal observation for t_M time steps, and the measurement matrix $\boldsymbol{\phi}_i = \begin{bmatrix} \phi_{i1}(t_1) & \cdots & \phi_{iN}(t_1) \\ \vdots & \ddots & \vdots \\ \phi_{i1}(t_M) & \cdots & \phi_{iN}(t_M) \end{bmatrix}$ signifies the dynamic interactions between node i and all N nodes (including self-interaction) in the network for t_M time steps. $\mathbf{X}_i$ is the connectivity vector for node i , which is sparse with only a few nonzero entities. For unweighted networks, $x_{ij} \in \{0,1\}$: $x_{ij} = 0$ implies that node i and node j ($j = 1, \dots, N$) are not connected and $x_{ij} = 1$ denotes the existence of an edge in between. Since the number of nonzero entities contained in the vector $\mathbf{X}_i$ is significantly less than N , the network is sparse. The sparse reconstruction problem is treated as a regularized least-squares problem, leveraging the acquired nodal observation data $\mathbf{Y}_i$ at each node i and a measurement matrix $\boldsymbol{\phi}_i$:

$$\min_{\mathbf{X}_i \in \mathbb{R}^N} \|\mathbf{Y}_i - \boldsymbol{\phi}_i \mathbf{X}_i\|_2^2 + \lambda \|\mathbf{X}_i\|_0 \quad (1)$$

where $\lambda > 0$ is a regularization parameter, and the ℓ_0 -norm $\|\mathbf{X}_i\|_0$ directly controls the sparsity or the number of non-zero elements in $\mathbf{X}_i$. As ℓ_0 -norm optimization problem is NP hard, it is typically relaxed via the ℓ_1 -norm $\|\mathbf{X}_i\|_1$. The solution only provides point estimates for the connectivity of node i to other nodes in the network. This procedure is repeated N times until all

the nodes are investigated.

In the same vein, the sparse learning approach was adopted in the reconstruction of power grid networks with node-level sensing data [7]. The reduction in measurement of nodal variables is achieved through compressed sensing that makes use of structural properties of the grid network. In [23], a rank-based nonlinear interdependence measure was developed to infer the coupling strength and link density of the underlying network, which addressed the problem associated with correlations and mutual information. The measure was applied to a system of coupled Lorenz dynamics as well as to multichannel electroencephalographic recordings from an epilepsy patient.

Improvements are possible when entering the Bayesian domain. In [24], compressed sensing and the Bayesian approach registered robust and accurate results in network reconstruction from potentially incomplete and noisy data. Huang et al. [25] incorporated a hierarchical prior model in Bayesian learning for network reconstruction on evolutionary game data, and the learning parameters were updated iteratively as the reconstruction of nodes progresses. It is noteworthy that the Bayesian approach is appealing in connectivity inference, which offers a distribution as opposed to the point estimate in the Lasso framework. However, accurate and robust inference of large-scale complex networks is still a confounding pursuit, especially considering the short and limited measurements corrupted by the ambient or measurement noise and other artifacts.

III. METHODOLOGY

Our network reconstruction framework consists of two parts: sparse Bayesian learning for recovery of node connectivity and sequential retrieval to select the next most informative node to investigate.

A. Sparse Bayesian Learning for Node Connectivity Recovery

Sparse Bayesian learning has garnered tremendous traction recently to account for the uncertainty associated with sparse solutions [24], [25], [26]. The key is the sparsity-promoting prior formulation, and different prior distributions have been investigated in literature, including normal product [27], Laplace [28], horseshoe [29], scale-mixture ℓ_1 norm [30] and inverse Gamma [31]. It is noted that the ℓ_1 prior imposes penalty on the sum of the magnitudes of estimated weights or connections, as opposed to the constraint on each edge in the spike-and-slab prior. Spike-and-slab and horseshoe priors also offer advantages over Laplace and inverse Gamma priors, particularly in enforcing sparsity. Unlike Laplace and inverse Gamma priors, which either overly shrink all coefficients or leave them largely unaffected due to their single scale, spike-and-slab priors use a mixture of densities with different scales. This mixture allows them to selectively shrink only a subset of coefficients while leaving others largely unchanged. The spike-and-slab priors distinguish between coefficients better modeled by the slab (which remain almost unchanged) and those better

suited to the spike (which are heavily shrunk towards zero). Moreover, as a flexible shrinkage method, Spike and Slab prior allows user-specified sparsity [32]. Both spike-and-slab and horseshoe priors exhibit similar selective shrinkage, but spike-and-slab priors have additional benefits. They allow direct adjustment of sparsity by modifying the spike weight, which controls the proportion of coefficients expected to be zero. Additionally, spike-and-slab priors can be expressed with latent binary variables indicating whether each coefficient is assigned to the spike or slab, facilitating the identification of relevant features. Lastly, spike-and-slab priors offer a closed-form convolution with the Gaussian distribution, an advantage for approximate inference methods using Gaussian approximations, unlike horseshoe priors [33], [34].

Here, the spike refers to the distribution with spike at $\mathbf{X}_i = \mathbf{0}$, and the slab determines the distribution of non-zero entities of $\mathbf{X}_i$. On a side remark, the sparsity induced by the Spike and Slab model has also been widely used in feature selection [32], [33].

In this study, it is adopted to specify the prior distribution of connectivity $\mathbf{X}_i$:

$$\begin{aligned} p(\mathbf{X}_i|\mathbf{z}_i) &= \prod_{j=1}^N p(x_{ij}|z_{ij}) \\ &= \prod_{j=1}^N [(1 - z_{ij})\delta(x_{ij}) \\ &\quad + z_{ij}\mathcal{N}(x_{ij}|m_{ij}, v_{ij})] \end{aligned} \quad (2)$$

where $\delta(\cdot)$ is the point probability mass centered at the spike 0, and m_{ij} and v_{ij} are the mean and variance of the slab distribution (normal here). $\mathbf{z}_i = [z_{i1}, \dots, z_{iN}]$, and $z_{ij} \in \{0,1\}$ is the latent binary variable indicating whether x_{ij} attains the deterministic value 0 ($z_{ij} = 0$) or is drawn from the slab distribution ($z_{ij} = 1$), where $j = 1, \dots, N$. In a hierarchical prior setting, z_{ij} follows a Bernoulli distribution, $z_{ij} \sim \text{Bernoulli}(z_{ij}|\gamma_{ij})$. The parameter $\gamma_{ij} \in [0,1]$ is the prior probability that x_{ij} deviates from zero. Therefrom, the distribution of $\mathbf{z}_i$ is the product of Bernoulli terms given $\boldsymbol{\gamma}_i = [\gamma_{i1}, \dots, \gamma_{iN}]$:

$$p(\mathbf{z}_i) = \prod_{j=1}^N \text{Bernoulli}(z_{ij}|\gamma_{ij}) \quad (3)$$

Given nodal observation $\mathbf{Y}_i$ and measurement matrix $\boldsymbol{\phi}_i$, the likelihood function is

$$p(\mathbf{Y}_i|\mathbf{X}_i, \boldsymbol{\phi}_i) = \prod_{t=t_1}^{t_M} \mathcal{N}(y_i(t)|\boldsymbol{\phi}_i\mathbf{X}_i, \sigma_0^2) \quad (4)$$

where σ_0^2 is the noise variance. The posterior distribution $p(\mathbf{X}_i, \mathbf{z}_i|\mathbf{Y}_i, \boldsymbol{\phi}_i)$ is given as

$$p(\mathbf{X}_i, \mathbf{z}_i|\mathbf{Y}_i, \boldsymbol{\phi}_i) = \frac{p(\mathbf{Y}_i|\mathbf{X}_i, \boldsymbol{\phi}_i)p(\mathbf{X}_i|\mathbf{z}_i)p(\mathbf{z}_i)}{p(\mathbf{Y}_i|\boldsymbol{\phi}_i)} \quad (5)$$

Consequently, the marginal posterior distribution for $\mathbf{X}_i$ can be updated as

$$p(\mathbf{X}_i|\mathbf{z}_i, \mathbf{Y}_i, \boldsymbol{\phi}_i) = \int p(\mathbf{X}_i, \mathbf{z}_i|\mathbf{Y}_i, \boldsymbol{\phi}_i) d\mathbf{z}_i \quad (6)$$

As the integration here is intractable, numerical approximations such as expectation propagation (EP) [32], [33], [34], which is a general approach for approximating the integral of functions factorized into multiple simple factors, can be used to estimate the posterior. We factorize the integral part of (6) into multiple simple factors, before using the EP method to approximate each factor. Since the evidence $p(\mathbf{Y}_i|\boldsymbol{\phi}_i)$ is a normalizing constant, we only need to factorize the numerator part $p(\mathbf{Y}_i|\mathbf{X}_i, \boldsymbol{\phi}_i)p(\mathbf{X}_i|\mathbf{z}_i)p(\mathbf{z}_i)$ of (5). Thus, we obtain the following representation for the joint distribution of $\mathbf{X}_i$, $\mathbf{z}_i$, and $\mathbf{Y}_i$ given $\boldsymbol{\phi}_i$:

$$\begin{aligned} p(\mathbf{X}_i, \mathbf{z}_i, \mathbf{Y}_i|\boldsymbol{\phi}_i) &= p(\mathbf{Y}_i|\mathbf{X}_i, \boldsymbol{\phi}_i)p(\mathbf{X}_i|\mathbf{z}_i)p(\mathbf{z}_i) \\ &= \prod_{t=t_1}^{t_M} p(y_i(t)|\mathbf{X}_i, \boldsymbol{\phi}_i) p(\mathbf{X}_i|\mathbf{z}_i)p(\mathbf{z}_i) \\ &= \prod_{h=1}^3 g_h(\mathbf{X}_i, \mathbf{z}_i) \end{aligned} \quad (7)$$

For ease of representation, we denote $g_1(\mathbf{X}_i, \mathbf{z}_i) = \prod_{t=t_1}^{t_M} p(y_i(t)|\mathbf{X}_i, \boldsymbol{\phi}_i)$, $g_2(\mathbf{X}_i, \mathbf{z}_i) = p(\mathbf{X}_i|\mathbf{z}_i)$, and $g_3(\mathbf{X}_i, \mathbf{z}_i) = p(\mathbf{z}_i)$. $g_h(\mathbf{X}_i, \mathbf{z}_i)$ can be approximated as a simpler form $\hat{g}_h(\mathbf{X}_i, \mathbf{z}_i)$ by EP. Then we have

$$\prod_{h=1}^3 g_h(\mathbf{X}_i, \mathbf{z}_i) \approx \prod_{h=1}^3 \hat{g}_h(\mathbf{X}_i, \mathbf{z}_i) = Q(\mathbf{X}_i, \mathbf{z}_i) \quad (8)$$

Here, $\hat{g}_h(\mathbf{X}_i, \mathbf{z}_i)$ (Gaussian and Bernoulli distributions) adheres to the exponential distribution family [35], ensuring that their product $Q(\mathbf{X}_i, \mathbf{z}_i)$ is also exponential, as dictated by the exponential closure property [34]. Consequently, the posterior distribution $p(\mathbf{X}_i, \mathbf{z}_i|\mathbf{Y}_i, \boldsymbol{\phi}_i)$ is obtained through the normalized $Q(\mathbf{X}_i, \mathbf{z}_i)$ with normalizing constant $p(\mathbf{Y}_i|\boldsymbol{\phi}_i)$.

Next, functions $\hat{g}_1$, $\hat{g}_2$, and $\hat{g}_3$ are iteratively updated to minimize the Kullback-Leibler (KL) divergence $D_{KL}(g_h Q^{\setminus h} || \hat{g}_h Q^{\setminus h})$ between $\hat{g}_h(\mathbf{X}_i, \mathbf{z}_i) Q^{\setminus h}(\mathbf{X}_i, \mathbf{z}_i)$ and $g_h(\mathbf{X}_i, \mathbf{z}_i) Q^{\setminus h}(\mathbf{X}_i, \mathbf{z}_i)$. Here, $Q^{\setminus h}(\mathbf{X}_i, \mathbf{z}_i) = \frac{Q(\mathbf{X}_i, \mathbf{z}_i)}{\hat{g}_h(\mathbf{X}_i, \mathbf{z}_i)}$ and $D_{KL}(g_h Q^{\setminus h} || \hat{g}_h Q^{\setminus h}) = \sum_{\mathbf{z}_i} \int \left[g_h Q^{\setminus h} \log \left(\frac{g_h Q^{\setminus h}}{\hat{g}_h Q^{\setminus h}} \right) + \hat{g}_h Q^{\setminus h} - g_h Q^{\setminus h} \right] d\mathbf{X}_i$.

The approximate form $\hat{g}_h(\mathbf{X}_i, \mathbf{z}_i)$ in (8) can be expressed as

$$\hat{g}_1(\mathbf{X}_i, \mathbf{z}_i) = \tilde{s}_1 \prod_{j=1}^N \exp \left\{ -\frac{(x_{ij} - \tilde{m}_{1j})^2}{2\tilde{v}_{1j}} \right\} \quad (9a)$$

$$\begin{aligned} \hat{g}_2(\mathbf{X}_i, \mathbf{z}_i) &= \tilde{s}_2 \prod_{j=1}^N \exp \left\{ -\frac{(x_{ij} - \tilde{m}_{2j})^2}{2\tilde{v}_{2j}} \right\} \exp \left\{ \log \left(\frac{\tilde{\gamma}_{2j}}{1 - \tilde{\gamma}_{2j}} \right) z_{ij} \right. \\ &\quad \left. + \log(1 - \tilde{\gamma}_{2j}) \right\} \end{aligned} \quad (9b)$$

$$\begin{aligned} \hat{g}_3(\mathbf{X}_i, \mathbf{z}_i) &= \tilde{s}_3 \prod_{j=1}^N \exp \left\{ \log \left(\frac{\tilde{\gamma}_{3j}}{1 - \tilde{\gamma}_{3j}} \right) z_{ij} \right. \\ &\quad \left. + \log(1 - \tilde{\gamma}_{3j}) \right\} \end{aligned} \quad (9c)$$

where $\{\tilde{s}_\alpha\}_{\alpha=1}^3$, $\{\tilde{\mathbf{m}}_b = (\tilde{m}_{b1}, \dots, \tilde{m}_{bN}), \tilde{\mathbf{v}}_j = (\tilde{v}_{b1}, \dots, \tilde{v}_{bN})\}_{b=1}^2$, and $\{\tilde{\gamma}_d = (\tilde{\gamma}_{d1}, \dots, \tilde{\gamma}_{dN})\}_{d=2}^3$ denote free parameters for the EP algorithm. For each iteration, the free parameters are updated to minimize the KL divergence $D_{KL}(g_h Q^{\setminus h} || \hat{g}_h Q^{\setminus h})$. When the change amplitude of all the free parameters is less than a threshold (1×10^{-6} in this work) or the number of iterations reaches a limit (1,000 here), estimate of $Q(\mathbf{X}_i, \mathbf{z}_i)$ in (8) is considered to have converged. Then, $Q(\mathbf{X}_i, \mathbf{z}_i)/p(\mathbf{Y}_i|\boldsymbol{\phi}_i)$ is considered as the approximate posterior distribution $p(\mathbf{X}_i, \mathbf{z}_i|\mathbf{Y}_i, \boldsymbol{\phi}_i)$ in (5), which also has the exponential form:

$$\begin{aligned} p(\mathbf{X}_i, \mathbf{z}_i|\mathbf{Y}_i, \boldsymbol{\phi}_i) &= \prod_{j=1}^N \mathcal{N}(x_{ij}|m_{ij}, v_{ij}) \text{Bernoulli}(z_{ij}|\gamma_{ij}) \end{aligned} \quad (10)$$

Here, $\mathbf{m}_i = [m_{i1}, \dots, m_{iN}]$, $\mathbf{v}_i = [v_{i1}, \dots, v_{iN}]$, and $\boldsymbol{\gamma}_i = [\gamma_{i1}, \dots, \gamma_{iN}]$ are updated based on $\tilde{\mathbf{m}}_1, \tilde{\mathbf{m}}_2, \tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2, \tilde{\gamma}_2$, and $\tilde{\gamma}_3$:

$$v_{ij} = [\tilde{v}_{1j}^{-1} + \tilde{v}_{2j}^{-1}]^{-1} \quad (11a)$$

$$m_{ij} = [\tilde{m}_{1j}\tilde{v}_{1j}^{-1} + \tilde{m}_{2j}\tilde{v}_{2j}^{-1}]^{-1} \quad (11b)$$

$$\gamma_{ij} = [\tilde{\gamma}_{1j} + \tilde{\gamma}_{2j}] \quad (11c)$$

With the estimated posterior distribution obtained from (10) by importing the estimated $\mathbf{m}_i$, $\mathbf{v}_i$, and $\boldsymbol{\gamma}_i$, an edge between node i and j exists if $p(z_{ij}|\mathbf{Y}_i, \boldsymbol{\phi}_i) > \varrho$, where $0 \leq \varrho \leq 1$ is a prescribed threshold. With only small data corrupted by noise, the estimated posterior distribution $p(z_{ij}|\mathbf{Y}_i, \boldsymbol{\phi}_i)$ could be sensitive to the value of ϱ , and $\varrho = 0.5$ is used in this study. Next, the updated parameters m_{ij} , v_{ij} , and γ_{ij} derived from node i are subsequently treated as the priors m_{ji} , v_{ji} , and γ_{ji} for x_{ji} of node j in (2) and (3), which will be further updated to obtain $\mathbf{X}_j$ given nodal observation $\mathbf{Y}_j$ and measurement matrix $\boldsymbol{\phi}_j$. This effectively eschews the pitfall of conflicting results in the conventional sparse learning approaches reported in [7], [8].

B. Sequential Retrieval of Node Connectivity

Nonetheless, it is computationally daunting to retrieve connectivity $\mathbf{X}_i$ within a short time interval for networks with large N . Ideally, we can evaluate only a subset of the nodes to uncover the network connectivity with acceptable reconstruction error, leveraging proceeding reconstruction efforts. To select the optimal subset, the submodularity property will be investigated.

For a network with a finite set of nodes $\mathbb{Q} = 1, 2, \dots, N$, let θ_k denote the selected set of k nodes for connectivity reconstruction, then the reconstruction error is manifested in terms of the sum of square error (SSE), $SSE(\theta_k) = \|\mathbf{Y}_i - \phi_i \hat{\mathbf{X}}_i\|_2^2$. $SSE(\theta_k)$ embodies the sum of the discrepancy between the nodal observation $\mathbf{Y}_i$ and its estimate, and $\hat{\mathbf{X}}_i$ is the approximation of $\mathbf{X}_i$ for node i via the proposed sparse Bayesian learning approach. It is evident that the null set $\emptyset$ results in the maximum SSE without any knowledge of the connectivity. Accordingly, a utility function $f: 2^{\mathbb{Q}} \rightarrow \mathbb{R}$ can be defined to map any subset of $\mathbb{Q}$ (conventionally represented as $2^{\mathbb{Q}}$) to a real number. Herein, utility over a set θ_k is defined as $f(\theta_k) = SSE(\emptyset) - SSE(\theta_k)$. The larger the node set θ_k selected to reconstruct, the smaller $SSE(\theta_k)$, leading to monotonically increasing $f(\theta_k)$. Therefore, $f(\theta_k)$ is a monotonic submodular function with the following two properties:

- 1) *Monotonicity*: $f(\theta_1) \leq f(\theta_2)$ for all $\theta_1 \subseteq \theta_2 \subseteq \mathbb{Q}$. That said, addition of extra nodes always brings a non-negative change of the utility function in the network level. It is observed that the equality approximately holds when a large number of nodes are included and the $SSE(\theta_k)$ almost converges (see Section IV).
- 2) *Submodularity*: For $\theta_1 \subseteq \theta_2 \subseteq \mathbb{Q}$ and any element $s \in \mathbb{Q} \setminus \theta_2$, $f(\theta_1) - f(\theta_1 \cup s) \geq f(\theta_2) - f(\theta_2 \cup s)$, where $\mathbb{Q} \setminus \theta_2$ is the set of nodes in $\mathbb{Q}$ but not contained in the set θ_2 . This is also known as diminishing returns property, in that addition of extra node s to a smaller set θ_1 delivers larger utility than to a larger set θ_2 (also see Section IV).

Simple greedy algorithms prove effective for near-optimal subset selection for maximization of monotonic submodular functions [36]. Starting with a set θ_k (here, $\theta_0 = \emptyset$), the next node s_{k+1} is selected via

$$s_{k+1} = \operatorname{argmax}_{s \in \mathbb{Q} \setminus \theta_k} f(\theta_k \cup s) \quad (12)$$

and $\theta_{k+1} = \theta_k \cup \{s_{k+1}\}$. Here, the node with the largest nodal observation $\mathbf{Y}_i$ is selected initially without any connection information. The physical implication is that this node usually represents the most critical generator in the power grid system or the most critical player in the UG network. To find the best subsequent (most informative) node s_{k+1} to maximize $f(\theta_k \cup s)$, it is necessary to estimate $f(\theta_k \cup s)$ while incorporating the uncertainty associated with the estimation for all $s \in \mathbb{Q} \setminus \theta_k$, without computing the objective function. We tackle with this via Bayesian optimization [37] and build a surrogate for $f(\theta_k \cup s)$. Bayesian learning utilizes the surrogate function to

incorporate prior belief about the function and uses an acquisition function to decide where to evaluate the surrogate function next. Here, we define improvement by a utility function over selecting new node s given the retrieval node set θ_k

$$I_m(s|\theta_k) = \max(0, \epsilon_s - \epsilon_{\min}) \quad (13)$$

where $\epsilon_{\min} = \min_i \left(\|\mathbf{Y}_i - \phi_i \hat{\mathbf{X}}_i^{\theta_k}\|_2^2 \right)$ represents the minimum estimated error for $i \in \theta_k$ and $\epsilon_s = \|\mathbf{Y}_s - \phi_s \hat{\mathbf{X}}_s^{\theta_k}\|_2^2$ represents the predictive error of node $s \in \mathbb{Q} \setminus \theta_k$ given the retrieved information of the set θ_k . In such a manner, the most informative node s can be identified and then the sparse Bayesian learning algorithm is applied to reconstruct the network connectivity.

In concreteness, we initialize $m_{ij} = 0, v_{ij} = 1$ for $i, j \in [1, \dots, N]$, for the weight matrix $\mathbf{m}$ and variance matrix $\mathbf{v}$ in specification of the priors for $\mathbf{X}$ in (2). Only the node i in the retrieved set θ_k are fully estimated by the sparse Bayesian learning algorithm and the unretrieved node $s \in \mathbb{Q} \setminus \theta_k$ are partially recovered due to the symmetric of the networks. The fully estimated weight and variance for node i are $\hat{\mathbf{m}}_i^{\theta_k} = [\hat{m}_{i1} \dots \hat{m}_{iN}]^T$ and $\hat{\mathbf{v}}_i^{\theta_k} = [\hat{v}_{i1} \dots \hat{v}_{iN}]^T$; the partially recovered (predictive) weight and variance for node s are $\hat{\mathbf{m}}_s^{\theta_k} = [0 \dots \hat{m}_{si} \dots 0]^T$ and $\hat{\mathbf{v}}_s^{\theta_k} = [1 \dots \hat{v}_{si} \dots 1]^T$. In (10), $\hat{x}_{ij}^{\theta_k}$ is parametrized as a normal distribution, and we derive the EI acquisition function that strikes the tradeoff between local exploitation and global exploration

$$EI(s) = \epsilon \psi \left(\frac{\epsilon}{\|\hat{\mathbf{v}}_s^{\theta_k}\|_2^2} \right) + \|\hat{\mathbf{v}}_s^{\theta_k}\|_2^2 \mathbf{r} \left(\frac{\epsilon}{\|\hat{\mathbf{v}}_s^{\theta_k}\|_2^2} \right) \quad (14)$$

where $\epsilon = \|\mathbf{Y}_s - \phi_s \hat{\mathbf{m}}_s^{\theta_k}\|_2^2 - \min_i \left(\|\mathbf{Y}_i - \phi_i \hat{\mathbf{m}}_i^{\theta_k}\|_2^2 \right)$, $\psi(\cdot)$ and $\mathbf{r}(\cdot)$ are the CDF and PDF of standard normal distribution, respectively. The most informative node s_{k+1} to be selected is determined based on $EI(s)$ for $s \in \mathbb{Q} \setminus \theta_k$:

$$s_{k+1} = \operatorname{argmax}_{s \in \mathbb{Q} \setminus \theta_k} EI(s) \quad (15)$$

IV. NUMERICAL RESULTS

We demonstrate the performance of our reconstruction framework in two representative cases: the synthetic unweighted UG network and the weighted IEEE-118 power grid system. The weight here represents the different susceptance of the transmission lines. The UG network topology is generated according to the Barabási-Albert (BA) model. The BA model generates random scale-free networks according to the preferential attachment principle. For a network with c_0 nodes, new nodes are added one at a time and connected to $c \leq c_0$ existing nodes with a probability

proportional to the node degree. Two different BA models (low average degree with $c = 1$ and high average degree $c = 2$) are utilized to generate the UG network, and the initial number of nodes is $c_0 = c + 1$. For both the UG and power grid cases, the number of nodes in all network structures is $N = 118$ and the timesteps is $t_M = 60$ for recording data.

For comparison, we include a baseline model Random Spike and Slab (RandomSS), which uses the Spike and Slab prior in the sparse Bayesian learning approach to reveal the node connectivity but only randomly selects the next node for reconstruction. By contrast, the proposed Sequential Spike and Slab model (SeqSS) selects the next most informative node via EI. For each network, we conduct the connectivity recovery over 30 experiments with random noise on nodal observation using the following metrics: the Frobenius norm for the connectivity discrepancy

$$Error_{\hat{A}} = \frac{\|\mathbf{A} - \hat{\mathbf{A}}\|_F}{\|\mathbf{A}\|_F} \quad (16)$$

where $\mathbf{A}$ and $\hat{\mathbf{A}}$ denotes the true and estimated connectivity matrices, respectively; and the Frobenius norm for the observation and weight discrepancy

$$Error_{\hat{Y}} = \frac{\|\mathbf{Y} - \hat{\mathbf{Y}}\|_F}{\|\mathbf{Y}\|_F} \quad (17)$$

and

$$Error_{\hat{m}} = \frac{\|\mathbf{m} - \hat{\mathbf{m}}\|_F}{\|\mathbf{m}\|_F} \quad (18)$$

Here, $\hat{\mathbf{m}}$ and $\hat{\mathbf{A}}$ with entry z_{ij} are estimated from (10) and (11c). $z_{ij} = 1$ (an edge exists between node i and j) if $\gamma_{ij} > 0.5$, and otherwise $z_{ij} = 0$. $\hat{\mathbf{Y}}$ is indirectly estimated from (10): $\hat{\mathbf{Y}}_i = \phi_i \hat{\mathbf{m}}_i$ for weighted network and $\hat{\mathbf{Y}}_i = \phi_i \hat{\mathbf{A}}_i$ for unweighted network.

A. UG Network

In a UG, one player (the offeror) proposes how to split a sum (e.g., money or credits), and another player (the responder) decides whether to accept or reject the offer. Here, we have N players or nodes in the UG network, and two nodes play with each other if an edge exists in between. In each round of the

game, node i plays two games with a connected node j , in the role of an offeror and a responder, respectively. As such, the proposal of a player i is denoted by (o_i, a_i) : it offers credit o_i to other interactive responders and accepts credit for at least a_i from other interactive offerors. Only when the responder accepts the proposed strategy, the sum will be split accordingly. The measurement matrix ϕ_i , or more specifically, the payoff ϕ_{ij} (the entry of measurement matrix ϕ_i) of player i with strategy (o_i, a_i) playing against player j with strategy (o_j, a_j) is represented as

$$\phi_{ij} = \begin{cases} o_j + 100 - o_i, & o_i \geq a_j \text{ and } o_j \geq a_i \\ o_j, & o_i < a_j \text{ and } o_j \geq a_i \\ 100 - o_i, & o_i \geq a_j \text{ and } o_j < a_i \\ 0, & o_i < a_j \text{ and } o_j < a_i \end{cases} \quad (19)$$

Subsequently, the cumulative payoff y_i (the entry of nodal observation $\mathbf{Y}_i$) for player i with strategy (o_i, a_i) at each round is $y_i = \sum_{j \in U_i} \phi_{ij}$, where U_i is the set of interactive node j with i . In the evolutionary game, player i updates the strategy in a randomized manner, i.e., $(o(t+1), a(t+1)) = (h(o(t)), h(a(t)))$:

$$h(\delta) = \begin{cases} \delta + \xi, & \delta + \xi \in [0, 100] \\ 100 \times \left\lfloor \frac{\delta + \xi}{100} \right\rfloor, & \delta + \xi \notin [0, 100] \end{cases} \quad (20)$$

Here, $\xi \sim \text{Uniform}(-50, 50)$ and $\lfloor \cdot \rfloor$ is a floor function ensuring that the updated values of o and a are within the split sum (100 in this study) in the updating policy.

The observable data for any player i in the UG problem are time series $(o_i(t), a_i(t))$ and $y_i(t)$ during $t \in [t_1, t_M]$. We add noise $\boldsymbol{\varepsilon} = [\varepsilon_{t_1} \ \dots \ \varepsilon_{t_m} \ \dots \ \varepsilon_{t_M}]^T$ to the cumulative payoff $\mathbf{Y}$. Here, $\varepsilon_{t_m} \sim \mathcal{N}(\varepsilon_{t_m} | 0, (\beta \times \max(\mathbf{Y}))^2)$, $t_m = t_1, \dots, t_M$, where $\beta = \{0.01, 0.02, 0.03\}$ is the scale to the maximum absolute value in $\mathbf{Y}$. Thus, $\mathbf{Y}_i = \phi_i \mathbf{X}_i + \boldsymbol{\varepsilon}$.

$$\begin{bmatrix} y_i(t_1) \\ \vdots \\ y_i(t_M) \end{bmatrix} = \begin{bmatrix} \phi_{i1}(t_1) & \dots & \phi_{iN}(t_1) \\ \vdots & \ddots & \vdots \\ \phi_{i1}(t_M) & \dots & \phi_{iN}(t_M) \end{bmatrix} \begin{bmatrix} x_{i1} \\ \vdots \\ x_{iN} \end{bmatrix} + \begin{bmatrix} \varepsilon_{t_1} \\ \vdots \\ \varepsilon_{t_M} \end{bmatrix} \quad (21)$$

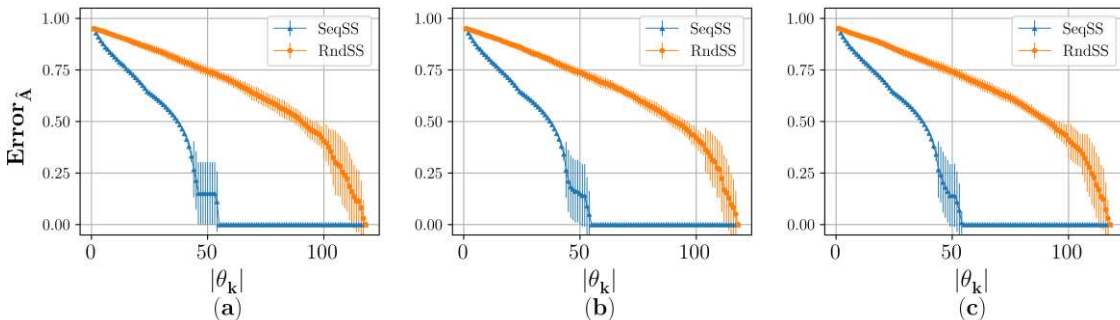


Fig. 1. Connectivity discrepancy of UG network for SeqSS and RandomSS algorithms with different noise levels (a) $\beta = 0.01$, (b) $\beta = 0.02$, and (c) $\beta = 0.03$ with $c = 1$ in the BA model.

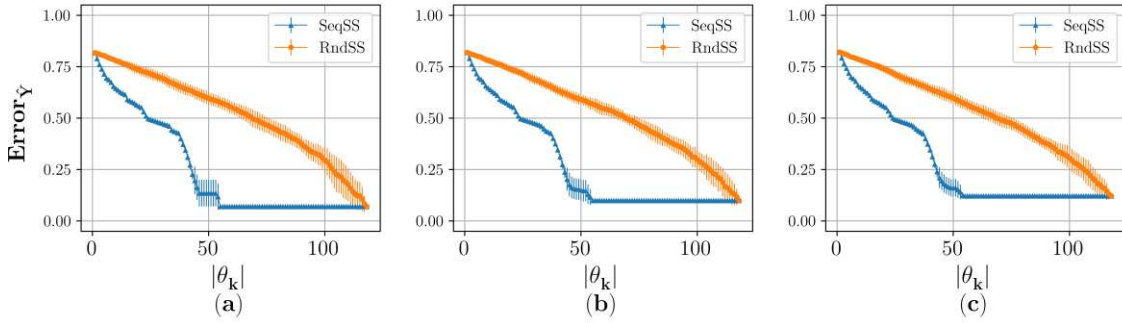


Fig. 2. Observation discrepancy of UG network for SeqSS and RandomSS algorithms with different noise levels (a) $\beta = 0.01$, (b) $\beta = 0.02$, and (c) $\beta = 0.03$ with $c = 1$ in the BA model.

Here, $x_{ij} = 1$ if $j \in U_i$, otherwise $x_{ij} = 0$. $\mathbf{Y}_i$ is generated by randomized repetitions of 30 experiments at each scale level β for the two BA network structures.

We compare the performance of SeqSS and RandomSS in terms of the connectivity discrepancy $Error_{\hat{A}}$ and the observation discrepancy $Error_{\hat{Y}}$ with different cardinality $|\theta_k|$ of the selected node set θ_k . The $Error_{\hat{A}}$ of the UG network with $c = 1$ in the BA model is depicted in Fig. 1. An error bar is also provided at each point denoting the standard deviation from the 30 experiments. SeqSS registers the same accuracy as RandomSS with edge construction for only $\sim 50\%$ of the nodes.

cardinality $|\theta_k|$ increases, the $Error_{\hat{Y}}$ edges down continuously. While a higher noise level renders a larger $Error_{\hat{Y}}$, SeqSS is overall robust to such perturbations.

The $Error_{\hat{A}}$ and $Error_{\hat{Y}}$ of the UG network with $c = 2$ in the BA network are displayed in Figs. 3 and 4, respectively. SeqSS has consistent performance on different network topologies.

It bears mentioning that the discrepancy variance (represented by the error bar) in SeqSS is notably smaller than that observed in RandomSS. This minor variability is attributed to the EI algorithm in SeqSS that consistently selects the most informative nodes.

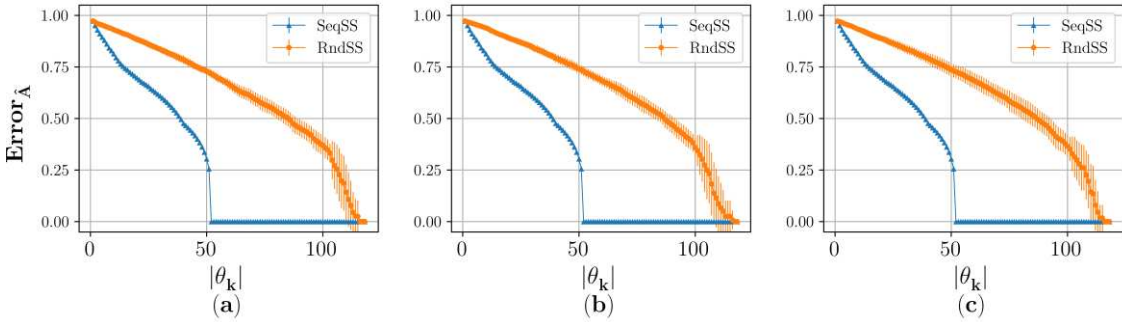


Fig. 3. Connectivity discrepancy of UG network for SeqSS and RandomSS algorithms with different noise levels (a) $\beta = 0.01$, (b) $\beta = 0.02$, and (c) $\beta = 0.03$ with $c = 2$ in the BA model.

The observation discrepancy $Error_{\hat{Y}}$ of the UG network with $c = 1$ in the BA model is depicted in Fig. 2. As the

B. Power Grid System

The topology of the power grid system could remain

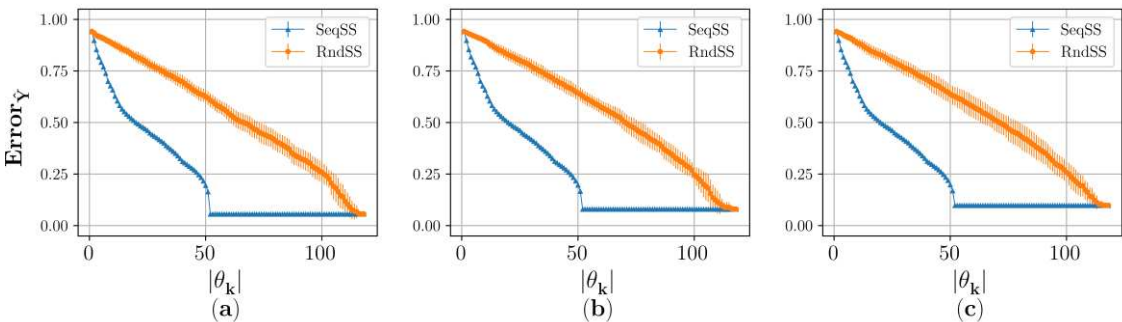


Fig. 4. Observation discrepancy of UG network for SeqSS and RandomSS algorithms with different noise levels (a) $\beta = 0.01$, (b) $\beta = 0.02$, and (c) $\beta = 0.03$ with $c = 2$ in the BA model.

unknown at the onset of blunt perturbations, such as inclement weather or operation glitches. With the simplified direct current (DC) approximation, the power flow P_{ij} from node i to j (e.g., generator or load) over the transmission line with reactance r_{ij} is given as:

$$P_{ij} = \frac{|V_i| \cdot |V_j|}{r_{ij}} \sin(\varphi_i - \varphi_j) \quad (22)$$

where $|V_i|$ and $|V_j|$ are the voltage magnitudes, and φ_i and φ_j are the phase angles of node i and j , respectively. Here, we simulate the phase angle variation on the power grid system: the phase angle $\varphi_i(t) = (\omega + \Delta\omega_i)t$, where $\omega = 2\pi \times 50$ is the

voltage magnitude $|V_i|$ is set to the unit value for all nodes for simplicity. Thereby, the effective power balance at node i is delineated as $y_i = \sum_{j \in U_i} P_{ij} = \sum_{j \in U_i} \frac{\sin(\varphi_i - \varphi_j)}{r_{ij}}$, where U_i is the set of connected node j with i . The sensing data from the phasor measurement units for node i include the phase angle φ_i and power flow y_i during $t \in [t_1, t_M]$. Similarly, we include noise $\boldsymbol{\varepsilon} = [\varepsilon_{t_1} \dots \varepsilon_{t_m} \dots \varepsilon_{t_M}]^T$ to the power flow $\mathbf{Y}$, and $\varepsilon_{t_m} \sim \mathcal{N}(\varepsilon_{t_m} | 0, (\beta \times \max(\mathbf{Y}))^2)$, $\beta = \{0.01, 0.02, 0.03\}$. Likewise, $\mathbf{Y}_i = \boldsymbol{\phi}_i \mathbf{X}_i + \boldsymbol{\varepsilon}$

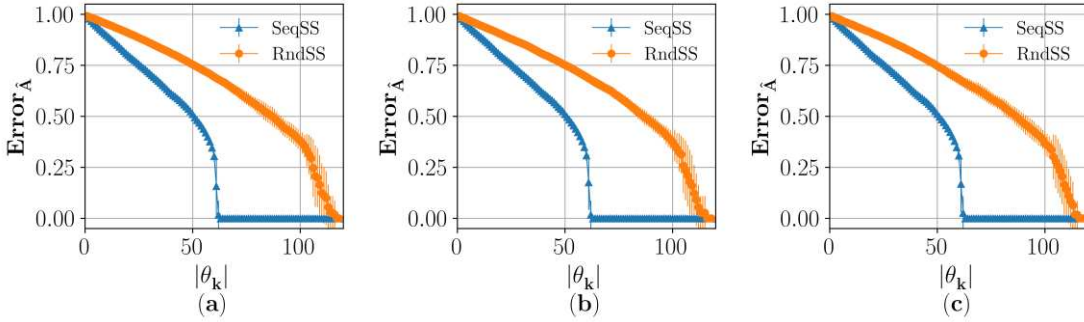


Fig. 5. Connectivity discrepancy of the power grid system for SeqSS and RandomSS algorithms with different noise levels (a) $\beta = 0.01$, (b) $\beta = 0.02$, and (c) $\beta = 0.03$ under IEEE-118 network.

angular frequency of grid operation and $\Delta\omega_i \sim \mathcal{N}(0, 20)$ is the random frequency perturbation for node i . Moreover, the

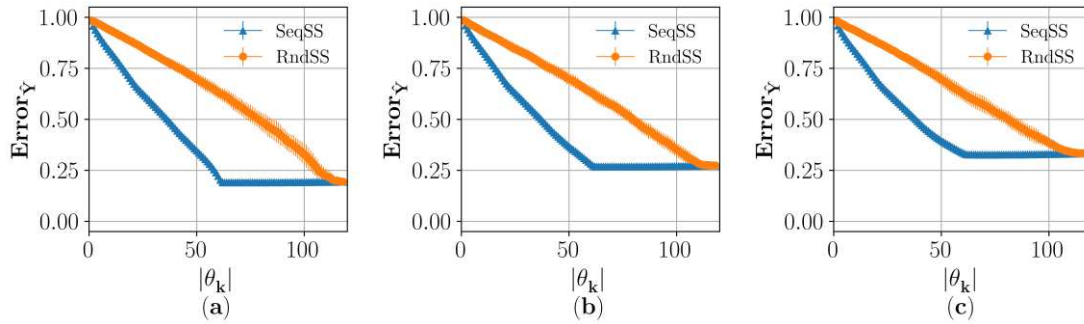


Fig. 6. Observation discrepancy of the power grid system for SeqSS and RandomSS algorithms with different noise levels (a) $\beta = 0.01$, (b) $\beta = 0.02$, and (c) $\beta = 0.03$ under IEEE-118 network.

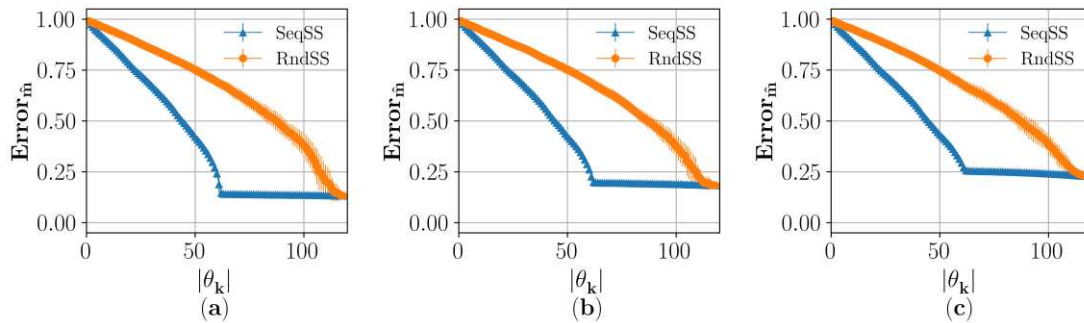


Fig. 7. Weight discrepancy of the power grid system for SeqSS and RandomSS algorithms with different noise levels (a) $\beta = 0.01$, (b) $\beta = 0.02$, and (c) $\beta = 0.03$ under IEEE-118 network.

$$\begin{bmatrix} y_i(t_1) \\ \vdots \\ y_i(t_M) \end{bmatrix} = \begin{bmatrix} \phi_{i1}(t_1) & \cdots & \phi_{iN}(t_1) \\ \vdots & \ddots & \vdots \\ \phi_{i1}(t_M) & \cdots & \phi_{iN}(t_M) \end{bmatrix} \begin{bmatrix} x_{i1} \\ \vdots \\ x_{iN} \end{bmatrix} + \begin{bmatrix} \varepsilon_{t_1} \\ \vdots \\ \varepsilon_{t_M} \end{bmatrix} \quad (23)$$

Here, $\phi_{ij}(t) = \sin(\varphi_i(t) - \varphi_j(t))$ and susceptance $x_{ij} = \frac{1}{r_{ij}} > 0$ if $j \in U_i$, otherwise $r_{ij} = \infty$ and $x_{ij} = 0$. Thus, x_{ij} is the weight of edge in the grid network. $\mathbf{Y}_i$ is generated by randomized repetitions of 30 experiments at each scale level β for the IEEE-118 network structure.

In this case, in addition to the connectivity discrepancy $Error_{\hat{\alpha}}$ and observation discrepancy $Error_{\hat{\gamma}}$, we also include the weight discrepancy $Error_{\hat{m}}$ to compare SeqSS and RandomSS. In Fig. 5, we show that SeqSS reaches the same level of $Error_{\hat{\alpha}}$ as RandomSS with only a fraction of the recovered nodes.

The observation discrepancy $Error_{\hat{\gamma}}$ is shown in Fig. 6. Same level of $Error_{\hat{\gamma}}$ is attained in SeqSS as that of RandomSS, with around half of the nodes interrogated. Similarly, the discrepancy variance in SeqSS is significantly smaller than that in RandomSS. In Fig. 7, the $Error_{\hat{m}}$ of the IEEE-118 network structure is exhibited. Here, with only a fraction of the recovered nodes, the weight discrepancy of SeqSS approaches the same level as RandomSS. Furthermore, different from the unweighted UG network, estimate of the weight matrix in the grid network entails more nodes to converge, and the accuracy is sensitive to the noise.

V. CONCLUSION AND DISCUSSION

In this study, we develop a sparse Bayesian approach based on the Spike and Slab prior for sequential reconstruction of network connectivity. We corroborate this approach on an unweighted UG network (with two different BA structures) and a weighted power grid system (the IEEE-118 network). Extensive studies imply that our proposed SeqSS algorithm identifies the network connectivity in an efficient fashion compared to RandomSS. Owing to the EI algorithm to select the most informative node, the SeqSS algorithm notches smaller discrepancy variance than that of RandomSS.

Note that the computational bottleneck in Bayesian solution for large-scale networks consists in the Bayesian approximation and node-by-node iteration. Our method is leaned towards the node iteration: we only utilize approximately 50% of nodes to recover the network structure, thereby reducing the computational burden associated with the Bayesian learning process. This study has the potential to significantly scale up connectivity reconstruction for massive networks, radically transform the monitoring and operation for various realistic networked systems, including the power grid, transportation and communication. This provides a new robust and efficient foundation for operational decision making, such as the monitoring and maintenance of the transportation or communication network or power grid system. We will also investigate innovations to accelerate approximation of Bayesian posterior in our future study.

Another important area for future investigation is the

problem of data distribution shifts in dynamic networks. This present study addresses connectivity recovery for static networks within a short time interval. Hence, only small data are recorded and they shall follow the same distribution. Remarkably, data distributions may shift in the regime of dynamic networks, which entails out-of-distribution generalization, as discussed in recent studies [38], [39], [40], [41], for accurate network inference. Expanding our focus to dynamic networks will improve the applicability and robustness of this study, offering new insights and capabilities in network inference.

ACKNOWLEDGMENT

This work was partially supported by the U.S. National Science Foundation (2119334, 1927418 and 1927425).

REFERENCES

- [1] J. Wan, G. Ichinose, M. Small, H. Sayama, Y. Moreno, and C. Cheng, "Multilayer networks with higher-order interaction reveal the impact of collective behavior on epidemic dynamics," *Chaos, Solitons & Fractals*, vol. 164, p. 112735, Nov. 2022, doi: 10.1016/j.chaos.2022.112735.
- [2] Y. Che and C. Cheng, "Active learning and relevance vector machine in efficient estimate of basin stability for large-scale dynamic networks," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 31, no. 5, p. 053129, May 2021, doi: 10.1063/5.0044899.
- [3] I. Klickstein, A. Shirin, and F. Sorrentino, "Locally optimal control of complex networks," *Phys. Rev. Lett.*, vol. 119, no. 26, p. 268301, Dec. 2017, doi: 10.1103/PhysRevLett.119.268301.
- [4] Y. Wang, "Resilience quantification for probabilistic design of cyber-physical system networks," *ASME J. Risk Uncertainty Part B*, vol. 4, no. 031006, Mar. 2018, doi: 10.1115/1.4039148.
- [5] S. Cao and H. Sayama, "Detecting dynamic states of temporal networks using connection series tensors," *Complexity*, vol. 2020, p. e9649310, Dec. 2020, doi: 10.1155/2020/9649310.
- [6] S. Hempel, A. Koseska, J. Kurths, and Z. Nikoloski, "Inner composition alignment for inferring directed networks from short time series," *Phys. Rev. Lett.*, vol. 107, no. 5, p. 054101, Jul. 2011, doi: 10.1103/PhysRevLett.107.054101.
- [7] F. Basiri, J. Casadiego, M. Timme, and D. Witthaut, "Inferring power-grid topology in the face of uncertainties," *Phys. Rev. E*, vol. 98, no. 1, p. 012305, Jul. 2018, doi: 10.1103/PhysRevE.98.012305.
- [8] X. Han, Z. Shen, W.-X. Wang, and Z. Di, "Robust reconstruction of complex networks from sparse data," *Phys. Rev. Lett.*, vol. 114, no. 2, p. 028701, Jan. 2015, doi: 10.1103/PhysRevLett.114.028701.
- [9] M. Timme and J. Casadiego, "Revealing networks from dynamics: an introduction," *J. Phys. A: Math. Theor.*, vol. 47, no. 34, p. 343001, Aug. 2014, doi: 10.1088/1751-8113/47/34/343001.
- [10] D. Yu, M. Righero, and L. Kocarev, "Estimating topology of networks," *Phys. Rev. Lett.*, vol. 97, no. 18, p. 188701, Nov. 2006, doi: 10.1103/PhysRevLett.97.188701.
- [11] D. Yu and U. Parlitz, "Inferring local dynamics and connectivity of spatially extended systems with long-range links based on steady-state stabilization," *Phys. Rev. E*, vol. 82, no. 2, p. 026108, Aug. 2010, doi: 10.1103/PhysRevE.82.026108.
- [12] J. Ren, W.-X. Wang, B. Li, and Y.-C. Lai, "Noise bridges dynamical correlation and topology in coupled oscillator networks," *Phys. Rev. Lett.*, vol. 104, no. 5, p. 058701, Feb. 2010, doi: 10.1103/PhysRevLett.104.058701.
- [13] J. F. Donges, Y. Zou, N. Marwan, and J. Kurths, "The backbone of the climate network," *EPL*, vol. 87, no. 4, p. 48007, Sep. 2009, doi: 10.1209/0295-5075/87/48007.
- [14] L. Gao, L. Sommerlade, B. Coffman, T. Zhang, J. M. Stephen, D. Li, J. Wang, C. Grebogi, and B. Schelter, "Granger causal time-dependent source connectivity in the somatosensory network," *Sci Rep*, vol. 5, no. 1, Art. no. 1, May 2015, doi: 10.1038/srep10399.
- [15] V. A. Makarov, F. Panetsos, and O. de Feo, "A method for determining neural connectivity and inferring the underlying network

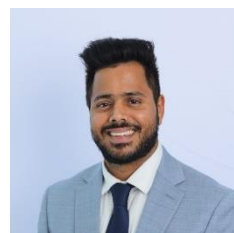
- dynamics using extracellular spike recordings,” *J Neurosci Methods*, vol. 144, no. 2, pp. 265–279, Jun. 2005, doi: 10.1016/j.jneumeth.2004.11.013.
- [16] X. Wu, C. Zhou, G. Chen, and J. Lu, “Detecting the topologies of complex networks with stochastic perturbations,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 21, no. 4, p. 043129, Dec. 2011, doi: 10.1063/1.3664396.
- [17] X. Wu, W. Wang, and W. X. Zheng, “Inferring topologies of complex networks with hidden variables,” *Phys. Rev. E*, vol. 86, no. 4, p. 046106, Oct. 2012, doi: 10.1103/PhysRevE.86.046106.
- [18] W.-X. Wang, Y.-C. Lai, and C. Grebogi, “Data based identification and prediction of nonlinear and complex dynamical systems,” *Phys. Rep.*, vol. 644, pp. 1–76, Jul. 2016, doi: 10.1016/j.physrep.2016.06.004.
- [19] A. Shamsan, X. Wu, P. Liu, and C. Cheng, “Intrinsic recurrence quantification analysis of nonlinear and nonstationary short-term time series,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 30, no. 9, p. 093104, Sep. 2020, doi: 10.1063/5.0006537.
- [20] X. Wu, Y. Zheng, Y. Che, and C. Cheng, “Pattern recognition and automatic identification of early-stage atrial fibrillation,” *Expert Syst. Appl.*, vol. 158, p. 113560, Nov. 2020, doi: 10.1016/j.eswa.2020.113560.
- [21] Y. Che, Z. (John) Zhang, and C. Cheng, “Physical–statistical learning in resilience assessment for power generation systems,” *Physica A*, vol. 615, p. 128584, Apr. 2023, doi: 10.1016/j.physa.2023.128584.
- [22] Z. Shen, W.-X. Wang, Y. Fan, Z. Di, and Y.-C. Lai, “Reconstructing propagation networks with natural diversity and identifying hidden sources,” *Nat Commun*, vol. 5, no. 1, Art. no. 1, Jul. 2014, doi: 10.1038/ncomms5323.
- [23] M. G. Leguía, C. G. B. Martínez, I. Malvestio, A. T. Campo, R. Rocamora, Z. Levnajić, and R. G. Andrzejak, “Inferring directed networks using a rank-based connectivity measure,” *Phys. Rev. E*, vol. 99, no. 1, p. 012319, Jan. 2019, doi: 10.1103/PhysRevE.99.012319.
- [24] K. Huang, Y. Jiao, C. Liu, W. Deng, and Z. Wang, “Robust network structure reconstruction based on Bayesian compressive sensing,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 29, no. 9, p. 093119, Sep. 2019, doi: 10.1063/1.5109375.
- [25] K. Huang, W. Deng, Y. Zhang, and H. Zhu, “Sparse Bayesian learning for network structure reconstruction based on evolutionary game data,” *Physica A*, vol. 541, p. 123605, Mar. 2020, doi: 10.1016/j.physa.2019.123605.
- [26] Y. Shu, W. Dan, and C. Cheng, “Spatiotemporal regularization in effective reconstruction of epicardial potential,” in *2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*, Jul. 2021, pp. 1–4, doi: 10.1109/BHI50953.2021.9508525.
- [27] Z. Zhou, K. Liu, and J. Fang, “Bayesian compressive sensing using normal product priors,” *IEEE Signal Process Lett*, vol. 22, no. 5, pp. 583–587, May 2015, doi: 10.1109/LSP.2014.2364255.
- [28] S. D. Babacan, R. Molina, and A. K. Katsaggelos, “Bayesian compressive sensing using laplace priors,” *IEEE Trans Image Process*, vol. 19, no. 1, pp. 53–63, Jan. 2010, doi: 10.1109/TIP.2009.2032894.
- [29] C. M. Carvalho, N. G. Polson, and J. G. Scott, “Handling Sparsity via the Horseshoe,” in *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics*, Apr. 2009, Accessed: Aug. 03, 2024. [Online]. Available: <https://proceedings.mlr.press/v5/carvalho09a.html>
- [30] T. Park and G. Casella, “The Bayesian Lasso,” *Journal of the American Statistical Association*, vol. 103, no. 482, pp. 681–686, Jun. 2008, doi: 10.1198/016214508000000337.
- [31] M. E. Tipping, “Sparse Bayesian learning and the relevance vector machine,” *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, Sep. 2001, doi: 10.1162/15324430152748236.
- [32] H. Ishwaran and J. S. Rao, “Spike and Slab variable selection: Frequentist and Bayesian strategies,” *Ann. Stat.*, vol. 33, no. 2, pp. 730–773, Apr. 2005, doi: 10.1214/009053604000001147.
- [33] D. Hernández-Lobato, J. M. Hernández-Lobato, and P. Dupont, “Generalized Spike-and-Slab priors for Bayesian group feature selection using expectation propagation,” *J. Mach. Learn. Res.*, vol. 14, no. 59, pp. 1891–1945, 2013.
- [34] J. M. Hernández-Lobato, D. Hernández-Lobato, and A. Suárez, “Expectation propagation in linear regression models with Spike-and-Slab priors,” *Mach Learn*, vol. 99, no. 3, pp. 437–487, Jun. 2015, doi: 10.1007/s10994-014-5475-7.
- [35] M. I. Jordan, “An Introduction to probabilistic graphical models.” Accessed: Dec. 27, 2023. [Online]. Available: <https://people.eecs.berkeley.edu/~jordan/prelims/>
- [36] R. Zhu, B. Yao, H. Yang, and F. Leonelli, “Optimal sensor placement for space–time potential mapping and data fusion,” *IEEE Sensors Letters*, vol. 3, no. 1, pp. 1–4, Jan. 2019, doi: 10.1109/LSNS.2018.2884205.
- [37] Y. Che, J. Liu, and C. Cheng, “Multi-fidelity modeling in sequential design for stability identification in dynamic time-delay systems,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 29, no. 9, p. 093105, Sep. 2019, doi: 10.1063/1.5097934.
- [38] A. Canatar, B. Bordelon, and C. Pehlevan, “Out-of-distribution generalization in kernel regression,” in *Proceedings of the 35th International Conference on Neural Information Processing Systems*, in NIPS ’21, Red Hook, NY, USA: Curran Associates Inc., Jun. 2024, pp. 12600–12612.
- [39] N. Ye, K. Li, H. Bai, R. Yu, L. Hong, F. Zhou, Z. Li, and J. Zhu, “OoD-Bench: Quantifying and Understanding Two Dimensions of Out-of-Distribution Generalization,” in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2022, pp. 7937–7948, doi: 10.1109/CVPR52688.2022.00779.
- [40] X. Zhao, S. Bian, Y. Zhang, Y. Zhang, Q. Gu, X. Wang, C. Zhou, and N. Ye, “Domain Invariant Learning for Gaussian Processes and Bayesian Exploration,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, pp. 17024–17032, Mar. 2024, doi: 10.1609/aaai.v38i15.29646.
- [41] L. Zhu, W. Yin, Y. Yang, F. Wu, Z. Zeng, Q. Gu, X. Wang, C. Zhou, and N. Ye, “Vision-Language Alignment Learning Under Affinity and Divergence Principles for Few-Shot Out-of-Distribution Generalization,” *International Journal of Computer Vision*, pp. 1–33, Mar. 2024, doi: 10.1007/s11263-024-02036-4.



Jinming Wan received the B.E. degree in Forest Chemical Engineering from the Northwest Agriculture & Forestry University, Yangling, China, in 2015, and the M.S. degrees in Engineering and Technology Management from Morehead State University, Morehead, KY, U.S, in 2018. His research focuses on modeling and analysis, uncertainty quantification, machine learning, dynamics of complex systems, and optimal control techniques.



Jun Kataoka received his B.S. (2016) at Chuo University, Tokyo, Japan, MSc (2017) at University of Warwick, Coventry, UK and is currently a Ph.D. candidate in the Systems Science and Industrial Engineering department at SUNY Binghamton. He is working on physics-informed machine learning and domain adaptation applications in smart manufacturing as part of his Ph.D. dissertation.



Jayanth Sivakumar received an M.S. in Computer Science and a Ph.D. in Systems Science from Binghamton University. His research focuses on synthetic data generation methodologies for structured datasets to tackle small dataset problems. He is currently a senior data scientist at

Citibank N.A. He leads the optimization of the target communication channel for Citi's retail offers.



Eric Peña received a B.S. in Engineering Physics from The Ohio State University and an M.S. in Systems Science from Binghamton University. He was a Clifford D. Clark Fellow, received an Advanced Graduate Certificate in Complex Systems modeling, and was a member of the Binghamton Center of Complex Systems (CoCo). His research interests include modeling the dynamics and statistical properties of complex systems.



Yiming Che received the B.S. degree in Industrial Engineering from Capital University of Economics and Business, Beijing, China, in 2017, the M.S. degree in Industrial Engineering from the Binghamton University, Binghamton, NY, USA, in 2018, and the Ph.D. degree in Systems Science from the Binghamton University, Binghamton, NY, USA, in 2023. He joined the School of Computing and Augmented Intelligence, Arizona State University, Tempe, AZ, as a post-doctoral research fellow supervised by Dr. Teresa Wu, in 2023. His current research interests include computer vision in medical imaging, Bayesian inference and active learning.



Hiroki Sayama (Member, IEEE) is a Professor in the Department of Systems Science and Industrial Engineering, and the Director of the Binghamton Center of Complex Systems, at Binghamton University, State University of New York, USA. He also serves as a non-tenure-track Professor in the School of Commerce at Waseda University, Japan, as well as an External Faculty member of the Vermont Complex Systems Center at the University of Vermont, USA. He received his B.Sc., M.Sc. and D.Sc. in Information Science, all from the University of Tokyo, Japan. His research interests include complex dynamical networks, human and social dynamics, collective behaviors, artificial life/chemistry, interactive systems, and complex systems education, among others. He is an expert on mathematical/computational modeling and analysis of various complex systems.



Changqing Cheng (Member, IEEE) is an associate professor in the Department of Systems Science and Industrial Engineering at Binghamton University. His research interests include nonlinear dynamics analysis, statistical learning, simulation and modeling for process monitoring, quality control, and performance optimization of complex systems.