14th International Conference on Applications of Statistics and Probability in Civil Engineering, ICASP14
Dublin, Ireland, July 9-13, 2023

A Bayesian Approach for Inferring the Topology of Interdependent Infrastructure Networks from Cascading Failure Data

Yu Wang
Ph.D. Candidate, Computer Science, Vanderbilt Univ., Nashville, USA

Jinzhu Yu
Assistant Professor, Civil and Environmental Eng., Univ. of Texas, Arlington, USA

Hib Baroud
Associate Professor, Civil and Environmental Eng., Vanderbilt Univ., Nashville, USA

ABSTRACT: Accurate information on the topology of interdependent infrastructure networks is necessary to understand the behavior of interdependent infrastructure networks under uncertainty and evaluate their vulnerability to disruptions. Such information is often missing or not available. To address this issue, we propose a Bayesian approach to infer the topology of interdependent infrastructure networks from cascading failure data. This approach utilizes the hierarchical stochastic block model to generate the initial networks along with their corresponding prior probability and employs the susceptible-infectious epidemic spreading model to calculate the likelihood of the observed cascading failure sequence. Specifically, we use Metropolis-Hastings algorithm with the proposal designed specifically for interdependent infrastructure networks, the infrastructure-dependent proposal, to obtain the posterior distribution of network topology. A case study on inferring the topology of a synthetic system of the U.K. interdependent power-gas system is conducted to demonstrate the effectiveness of our approach in reconstructing the topology of interdependent infrastructure networks.

1. INTRODUCTION

1.1. Motivation and Problem Statement

Networks have recently become a crucial tool for describing features and analyzing behaviors of complex systems, where individual entities are represented as nodes and their connections or interactions are denoted as links. Examples include modeling online social media as social networks to quantitatively measure central roles of entities, (Kolaczyk, 2017), framing infrastructure networks as multi-networks and using flow theory to assess their performance during disasters (Ouyang et al., 2012), and abstracting drug-disease interactions as the biological networks to discover related drugs or classify proteins according to their biological function. Such network-based tasks are performed using corresponding real networks as platforms where complete information of the network topology is required (Ouyang et al., 2012). However, for some of these real-world networks, such as infrastructure networks, complete topology information is not available either due to security concerns or due to lack of shared data and knowledge across different types of infrastructure networks (e.g., water, power, transportation, and gas). As a result, researchers have turned to network simulation methods to generate network topology used for downstream tasks (Fu et al., 2016). Even though networks generated by simulation methods are similar to the original networks, strong priors and constraints are also embedded in the simulated network topology at the same time. Furthermore, network simulation methods are only designed for generating limited types of networks, which leads to a nar-

row range of real-world applications (Erdős et al., 1960). Due to these limitations and the advances in data collection and storage, recent studies have focused on network reconstruction methods to reconstruct the network topology given partial information.

1.2. Contributions

The ability to learn the topology of interdependent infrastructure networks will enable research advances in data-driven approaches for infrastructure performance modeling. Given the limitations of parametric estimation methods and the computational intractability of traditional nonparametric reconstruction methods, we devise a novel network reconstruction approach which considers topology constrains of real-world infrastructure networks, and thus, significantly decreases the space of the candidate set of networks. We apply this approach in a Bayesian inference framework to reconstruct the network topology based on a specific dynamic process occurring in interdependent infrastructure networks which is the cascading failure effect resulting from disruptions. A case study on reconstructing the topology of a synthetic system of the U.K. gas-power networks is conducted to demonstrate the effectiveness of our approach.

The rest of the paper is organized as follows. Section 2 introduces the general Bayesian framework in reconstructing network topology based on the disrupted network performance. The Metropolis-Hastings algorithm for network reconstruction and the infrastructure-dependent proposal are described in Section 3 with numerical results provided in Section 4 and concluding remarks in Section 5.

2. Bayesian Network reconstruction

In this section, a Bayesian framework to reconstruct the network topology based on the disrupted network performance is proposed. In graph theory, the information of the network topology is completely represented by its adjacency matrix $\mathbf{A}$. Therefore applying Bayes' theorem to reconstruct the network topology based on disrupted network performance $\mathbf{C}$ is achieved by updating the knowledge on the adjacency matrix $\mathbf{A}$, i.e., the posterior of $\mathbf{A}$:

$$P(\mathbf{A}|\mathbf{C}) = \frac{P(\mathbf{C}|\mathbf{A})P(\mathbf{A})}{P(\mathbf{C})} \propto P(\mathbf{C}|\mathbf{A})P(\mathbf{A}) \qquad (1)$$

where the likelihood $P(\mathbf{C}|\mathbf{A})$ measures the probability that the disruption $\mathbf{C}$ occurs on the network of particular topology $\mathbf{A}$; $P(\mathbf{A})$ is the prior information on the network topology; $P(\mathbf{C}) = \sum_{\mathbf{A}} P(\mathbf{C}|\mathbf{A})P(\mathbf{A})$ is the normalization term that indicates the total evidence for the disruption data $\mathbf{C}$. The calculation of the likelihood $P(\mathbf{C}|\mathbf{A})$ and the network prior $P(\mathbf{A})$ are introduced in the following section.

2.1. Graph Prior Model

Different types of networks with their prior probabilities $P(\mathbf{A})$ are generated by corresponding graph models. The Erdos-Renyi (ER) model generates networks with null hypothesis while two other models, the Barabasi-Albert (BA) model and the Watts-Strogatz (WS) model impose constraints on the generated networks. The former generates scale-free networks by preferential attachment mechanism while the latter generates small-world networks with short average path lengths and high clustering. However, most of real-world interdependent infrastructure networks are grouped by clusters and composed of hierarchical levels, which requires a more complex and realistic model such as the Hierarchical Stochastic Block model (HSBM) (Amini et al., 2022). The HSBM splits networks into multiple blocks and within each block, divide nodes into hierarchical levels. For example, the HSBM can be applied to model a system of power-water-gas networks using three blocks, each corresponding to water, power and gas networks and three layers within each block that each represent supply, transmission and demand nodes). The HSBM is used as the prior graph model to calculate the prior probability $P(\mathbf{A})$ since the topology of interdependent infrastructure networks can be fully characterized by HSBM. Consider a multiplex network with B blocks such that the number of nodes in each block is $n_b, b \in \mathbf{B} = \{1,2,...,B\}$. Within each single block b, a hierarchical structure is added so that the n_b nodes in block b are further divided into l levels. We assume that different

blocks share the same hierarchical structure. Denoting the block and layer labels of the node i as b_i and l_i, each element $\mathbf{A}_{ij}$ of the adjacency matrix $\mathbf{A}$ is an independent Bernoulli random variable with probability p_{ij}, a function of the block and layer labels, b_i, b_j, l_i, l_j. Therefore, the prior probability $P(\mathbf{A})$ conditioned on the block and layer assignments $\mathbf{B}, \mathbf{L}$ of the multiplex network is calculated as:

$$P(\mathbf{A}|\mathbf{B},\mathbf{L}) = \prod_{\forall i,j\in N, i\neq j} p_{ij}^{\mathbf{A}_{ij}}(1-p_{ij})^{1-\mathbf{A}_{ij}} \quad (2)$$

$$p_{ij} = f(b_i, b_j, l_i, l_j), \ \ \forall i,j \in N, \ \ i \neq j \quad (3)$$

where N is the set of nodes in the multiplex network and f is the function calculating the edge probability p_{ij} between nodes i, j based on their block b_i, b_j and layer l_i, l_j. In interdependent infrastructure networks, b corresponds to individual infrastructures such as power grid or water and gas distribution networks, and l corresponds to different types of facilities such as power supply nodes, i.e., gate stations, or water transmission nodes, storage tanks.

2.2. *Cascading Failure Model*

The knowledge on network topology is updated using the likelihood $P(\mathbf{C}|\mathbf{A})$ with a target topology that reproduces the diffusion of the cascading failure data. We consider the susceptible-infected (SI) model to quantitatively analyze the cascading failure (Peixoto, 2019). The set of information cascades $\mathbf{C}$ contains independent cascading scenario $\mathbf{C}^i, i \in \{1,2,...,C\}$ and each individual cascading scenario is a failure sequence where $\mathbf{C}^i_{tj} = 1$ indicates that node j has already failed at time t in the ith cascade. Node j fails when at least one of its adjacent nodes fails and this failure successfully propagates to node j. Consider the failure propagation of node j from time step t to $t+1$, shown in Fig. 1. The probability of node j failing, shown in Eq. 4, is calculated as a function of the complementary event that the failure of nodes $1,2,,...,v-1$ does not propagate to node j along corresponding links.

$$\begin{aligned} P(\mathbf{C}^i_{t+1,j} = 1 | \mathbf{C}^i_{t,j} = 0, \mathbf{C}^i_{t,1} = 1..., \mathbf{C}^i_{t,v} = 0) \\ = 1 - (1-q_{1j})(1-q_{2j})\cdots(1-q_{v-1j}) \end{aligned} \quad (4)$$

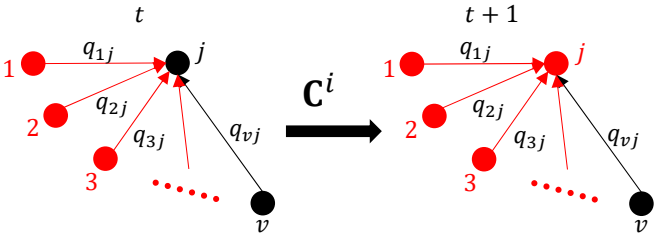

Figure 1: Failure propagation of node j from time step t to $t+1$. Failed nodes are shown in red while operational nodes are shown in black.

where q_{ij} denotes the probability of failure propagation along the link ij from node i to j.

Due to the independence between any two cascading scenarios $\mathbf{C}^i, \mathbf{C}^j (i \neq j), \forall i,j \in \{1,...,C\}$, the likelihood of cascading failure data $\mathbf{C}$ for the entire network considers the failure propagation for a single node and a single time step (Eq. 4) across all cascading failure scenarios, over all time steps for the entire disruption duration, and among all nodes in the network:

$$P(\mathbf{C}|\mathbf{A}) = \prod_{i=1}^{C}\prod_{t=1}^{T(\mathbf{C}^i)}\prod_{j=1}^{|N|} P(\mathbf{C}^i_{t+1,j}|\mathbf{C}^i_t) \quad (5)$$

where $T(\mathbf{C}^i)$ is the number of the time steps that the cascading failure scenario $\mathbf{C}^i$ lasts; $P(\mathbf{C}^i_{t+1,j}|\mathbf{C}^i_t)$ is the probability of the status (failed/not failed) of node j at time step $t+1$. Considering the uninfected status, Eq. 4 becomes:

$$\begin{aligned} &P(\mathbf{C}^i_{t+1,j}|\mathbf{C}^i_t) \\ &= (1 - \prod_{k\in\{N\setminus j\}} (1 - q_{kj}\mathbf{A}_{kj}\mathbf{C}^i_{t,k}(1-\mathbf{C}^i_{t-1,k})))^{\mathbf{C}^i_{t+1,j}(1-\mathbf{C}^i_{t,j})} \\ &(\prod_{k\in\{N\setminus j\}} (1 - q_{kj}\mathbf{A}_{kj}\mathbf{C}^i_{t,k}(1-\mathbf{C}^i_{t-1,k})))^{(1-\mathbf{C}^i_{t+1,j})(1-\mathbf{C}^i_{t,j})} \end{aligned} \quad (6)$$

where the first half represents the probability of failure propagating to node j while the second half represents its complementary event, the probability of failure not propagating to node j. The complementary property is guaranteed since only one of the two binary multiplication terms $\mathbf{C}^i_{t+1,j}(1-\mathbf{C}^i_{t,j})$ and $(1-\mathbf{C}^i_{t+1,j})(1-\mathbf{C}^i_{t,j})$ can be 1 and the other is 0. The term $\mathbf{C}^i_{t,k}(1-\mathbf{C}^i_{t-1,k})$ ensures that only nodes that have failed in the preceding time step will influence the failure of nodes at the current

time step, which characterizes the cascading mechanisms of the infrastructures such as progressive collapse of building structures and gradual outage of power stations (Adam 2018). For cases where nodes that have failed several time steps ago can still affect nodes at later time steps such as news diffusion and epidemic spread in social networks, one can remove the term $\mathbf{C}^i_{t-1,k}$ to incorporate these cases into the cascading failure model.

Conditioning the network topology on the graph prior model from Eq. 2, the Bayesian updating equation 1 is reformulated as:

$$P(\mathbf{A}|\mathbf{C},\mathbf{B},\mathbf{L}) \propto P(\mathbf{C}|\mathbf{A},\mathbf{B},\mathbf{L})P(\mathbf{A}|\mathbf{B},\mathbf{L})P(\mathbf{B},\mathbf{L}) \tag{7}$$

When information on blocks and layers of infrastructure networks, $\mathbf{B}$ and $\mathbf{L}$, is accessible, then $P(\mathbf{B},\mathbf{L}) = 1$. The independence of cascading failure $\mathbf{C}$ from $\mathbf{B},\mathbf{L}$ results in $P(\mathbf{C}|\mathbf{A},\mathbf{B},\mathbf{L}) = P(\mathbf{C}|\mathbf{A})$. So the posterior of the adjacency matrix is further simplified to:

$$P(\mathbf{A}|\mathbf{C},\mathbf{B},\mathbf{L}) \propto P(\mathbf{C}|\mathbf{A})P(\mathbf{A}|\mathbf{B},\mathbf{L}) \tag{8}$$

where the graph prior $P(\mathbf{A}|\mathbf{B},\mathbf{L})$ is calculated using Eq. 2 and the likelihood $P(\mathbf{C}|\mathbf{A})$ is given by Eq. 5-6. Since Eq. 8 usually does not have any closed form solution and the number of candidate graphs grows exponentially with the increase of the network size, Metropolis-Hastings algorithm is leveraged to generate samples of the posterior distribution of interest, which is introduced in the following section.

3. Metropolis-Hastings Algorithm for Network Reconstruction

In the Metropolis-Hastings algorithm, the next sample value is selected from a proposal distribution parametrized by the current sample value and is either accepted or rejected with the accept ratio determined by the value of function of the current and candidate sample values with respect to the desired distribution. As more samples are accepted, the distribution of values becomes a close approximation of the desired distribution. In applying the Metropolis-Hastings algorithm to construct the posterior distribution of network topology, each sample value is a network adjacency matrix and we propose a new adjacency matrix $\mathbf{A}'$ from the previous adjacency matrix $\mathbf{A}$ using proposal distribution $Q(\mathbf{A}'|\mathbf{A})$. The algorithm accepts the shift from $\mathbf{A}$ to $\mathbf{A}'$ with the ratio α, Eq. 9.

$$\alpha = \min\{1, \frac{P(\mathbf{C}|\mathbf{A}')P(\mathbf{A}'|\mathbf{B},\mathbf{L})}{P(\mathbf{C}|\mathbf{A})P(\mathbf{A}|\mathbf{B},\mathbf{L})} \cdot \frac{Q(\mathbf{A}|\mathbf{A}')}{Q(\mathbf{A}'|\mathbf{A})}\} \tag{9}$$

The performance of the algorithm depends on the choice of the proposal distribution $Q(\mathbf{A}'|\mathbf{A})$. A common proposal used when implementing the Metropolis-Hastings algorithm on networks is to change a random node pair either by creating or removing old edges. The main drawback of this proposal lies in the possibility to select invalid network samples. For example in interdependent water and power networks, water delivery stations provide water to power plants for cooling purposes rather than to power substations. However, the traditional proposal may sample a pair of nodes that include a water delivery station and a power substation, and connect them with an edge. Such proposal contradicts the topology and operation of real-world infrastructure networks. In order to propose more realistic networks, we modify the original random sampler method to devise an infrastructure-dependent proposal that imposes additional restrictions on the topology of the proposed infrastructure networks.

4. Infrastructure-Dependent Proposal

Different types of networks have different topology features. The infrastructure-dependent proposal samples networks considering topology constraints of the infrastructure networks. We first analyze the basic structure of the interdependent infrastructure networks, from which we abstract the topology constraints 1-7 used for constructing the infrastructure-dependent proposal.

Interdependent infrastructure networks are comprised of multiple single networks with interdependent links representing their mutual interactions. Each individual network has a hierarchical structure with three levels corresponding to three types of nodes representing supply, transmission and demand facilities. Suppose that we have a set of networks M and a set of their interdependency I, supply, transmission and demand nodes are denoted as

s, t, d. Based on the operations of individual networks and their interdependencies, we determine seven constraints on the topology of the proposed networks.

Typically, resources are generated or extracted from the nature at supply nodes and transported via transmission nodes to demand nodes where resources are further distributed to local residents. Therefore, every supply node $i \in \mathbf{N}_{\mathrm{s}}^m$ is connected by a path to at least one demand node $j \in \mathbf{N}_{\mathrm{d}}^m$, and vice versa, according to constraints 1-2. Every transmission node $i \in \mathbf{N}_{\mathrm{t}}^m$ is connected by a path from at least one supply node $j \in \mathbf{N}_{\mathrm{s}}^m$ and to at least one demand node $k \in \mathbf{N}_{\mathrm{d}}^m$, corresponding to constraints 3-4. Similarly, in interdependent networks $m \in \mathbf{I}$, we have the same constraints for connectivity from the supply nodes to demand nodes, 6-7. The resources move from the supply nodes at the high level to the demand nodes at the bottom level without flowing back, which results in only forward edges and thus no cycles, 5. The infrastructure-dependent proposal is formed by improving the traditional random proposal by considering the seven constraints.

1. $\forall m \in \mathbf{M}, \forall i \in \mathbf{N}_{\mathrm{s}}^m, \exists j \in \mathbf{N}_{\mathrm{d}}^m, i \overset{\text{path}}{\rightarrow} j$
2. $\forall m \in \mathbf{M}, \forall i \in \mathbf{N}_{\mathrm{d}}^m, \exists j \in \mathbf{N}_{\mathrm{s}}^m, j \overset{\text{path}}{\rightarrow} i$
3. $\forall m \in \mathbf{M}, \forall i \in \mathbf{N}_{\mathrm{s}}^m, \exists j \in \mathbf{N}_{\mathrm{t}}^m, i \overset{\text{path}}{\rightarrow} j$
4. $\forall m \in \mathbf{M}, \forall i \in \mathbf{N}_{\mathrm{t}}^m, \exists j \in \mathbf{N}_{\mathrm{s}}^m, j \overset{\text{path}}{\rightarrow} i$
5. $\forall m \in \mathbf{M}$, no cycles
6. $\forall m \in \mathbf{I}, \forall i \in \mathbf{N}_{\mathrm{s}}^m, \exists j \in \mathbf{N}_{\mathrm{d}}^m, i \overset{path}{\rightarrow} j$
7. $\forall m \in \mathbf{I}, \forall i \in \mathbf{N}_{\mathrm{d}}^m, \exists j \in \mathbf{N}_{\mathrm{s}}^m, j \overset{path}{\rightarrow} i$

In the Metropolis-Hastings algorithm, we select a node pair (i, j) at random at each iteration, add or remove the edge between that node pair (i, j) and get a new network as the candidate, the candidate network is accepted if the addition or removal of this edge does not violate the seven constrains. Otherwise, this candidate is rejected, and we propose a new node pair and check the feasibility of the proposed topology. All node pairs are chosen with equal probability, so in Eq. 9, $Q(\mathbf{A}'|\mathbf{A}) = \frac{2}{N(N-1)}$ for all $\mathbf{A}$ and $\mathbf{A}'$ that differ by one edge (Gray et al. 2019). Therefore, the transition of the infrastructure-dependent proposal is symmetric and Eq. 9 is further simplified to:

$$\alpha = \min\{1, \frac{P(\mathbf{C}|\mathbf{A}')P(\mathbf{A}'|\mathbf{B},\mathbf{L})}{P(\mathbf{C}|\mathbf{A})P(\mathbf{A}|\mathbf{B},\mathbf{L})}\} \qquad (10)$$

where the likelihoods $P(\mathbf{C}|\mathbf{A}'), P(\mathbf{C}|\mathbf{A})$ are calculated using Eq. 5-6 and the priors $P(\mathbf{A}'|\mathbf{B},\mathbf{L}), P(\mathbf{A}|\mathbf{B},\mathbf{L})$ are calculated using Eq. 2-3.

5. Experiment Evaluation

In this section, the proposed Bayesian inference approach is applied along with the new infrastructure-dependent proposal to infer the topological structure of a synthetic interdependent power-gas networks. Results demonstrate the capability of our approach in reconstructing the network topology.

5.1. Data

The interdependent power-gas networks are generated by the HSBM where the whole system has two blocks, each corresponding to the power and gas networks. The gas network is simplified from the National Transmission System (NTS) of U.K., the geographical boundary of which spans between (50.01°N, 58.62°N) in latitude and (−5.68°W, 1.66°E) in longitude. The gas network has 9 supply nodes, 29 transmission nodes and 9 demand nodes corresponding to gas storage facilities, compressor and gas terminal, respectively. The network topology is directly obtained fromQadrdan et al. (2010). For the power network, the topology of the IEEE RTS96 24-bus system for a single area is adopted which includes 10 supply nodes, 4 transmission nodes and 10 demand nodes, corresponding to power generators, 12 or 23kV substations and electric load, (Zlotnik et al., 2016)).

Due to security concerns and the use of a synthetic power network, the location of network components are unknown. Therefore, we simulate the node locations by applying the annealing simulation algorithm to optimize the overall distance between local residents and their nearest facilities.

We consider two types of interdependencies for the power and gas networks. First, power demand nodes (electric load facilities) provide electricity to gas supply nodes (gas storage facility) for their regular operation such as extracting natural gas from underground. Second, gas demand nodes (gas terminal) provide natural gas to power supply nodes (power generator) for the purpose of generating electricity. Both types of interdependencies are considered by accounting for the distance between each facility. The synthetic interdependent gas and power networks to be reconstructed is visualized in Fig. 2. We denote the target adjacency matrix, i.e. the adjacency matrix of this synthetic interdependent power-gas networks, that we aim to infer as $\mathbf{A}^*$ hereafter. In applying the Metropolis-Hastings algorithm to estimate the posterior distribution of the adjacency matrix $\mathbf{A}$, an adjacency matrix $\mathbf{A}^0$ is initialized as the starting point to begin the sampling process. To ensure a nonzero value of the likelihood $P(\mathbf{C}|\mathbf{A})$, the initial adjacency matrix $\mathbf{A}^0$ is required to structure the observed cascading scenarios $\mathbf{C}$, i.e. $P(\mathbf{C}|\mathbf{A}^0) > 0$. Therefore, we generate $\mathbf{A}^0$ by connecting nodes that fail at consequential time steps in the cascading scenarios $\mathbf{C}$. The cascading scenarios $\mathbf{C}$ encode the information of the network topology and prompts the MCMC chain to converge towards the adjacency matrix of the synthetic interdependent networks $\mathbf{A}^*$. The data of cascading scenarios $\mathbf{C}$ is simulated using the SI epidemic model where we set some initial failed nodes and perform the failure propagation with the conditional failure probability q in Eq. 6 as 0.4. Setting the conditional failure probability either too high or too low results in either very fast or slow failure propagation, both of which result in fewer failure sequences. In order to obtain enough cascading data to update our knowledge on the network topology, by draft trails, we find $q = 0.4$ is a medium value for generating cascading scenarios with reasonably long failure sequences. To demonstrate the dependency of the performance of reconstructing network topology on the amount of cascading scenarios $\mathbf{C}$, we design three experiments: (i) E_6^{10}: 10 cascading failure scenarios, each of which has at least 6 time steps (ii) E_8^{20}: 20 cascading failure scenarios, each of which has 8 time steps, and (iii) E_{10}^{30}: 30 cascading failure scenarios, each of which has 10 time steps.

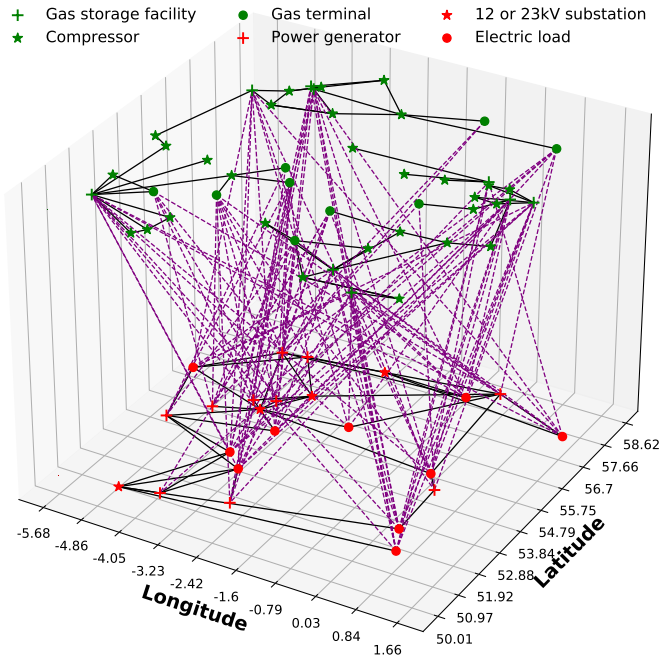

Figure 2: The synthetic interdependent gas-power networks in the U.K. The nodes in green and red represent facilitiesof the gas and power networks. Purple line refer to interdependent links between gas and power nodes

5.2. *Results*

In order to determine an appropriate number of iterations for the convergence of the Metropolis-Hastings algorithm, we calculate the features for each newly proposed network and compare their values with those from the target network to measure how far the current network proposal is from the target network. Since the average degree indicates the connectivity level of the network and is the basic measure in computing many high-level network features such as the clustering coefficient and the diameter, we select the average degree as the representative network feature and draw its trace plot for the three experiments, Fig. 3. The average degree of the reconstructed networks in E_6^{10} and E_8^{20}, the blue and orange chains, converge roughly around 0.042 while in E_{10}^{30}, the green chain, the value converges precisely around 0.04, which is

closer to the average degree of the target interdependent networks represented by the black dashed line. Increasing the number of cascading scenarios with longer time steps in the 3rd experiment embeds abundant node-pair information that covers a wider range of network topology compared to less information from fewer cascading scenarios with shorter time steps in the 1st and 2nd experiments. Due to the lack of data in E_6^{10} and E_8^{20}, the posterior of edges between certain node pairs remains the same as the prior whereas in E_{10}^{30}, the posterior of these edges is updated with data of the corresponding node pairs, drawing the distribution closer to the target posterior. Additionally, the increasing amount of cascading failure data in E_{10}^{30} incorporates more knowledge of the network topology into the prior and after each iteration, driving the the posterior closer to objective. As a result, compared to the first two experiments, the value of the average degree converges to the target with fewer iterations in E_{10}^{30}, i.e., less iteration numbers in the warm-up state. To evaluate the performance of the network

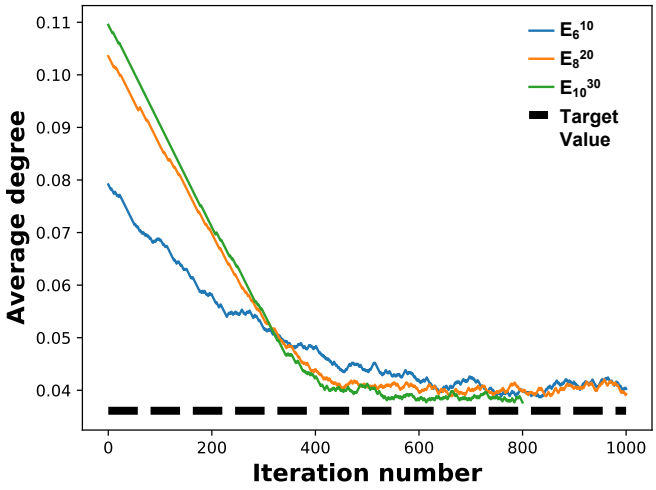

Figure 3: Trace plot of the average degree under three experiments

reconstruction, we further convert it into a problem of predicting network adjacency matrix. We filter out the network samples in the warm-up stage and obtain the posterior distribution of the network topology. Suppose that the set of networks in the posterior distribution is $\mathscr{A}$, the edge probability p_{ij} for each pair of nodes (i, j) is calculated by counting the number of graphs with the edge between i, j and dividing it by the total number of graphs in the posterior, Eq. 11.

$$p_{ij} = \frac{\sum_{\mathbf{A}\in\mathscr{A}} I\{\mathbf{A}_{ij} = 1\}}{|\mathscr{A}|} \tag{11}$$

In Eq. 11, $I\{\mathbf{A}_{ij}\}$ is the binary term, which equals to 1 when $\mathbf{A}_{ij} = 1$ and 0 when $\mathbf{A}_{ij} = 0$.

We plot the edge probability for each pair of nodes in the network and display them on the heatmap of the adjacency matrix shown in Fig. 4. The top-left corner of the figure shows the original networks where the edges are densely distributed in certain areas with other blank areas showing no connections, which clearly reflects the hierarchical block structure as described by the HSBM model. The constraints of the infrastructure-dependent proposal ensures that in all three experiments, the adjacency heatmaps conform to the block structure in the original network. Since E_{10}^{30} uses more cascading failure data to update the network topology, we collect more information about the network topology and thus have higher confidence in the predicted edges. Therefore, the heatmap of the adjacency matrix is darker in E_{10}^{30} than in the other two experiments, indicating higher edge probability.

Based on the adjacency heatmap, we set up a probability threshold p and classify node pairs into two categories: node pair connected by an edge if $p_{ij} \geq p$ and node pair not connected if $p_{ij} < p$. We further count the number of edges for each pair of nodes that are correctly classified and plot the precision-recall curve in Fig. 5. The best F1-score for each of the three experiments is 0.42, 0.82 and 0.85, respectively, which is consistent with the previous observation that as the amount of cascading data increases, the performance of the reconstruction improves.

6. Conclusion and Future Work

This paper presents a Bayesian estimation approach for reconstructing the posterior distribution of the topology of interdependent infrastructure networks based on data of cascading failures. The HSBM model is used to capture the clustering and hierarchical structure of the interdependent infrastructure networks. The intractability caused by

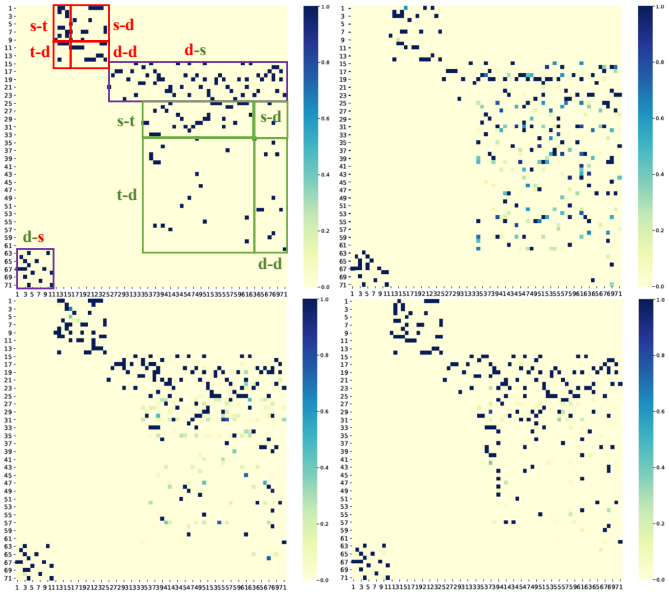

Figure 4: The heatmap of the adjacency matrix where both of the x and y axes represent the node number. Top-left: the original networks with hierarchical block structure (s-t: the supply to transmission edges, s-d: the supply to demand edges, t-d: the transmission to demand edges, d-d: the demand to demand edges, green color represents nodes in the gas network and the red color represents nodes in the power network, the cross-network edges are represented by d-s with different colors); top-right: E_6^{10}; bottom-left: E_8^{20}; bottom-right: E_{10}^{30}.

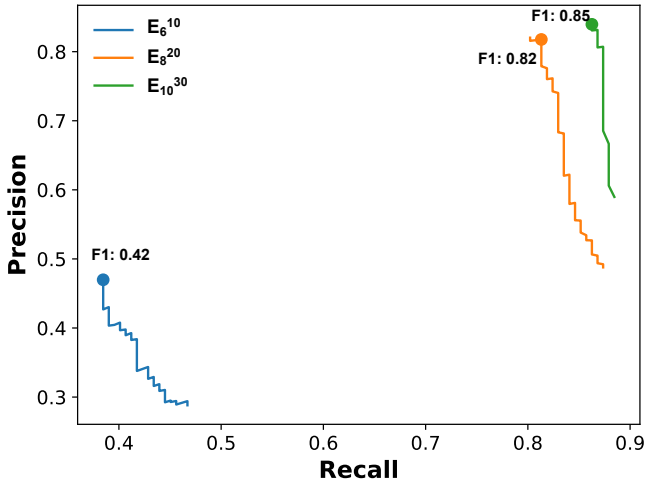

Figure 5: The precision-recall curves for the three experiments

the exponential growth of the candidate graphs is solved by a new infrastructure-dependent proposal where we improve the traditional random proposal by adding seven topological constraints of the inter-dependent infrastructure networks.

Future work will explore the expansion of the proposed reconstruction approach by incorporating node features that relax the assumption of constant propagation probability of failure across all pairs of nodes.

7. REFERENCES

Amini, A., Paez, M., and Lin, L. (2022). "Hierarchical stochastic block model for community detection in multiplex networks." *Bayesian Analysis*, 1(1), 1–27.

Erdős, P., Rényi, A., et al. (1960). "On the evolution of random graphs." *Publ. Math. Inst. Hung. Acad. Sci*, 5(1), 17–60.

Fu, G., Wilkinson, S., and Dawson, R. J. (2016). "A spatial network model for civil infrastructure system development." *Computer-Aided Civil and Infrastructure Engineering*, 31(9), 661–680.

Kolaczyk, E. D. (2017). *Topics at the Frontier of Statistics and Network Analysis:(re) visiting the Foundations*. Cambridge University Press.

Ouyang, M., Dueñas-Osorio, L., and Min, X. (2012). "A three-stage resilience analysis framework for urban infrastructure systems." *Structural safety*, 36, 23–31.

Peixoto, T. P. (2019). "Network reconstruction and community detection from dynamics." *Physical review letters*, 123(12), 128301.

Qadrdan, M., Chaudry, M., Wu, J., Jenkins, N., and Ekanayake, J. (2010). "Impact of a large penetration of wind generation on the gb gas network." *Energy Policy*, 38(10), 5684–5695.

Zlotnik, A., Roald, L., Backhaus, S., Chertkov, M., and Andersson, G. (2016). "Coordinated scheduling for interdependent electric power and natural gas infrastructures." *IEEE Transactions on Power Systems*, 32(1), 600–610.