Analysis of Frequency Dynamics in Power Grid: A Bayesian Structure Learning Approach

Hannan Ma and Husheng Li

***Abstract*—The oscillation of frequency in power grid is studied in this paper. The possibility association of frequencies measured at different locations are modeled by a Bayesian network with the *logical structure* learned using Bayesian structure learning and real measurements in the U.S. power grid. Frequency data analysis and the detection of incorrect frequency measurements (caused by equipment error or malicious attack) are performed over the logical Bayesian network structure. Such application of Bayesian network is a powerful mathematical tool in computational intelligence. Without the physical power network topology information, a two-branch search-and-score structure learning algorithm with L-1 regulation is proposed to learn the logical structure, achieving around 97% correct prediction rate for future frequency and 92% detection rate for false frequency data with 2% false alarm rate. The tool of epidemic propagation over this logical network is also exploited to analyze the propagation of frequency changes. Using the Kolmogorov-Smirnov test, such logical structure is demonstrated to be well approximated by the Small World network model. And the propagation of frequency changes is demonstrated to be described by the Susceptible-Infectious-Susceptible (SIS) model quite well. The Bayesian structure obtained from the real measurement is statistically validated using a 5-fold training data and the Pearson system.**

***Index Terms*—Artificial intelligence, dynamics, frequency estimation, inference mechanisms, security.**

I. Introduction

MODERN electricity power infrastructure is vulnerable against many forms of natural or malicious physical events nowadays [1]. It is required in the U.S. power grid that the frequency deviation from the standard 60 Hz be within 0.2 Hz. A significant frequency deviation may cause system instability or may damage electronic devices. As a consequence, the dynamics of frequency become one of the most important metrics in power grids [8]. Hence it becomes the key task to analysis the frequency dynamics and to understand the propagation of frequency oscillation.

System dynamics could be fully modeled using Swing Equation [2], but it requires physical topology information and detailed system parameter which are sometimes unavailable to the public. Furthermore, such methods are too complicated to gain insights in large scale power network. Most of past works are based on complicated modeling of physical components in power system. To name a few, [9] used nonlinear system theory to predict the periodic fluctuation of power system voltage flickers. In [10], [11], statistical methods were applied for long time frequency change rate estimation and dynamic stability analysis using Kalman filter.

In a sharp contrast, our study is a data driven method and uses a mathematical phenomenological model. *Instead of focusing on the physical topology of power system, in this paper we focus on how to obtain and apply a logical network structure. We adopt a data driven method and apply the Bayesian network, a powerful mathematical model in computational intelligence and machine learning [5]. Using the structure learning algorithm with real frequency measurements we obtain a logical structure for the frequency dynamics. Such Bayesian network structure is then used to predict the future frequency and detect possible false data. We also apply the epidemic propagation theory in social networks [3], [4], to show this logical structure could model the spread of significant frequency changes over the power network.*

In the last decade, Bayesian network has been applied in fault diagnosis and reliability assessment of power grids, e.g., in [12], [13]. Different from them, we focused on learning the Bayesian network for a large number of power network nodes and understanding how such structure is related to the power network dynamics. There are also studies using complex networks in power systems. Critical nodes in complex networks are analyzed to study the vulnerability, e.g., in [14], [15]. Although the topology also plays an important role in our model, we are more focused on analyzing the dynamics over the learnt topology. [16] presented an excellent study on dynamics of cascading failures in power network using topology analysis and a method similar to structure learning. The difference of our paper is that we are more data driven and more focused on frequency dynamics with epidemic model instead of cascading link failure. Note that [17] analyzing human brain MRI images had a similar procedure of this paper, but with less discussion on structure learning, prediction and further application. Although Bayesian network is related to soft computing, artificial intelligence and data mining, the application of Bayesian network is more of a computational intelligence case. The application of Bayesian network in neuroscience and bioinformatics have long been enlisted into the scope of the Journal of Computational Intelligence in Bioinformatics, the Computational Intelligence and Neuroscience published by *Hindawi*, and the book *Innovations in Bayesian Netowrks: Theory and Applications* is one of *Springer*'s "studies in Computational Intelligence" books.

Using frequency measurements in the Frequency Monitoring Network (FNET) [6] with Frequency Disturbance Recorders (FDRs) shown in Fig. 1, we obtained the logic structure using

Manuscript received August 31, 2011; revised December 31, 2011, July 31, 2012; accepted October 14, 2012. Date of publication February 06, 2013; date of current version February 27, 2013. Paper no. TSG-00472-2011.

H. Ma is with the Department of Electrical Engineering and Computer Science, the University of Tennessee, Knoxville, TN 37996 USA (e-mail: hma5@utk.edu).

H. Li is with the Department of Electrical Engineering and Computer Science, the University of Tennessee, Knoxville, TN 37996 USA and is also an International Scholar of Kyung Hee University, Korea (e-mail: husheng@eecs.utk.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSG.2012.2226066

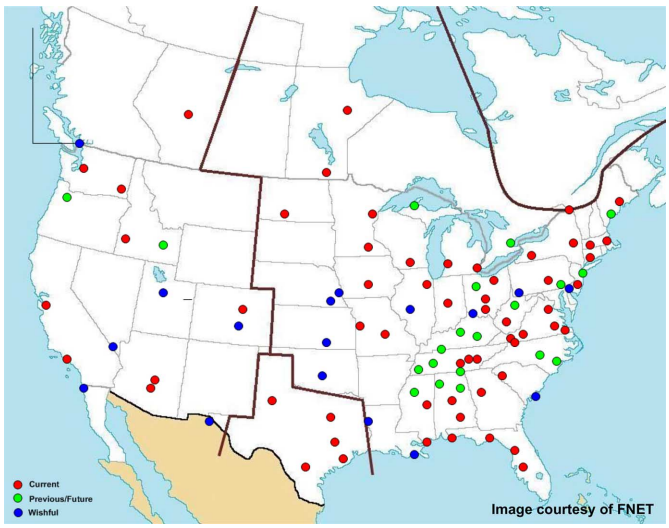


Fig. 1. The locations of FDRs in North America.

Bayesian Structure Learning (BSL) [5], with each node represents the frequency at a FDR and each edge models the possibility association between two FDRs. The K-2 algorithm [7] is a popular BSL algorithm which uses a predetermined order for nodes and finds directed edges in a sequential manner. We will show that the K-2 algorithm tends to generate many unnecessary edges in modeling the power network, resulting in *over learning* [5]. Hence, we propose an iterative BSL algorithm with an L-1 regulation to remove the over-learned edges, which attains a significant performance gain and does not require a predetermined order. The resulting structure is validated using 5-folds in training data.

Based on the Bayesian network structure learned from the data, we carried out the following two tasks for analyzing the frequency dynamics in power grids:

- The short-term future frequencies are predicted based on current measurements and the Bayesian network structure. The correct rate of prediction is shown to be around 97%. We also tested the performance of detecting false data caused by sensor malfunction or intentional attack using the Bayesian network structure. The detection rate is shown to be 92% with a 2% false alarm rate.
- We model the propagation of frequency fluctuation as the spread of epidemic disease and show the logical structure describes the frequency dynamics. The Kolmogorov-Smirnov (K-S) test shows that this Bayesian structure can be approximated by a Small World network model [3]. Then we describe the frequency change propagation using Susceptible-Infectious-Susceptible (SIS) model [4] in a Small World network. The validity of the model is well demonstrated using numerical results showing that both the SIS model and the logical network are consistent with the frequency measurement data.

In summary, our main contribution is using Bayesian network structure learning to obtain a representation of the logical structure of frequency dynamics in power grid, which is used to predict future frequency, detect attacks, and model frequency dynamics.

The remainder of this paper is organized as follows. A brief introduction to Bayesian network is given in Section II. The proposed learning algorithm is then explained in Section III (a brief introduction to conventional K-2 structure learning algorithm is in Appendix). Then the Bayesian network structure learned from the proposed algorithm and real measurements is in Section IV. The structure is then applied for frequency prediction and detection in Sections V, VI. The application in epidemic propagation modeling is in Section VII. The conclusions are drawn in Section VIII.

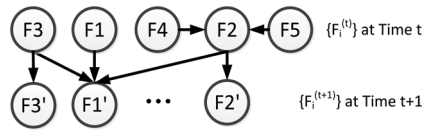


Fig. 2. A Bayesian network to model frequency dynamics in power grid.

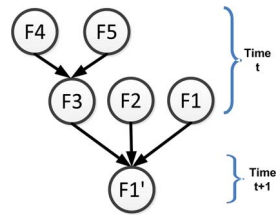


Fig. 3. The Bayesian network to infer the state of node F1.

II. Introduction to Bayesian Network

We use Bayesian network to model the frequency dynamics in power grid in both spacial and time domains. Details about Bayesian network could be found in [5].

For N locations in the power grid, the frequency measurement of each FDR at one time slot is modeled as a random variable and represented by a node F_i in Bayesian network. The possibility associations among frequency measurements at different sampling times at different locations are represented by directed edges between nodes. For example, the Bayesian network shown in Fig. 2 is a directed acyclic graph consisting of 10 nodes (with $N = 5$ over two different time slots) and directed edges. Focusing on F_1^{t+1} that represents the frequency for node F_1 at time $t+1$, we denote by $Pa_1^t = \{F_1^t, F_2^t, F_3^t\}$ its parent nodes that directly point to F_1^{t+1}. A subgraph of Fig. 2 shown in Fig. 3 could be used to infer the frequency value of F_1^{t+1}. Fig. 3 shows that F_1^{t+1} is dependent on its parents and is conditionally independent of non-parent nodes F_4^t and F_5^t. Mathematically, given the values of the parents, we have

$$Pr\left(F_1^{t+1} | \left\{F_i^t\right\}_{i=1}^5\right) = Pr\left(F_1^{t+1} | Pa_1^t\right). \quad (1)$$

Furthermore, the joint distribution could be factorized into the product of conditional probability distributions (CPDs) which are the parameters of the Bayesian network, i.e.,

$$Pr\left(\left\{F_1^{t+1}\right\} \cup \left\{F_i^t\right\}_{i=1}^5\right) = Pr\left(F_1^t\right) \cdot Pr\left(F_2^t\right) \cdot Pr\left(F_4^t\right) \cdot Pr\left(F_5^t\right) \cdot Pr\left(F_1^{t+1} | F_1^t, F_2^t, F_3^t\right) \cdot Pr\left(F_3^t | F_4^t, F_5^t\right). \quad (2)$$

With such a Bayesian network structure and the corresponding CPDs, we can accomplish the following tasks:

- Frequency Prediction: With frequency measurements at time t as Pa_1^t, probabilities of future frequencies at $t+1$ F_1^{t+1} could be inferred by decision rule:

$$\hat{F}_1^{t+1} = \arg\max_f Pr\left(F_1^{t+1} = f | F_1^t, F_2^t, F_3^t\right). \quad (3)$$

- False Data Detection: the frequency prediction $\hat{F}_1^{t+1}$ is of high accuracy. If there is a large gap between the received measurement F_1^{t+1} and the prediction $\hat{F}_1^{t+1}$, we can claim that there is probably something wrong in the received frequency report.
- Modeling the Frequency Dynamics: If considering the propagation of frequency fluctuations as epidemic spreading over a social network, then we prove such social network is the logical Bayesian network structure. This gives insights on frequency dynamics in power grids.

These applications will be studied in Sections V, VI, and VII.

III. Two-Branch Iterative Structure Learning Algorithm

In this section we propose a novel structure learning algorithm that iteratively searches in two directions among all the nodes to look for the best parents. A brief introduction to the conventional K-2 algorithm is given in Appendix I for self-containedness. Readers unfamiliar with the K-2 algorithm can read the Appendix to understand the basic principles of the algorithm, which helps to understand the discussion in this section. Readers that are only interested in the applications and conclusions in power grid can skip this section.

The new algorithm has the following features:

- The algorithm adopts random ordered inputs. Hence we do not need to find an optimal order or bear the risk caused by sup-optimal ordering, which is also discussed in Appendix I.
- A two-branch iterative algorithm is proposed to enlarge the candidate parents searching space, with mutual information is used to optimize the searching space in order to control the added computational complexity. In this balanced way more nodes that are highly probable to be parents are included in the searching space to describe the cause-effect possibility associations.
- An L-1 regulation [23] is applied to constrain on the structure complexity and to eliminate the possibility of over learning the structure.

We expand the searching space of potential parents for F_i to all other nodes for a random input order, which is given by

$$\hat{P}a_i = \{F_j\}_{j=1}^{i-1} \cup \{F_j\}_{j=i+1}^{N} \triangleq \hat{P}a_i^- \cup \hat{P}a_i^+. \tag{4}$$

This enlarges the searching space but increases computational complexity. To optimize it and reduce complexity, candidate nodes in $\hat{P}a_i^-$ and $\hat{P}a_i^+$ are sorted by their mutual information based on the Minimum Description Length principle [18]:

$$I(F_i; F_j) = \sum_{f_i} Pr(F_i = f_i) \log \frac{P(F_i|F_j)}{P(F_i)}, \tag{5}$$

where $F_j \in \hat{P}a_i^-$ or $\hat{P}a_i^+$ and all the terms in (5) are needed to compute the Cooper-Herskovits (CH) score function (13) (the definition of the CH score function can be found in Appendix I); therefore the only extra thing to do is computing (5) and sorting. A threshold ε_I is set manually so that all nodes have at least *MaxParent* number of candidate parents: nodes F_i and F_j are claimed to be independent if $I(F_i; F_j) < \varepsilon_I$, and independent nodes are excluded from $\hat{P}a_i$.

For each F_i, the algorithm scans each $F_j \in \hat{P}a_i$. One more parent is added to Pa_i after each scan until adding any more edges cannot increase the score or the maximum number of parents $maxParents$ has been reached. If any cycle is formed, we break the edge that produces the least score. If no more edges to add or no cycle detected, the algorithm gives the structure dag which is the connection matrix $dag = \{d_{ij} | d_{ij} = 1$ if having edge $F_i \rightarrow F_j\}$.

The proposed algorithm conducts a two-way scan in input nodes, namely a forward branch scanning from F_1 to F_N and a backward branch scanning backwards. As shown in Fig. 4, after the random order is generated by an interleaver, in the forward branch after scanning node F_i we have dag_i^f with structure information among nodes in $\hat{P}a_i^-$ known and connections for nodes in $\hat{P}a_i^+$ remain unknown. At the same time, in the backward branch, $F_{j=N-i+1}$ is scanned yielding dag_j^b with structure information for nodes $\{F_k\}_{k=j+1}^{N} \subseteq \hat{P}a_i^+$, which is useful for the forwarding branch. Meanwhile, information in dag_i^f regarding nodes in $\hat{P}a_i^- \subseteq \hat{P}a_j$ is useful for the backward branch. Hence both branches could share structure information after scanning each node and merge the structure information from both branches by Boolean algebra conjunction or binary logic *and*: $dag^{(ITE)} = dag_i^f \oplus dag_j^b$. The pseudo code is in Algorithm 1 with all symbols defined in previous and in Appendix I. This iterative algorithm learns the correct Bayesian network for ALARM data-set-3 [7], a benchmark testing problem for BSL algorithms, consisting 37 nodes and 46 edges. It also yields an average of 8.39% gain in the total score over the conventional K-2 algorithm, which implies more accuracy in the structure could be achieved by the proposed algorithm.

Algorithm 1 Pseudo Code for Two-Branch Iterative Algorithm

Input: Randomly ordered Nodes V, training data D
Output: network structure Dag
for Iteration $\leftarrow$ ITE **do**
  for all nodes $F_i \in V$, forward branch **do**
    initialization Parent set $Pa_i \leftarrow \phi$;
    while Pa_i got updated **do**
      initialization $PaCount_i \leftarrow 0$;
      while $PaCount_i < MaxParent$ **do**
        for all candidate parent nodes $F_j \in \widehat{Pa_i}$ **do**
          temporary parents $\hat{P}a_{ij} \leftarrow Pa_i \cup F_j$;
          get structure $G_j = \{V_j, E_j\}$ where
          $V_j \leftarrow \hat{P}a_{ij} \cup F_j$, $E_j \leftarrow dag^{ITE}$;
          score $g(F_i, \hat{P}a_{ij}^{G_j})$ on G_j;
        end for
        if The score is enhanced **then**
          Accept new parents and update parent Pa_i
          $F_k \leftarrow \arg\max_k g(F_i, \hat{P}a_{ik}^{G_k})$;
          $PaCount_i \leftarrow PaCount_i + 1$;
        end if
      end while
      Break cycles if any;
      Do similar step in the backward branch, merge dag
    end while
  end for
    Decide whether the iteration should be terminated;
end for

L-1 Regulation has been applied in structure learning [23]–[25] in order to control the size and degree of the obtained

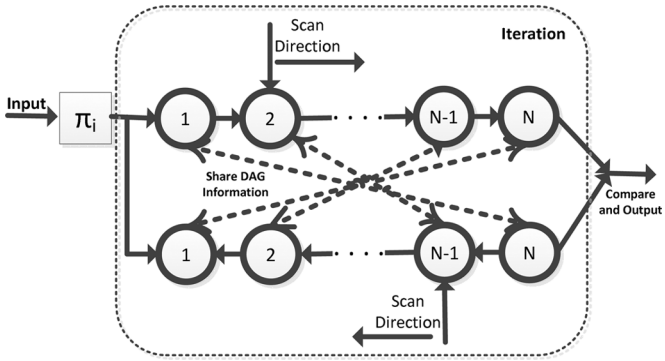


Fig. 4. Structure for the iterative two-stage algorithm.

structure and to resolve the over learning problem. In the proposed two-branch algorithm, the local structure learning on F_i is to find Pa_i to maximize the L-1 regulated CH function:

$$g(F_i, Pa_i) = \arg\max_{Pa_i} \left(\log CH(F_i, Pa_i) + N_{\text{params}}\right), \tag{6}$$

where $N_{\text{params}} = \sum_{k=1}^{M_i} N_{ijk}$ in (13), both defined in Appendix I. Here we use logarithm to make the final score to be sum of all local scores; and we apply numerical algorithm in [25] to solve this optimization problem.

IV. Bayesian Structure Learning Results

The proposed two-branch BSL algorithm with L-1 regulation is applied on real frequency measurements. We compare the logical structure learned by three different BSL algorithms: the conventional K-2 algorithm that runs 500 random input orders and choose the structure with highest score; the iterative BSL algorithm without using the L-1 regulation, which is essentially the two-branch iterative version of K-2; and the iterative BSL with L-1 regulation. We demonstrate that the iterative BSL with L-1 regulation achieves the best performance. The resulting structure will be applied in the next two sections for analyzing the frequencies in power grids. FNET provides real time frequency measurements, collected from $N = 89$ FDRs located across North America around the same time every day averaged over one week and quantized into 7 levels with edges {59.9 Hz, 59.95 Hz, 59.99 Hz, 60.01 Hz, 60.05 Hz, 60.1 Hz}. The sampling interval is 100 ms. The Bayesian network is trained using 8000 training samples with $MaxParent = 10$ and 5-fold validation. Another 1000 samples collected after training data are then used to test prediction performance.

The structures of conventional K-2 algorithm failed the cross validation. Fig. 5 has the highest score among 500 random orders. Compared with Fig. 7 we observe it lacks several important edges. Compared with FDR map Fig. 1, the node in Manitoba Canada is wrongly isolated.

The structure of iterative K-2 algorithm without L-1 regulation is in Fig. 6. Compared with the L-1 regulated version in Fig. 7, it is obvious that without constraints on structure complexity the greedy nature of the K-2 algorithm over learned the structure. With L-1 regulation, we can observe from the structure in Fig. 7 that the western and eastern America power grid components are roughly separated. The structure is overlaid onto the map of North America shown in Fig. 8. According

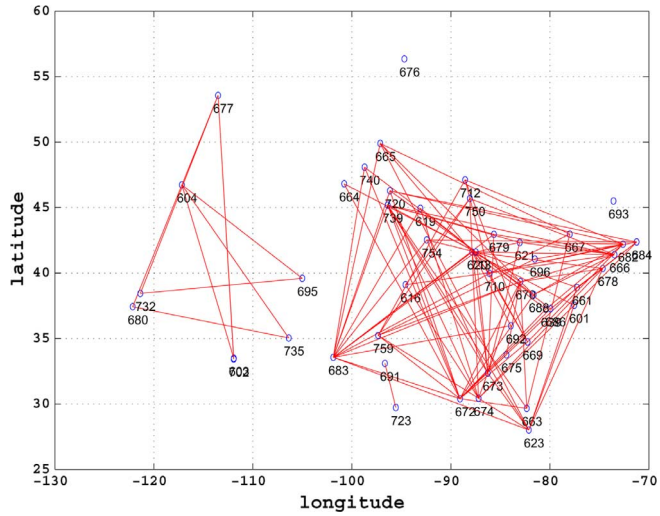


Fig. 5. Structure learned from conventional K-2 algorithm.

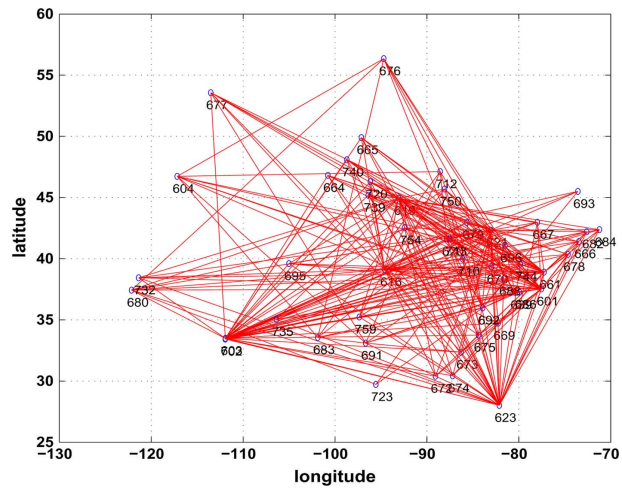


Fig. 6. Structure learned from iterative version of K-2 algorithm without L-1 regulation.

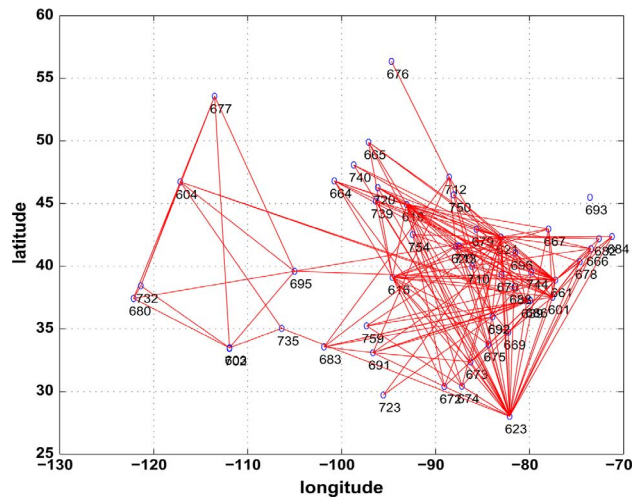


Fig. 7. Structure learned from L_1 regulation algorithm.

to the dynamic clustered interconnection units from FNET [6], the eastern and western U.S. interconnection units are clustered and separated. FDR#693 in Quebec, Canada, is in Quebec unit according to FNET and is isolated in the structure. FDR#676

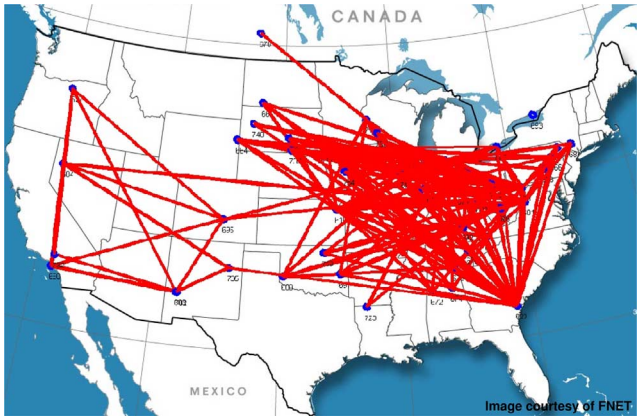


Fig. 8. Structure learned by the two-stage algorithm with the L-1 regulation overlaid with the map of North America.

TABLE I
COMPARISON OF DEGREE FOR STRUCTURE COMPLEXITY

FDR ID	L-1 Regulation	Iterative without L-1	Conventional K-2
AvgDegree	4.72	8.48	6.44
616	7	11	17
693	0	0	4
732	1	5	7

around Manitoba, Canada, is included in the eastern interconnection units according to FNET. The structure successfully recovered these relationships. This consistency between the structure learning and the power grid topology demonstrates that the structure learned with the proposed iterative two-branch algorithm with $L-1$ regulation is more reliable.

The degrees for selected typical nodes with high, low and average degrees are listed in Table I for comparison. We observe that the L-1 regulation removes many unnecessary edges in the learned structure.

V. APPLICATION 1: FREQUENCY PREDICTION

We apply inference over the learned logical structure to predict the short-term future frequencies, using Belief Propagation [5] and decision rule (3). With the Bayesian network similar to Fig. 2 showing both spatial and temporal relationships, given the current measurement at time t, we predict the quantized frequency level for the next measurement at time $t+1$ by computing $P(F_i^{t+1} = f | F_j^t, j = 1, \ldots, N)$ and choose the one yielding such highest probability as the prediction $\hat{F}_i^{t+1}$. We define F_i is unpredictable if $\max_f \{P(F_i^{t+1} = f | F_j^t, j = 1, \ldots, N)\}$ is less than 0.6 since the belief on the value of F_i^{t+1} is vague.

Simulation shows the structure obtained from the conventional K2 achieves an Average Prediction Error (APE), which is the frequency to approximate the probability of inference error $P(\hat{F}_i^{t+1} \neq F_i^{t+1} | F_i^{t+1} \text{is predictable})$, is 5.428% while our proposed learning algorithm with the L-1 regulation achieves an APE of 3.280%. It is reliable to predict frequency in the future 3 minutes given past 5 minutes' observation This means that we can predict the frequency in the near future (in the order of hundreds of milliseconds in our experiment) with high accuracy. Our proposed learning algorithm achieves a 39% relative performance gain over the traditional learning algorithm. The APE and Predictable Rate ($\text{PR} = \frac{(NumPredictable)}{(N)}$) for typical nodes with low or high degree centrality are listed in Tables II and III respectively, with sensitivity listed in parenthesis. Sensitivity is computed in the same manner as [17], [20] as the worst case error rate for all other possible parents value on each node, and high sensitivity reflects more robustness and reliability of the logical structure. We observe that our proposed learning algorithm outperforms the traditional ones in APE, PR and sensitivity, which further demonstrates its validity.

TABLE II
COMPARISON: PREDICTION RATE AND SENSITIVITY

FDR ID	Conventional K-2	Iterative without L-1	L-1 Regulation
664	96.85% (65.3%)	93.47% (56.2%)	96.69% (70.4%)
665	96.40% (73.2%)	96.45% (70.3%)	97.26% (75.8%)
672	97.75% (86.3%)	98.10% (85.9%)	100.0% (90.5%)
675	95.59% (80.1%)	98.10% (87.9%)	98.12% (88.7%)
682	97.94% (70.2%)	98.15% (74.7%)	98.24% (75.0%)
686	98.16% (100%)	99.35% (99.0%)	100.0% (100%)

TABLE III
COMPARISON: FREQUENCY PREDICTABLE RATE

FDR ID	Conventional K-2	Iterative without L-1	L-1 Regulation
664	88.50%	92.70%	98.40%
665	96.40%	98.50%	100%
672	97.70%	100%	100%
675	89.50%	90.80%	97.10%
682	97.90%	100%	100%
686	98.10%	99.20%	100%

VI. APPLICATION 2: FALSE DATA DETECTION

The resulting Bayesian network structure also demonstrates that the frequencies at different locations of the power grid are statistically correlated. This information redundancy can be applied to detect mistakes in the frequency reports, similarly to the error detection in channel coding. In particular it can be used to detect possible false data attack that modifies the values of the frequency measurement reports.

We assume that some FDRs are under attack and are sending fake frequency data which is uniformly drawn from the 7 quantization levels.[1] This also happens when the FDR is malfunctioning. Given N FDRs, we do not know which nodes are reliable and which are under attack. Our strategy for detecting possible attacks is to use all past observations (containing unidentified error measurements) as evidence, and predict the measurements of other FDRs using decision rule (3). We predict the frequency of one FDR and then compare it with the report. If the FDR is predictable and a large gap is detected, we claim that the report is unreliable and the FDR is under attack or is malfunctioning. Case of detection is when the prediction equals real frequency for malfunctioned FDR, and false alarm is the case when prediction is wrong by a large gap for normal functioned FDR.

In our experiment for $N = 89$ nodes, we randomly set 5% of them under attack. We define a large gap between the prediction (P) and the report (R) using relative Euclidean distance with threshold γ_{th}: $|P - R| > (P + R)/(2) \cdot \gamma_{\text{th}}\%$.

The performance of the false data detection can be characterized by Receiver Operation Characteristic (ROC) curves [30] in which both the detection and false alarm rates are plotted for

[1]There could be many other types of attack. We consider only one for simplicity.

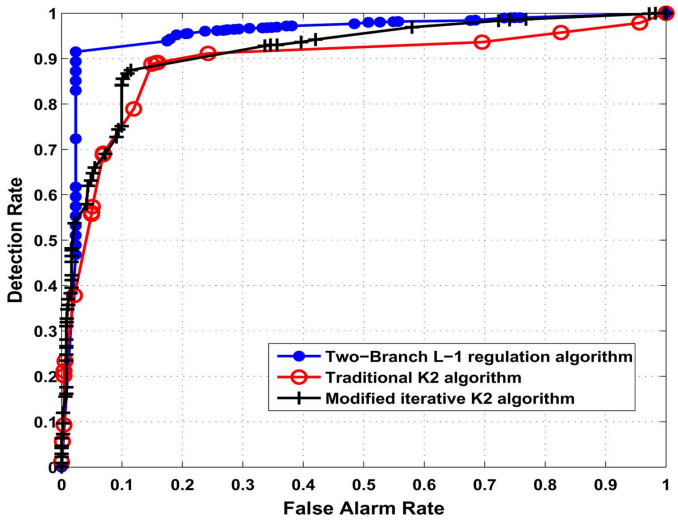


Fig. 9. ROC curve of frequency attack detection.

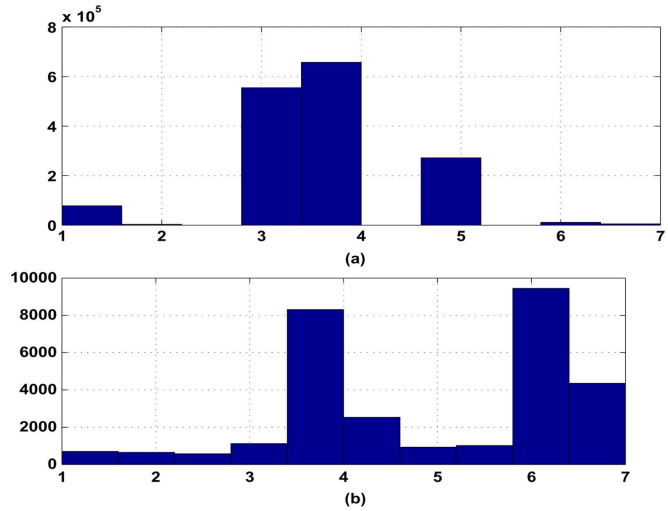


Fig. 10. Frequency measurement distributions.

different detection threshold $\gamma_{\rm th}$, as shown in Fig. 9. We observe that the proposed learning algorithm with the L-1 regulation has the best performance among all three algorithms. Given threshold set to $(1)/(7) \approx 15\%$, the proposed algorithm can achieve a very good detection rate of false data 92% with a reasonable false alarm rate 2%.

VII. Application 3: Epidemic Propagation Model

In this section we will model the frequency fluctuation as an epidemic propagation over the network structure learned from real frequency measurements. In particular, we choose the SIS epidemic propagation model [3] and use the following two approaches to show that the logical structure models the frequency dynamic well using the SIS model:

- Direct model fitting: We optimize the parameter and directly fit the SIS state transition model and its differential equation to the frequency data;
- Network structure oriented fitting: Starting from the logical Bayesian structure, we derive the frequency dynamics.

These two approaches will be studied in the following two subsections.

A. Direct Model Fitting

To study the frequency oscillation, we use a segment of measurements with a sudden change of frequencies in the FNET, different from the training or the testing data used previously in learning. One minute observation of 33 FDRs located in the eastern America during a generator trip event are plotted in Fig. 11 with frequency against observation time, where the time unit is 100 ms. Same as before, a logical structure is learned from 300 data samples with sampling rate 200 ms during this one minute time with two-fold validation, and the rest 300 samples are used to test prediction rate. In this way we achieve an average correct rate of 92.67% with average sensitivity 77.3%. Note that it is not applaudable to use the first 300 sample to train as these training samples do not represent the whole frequency dynamic behavior. Fig. 10 shows the frequency reading distributions of previous 10 k normal working data in (a) and generator trip data in (b). This shows that the prediction performance is good for both low and high frequency deviation. SIS model defines two population types: *susceptible people* (S) defined as people that could be infected by the disease, and *infected people* (I) defined as people that are infected by the disease. Susceptible people contact their friends in the social network with a contact rate β and could become sick; meanwhile infected people recover with a rate γ and become susceptible again.

Eventually all frequencies become higher. Thus using a frequency threshold $f_{\rm th} = 60.00$ Hz we define two states for each FDR node, $F_i \in S$ as susceptible nodes if its frequency $f_i \leq f_{\rm th}$ and $F_i \in I$ as infected nodes if $f_i > f_{\rm th}$. The SIS model could therefore be described using the differential equations.[2]:

$$\frac{ds(t)}{dt} = -\beta i(t)s(t) + \gamma i(t), \tag{7}$$

$$\frac{di(t)}{dt} = \beta i(t)s(t) - \gamma i(t), \tag{8}$$

where s and i are the fractions of susceptible and infected nodes, defined as $s(t) = (S(t))/(N)$ and $i(t) = (I(t))/(N)$ with $S(t) + I(t) = N$.

Given the data in Fig. 11, parameters γ and β are optimized in the sense of minimizing the mean square error of $i(t)$ between the one from the real data and the one generated by SIS model. The best parameters are given by $(\beta = 0.042, \gamma = 0.252)$. Fig. 12 shows that the SIS model in (7), (8) approximate the FNET frequency data quite well.

B. SIS Model: Network Structure Oriented Fitting

In this subsection, we only consider the skeleton of previously obtained Bayesian network (an undirected graph with the same connection). In the previous logical structure for these $N = 33$ FDR nodes, there are $M = 207$ edges. We consider the degree of an arbitrary node as a random variable k. Then we has M degree samples $\mathcal{K} = \{k_i\}_{i=1}^{M}$ and could focus on the degree distribution.

1) Small World and Scale Free Networks: We consider two popular models for complex networks: the Small World network

[2]For simplicity, here we use the simple model for homogeneous networks in the analysis, similar to [31], [32] Actually the Bayesian network structure obtained in the paper is heterogeneous. A rigorous modeling for heterogeneous networks can be found in [4], [3], which will be adopted in our future work. Anyway, our numerical results show that the simple homogeneous network can also well approximate the real dynamics in power grids.

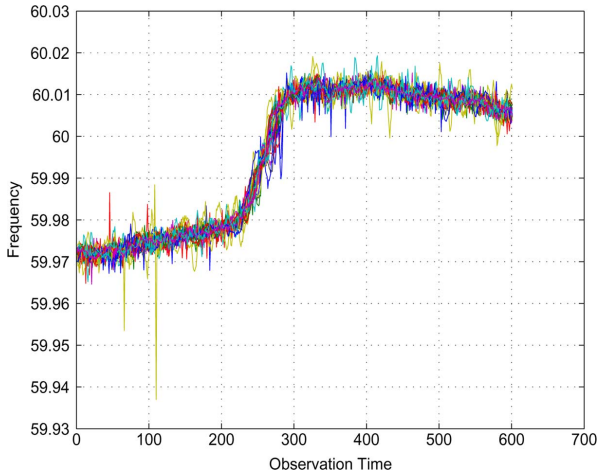


Fig. 11. A sudden frequency change in the U.S. power network.

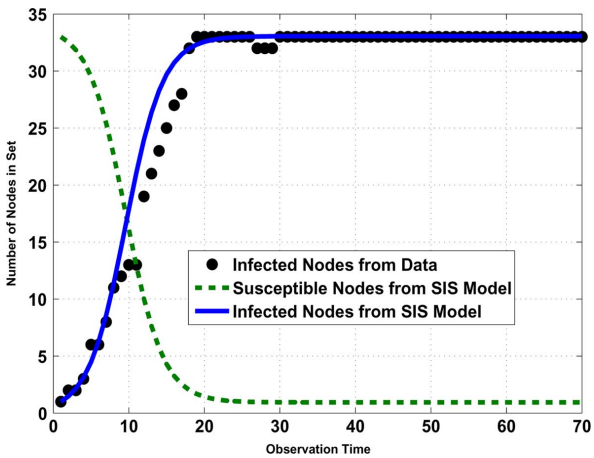


Fig. 12. Optimized SIS model data by direct model fit compared with the FNET frequency data.

model [26] and Scale Free network [3], [27], which are distinguished by their degree distributions:

- The Small World model's degree has discrete poisson distribution with probability mass function (pmf) as:

$$p_\lambda(k) = e^\lambda \frac{\lambda^k}{k!}, \tag{9}$$

with parameter λ and cumulative distribution function (CDF) $\mathcal{F}_\lambda(k)$. Given degree data $\mathcal{K} = \{k_i\}_{i=1}^M$ it is easy to verify the Maximum Likelihood Estimation (MLE) of parameter λ is $\hat{\lambda}$:

$$\hat{\lambda} = \frac{1}{M}\sum_{i=1}^{M} k_i, \tag{10}$$

where M is number of degree samples.

- The Scale Free model's degree has discrete power law distribution [4] with pmf given by

$$p_\alpha(k) = \frac{(\alpha-1)L(\mathcal{K},\alpha)}{\min(\mathcal{K})}\left(\frac{k_i}{\min(\mathcal{K})}\right)^{-\alpha}, \tag{11}$$

with the parameter α and corresponding CDF being $\mathcal{F}_\alpha(k)$, where $L(\mathcal{K},\alpha)$ is a normalizing constant depending on both the parameter α and degree sample data $\mathcal{K}$. In our case we ignore any isolated nodes in our structure; hence

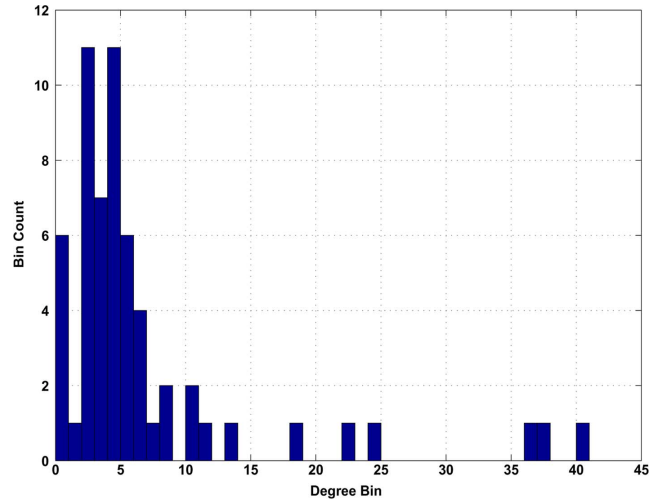


Fig. 13. Degree distribution of the Bayesian network learned from learning algorithm with L_1 regulation.

we have $\min(\mathcal{K}) \geq 1$. The MLE of parameter α could be numerically obtained by solving:

$$\hat{\alpha} = 1 + M \Big/ \left(-M\frac{\frac{d}{d\hat{\alpha}}L(\mathcal{K},\hat{\alpha})}{L(\mathcal{K},\hat{\alpha})} + \sum_{i=1}^{M}\log k_i\right). \tag{12}$$

2) Kolmogorov-Smirnov Test: For the logical structure learned from data, its degree distribution is shown in Fig. 13. We use the MLE to estimate parameters obtained from the Bayesian network and carry out the following two hypothesis testings a and b for its degree's CDF $\mathcal{F}_0(k)$:

- H_{a0}: $\mathcal{F}_0(k)$ follows the Poisson distribution $\mathcal{F}_\lambda(k|\hat{\lambda})$;
- H_{a1}: $\mathcal{F}_0(k)$ does not follow the Poisson distribution $\mathcal{F}_\lambda(k|\hat{\lambda})$;
- H_{b0}: $\mathcal{F}_0(k)$ follows the power law distribution $\mathcal{F}_\alpha(k|\hat{\alpha})$;
- H_{b1}: $\mathcal{F}_0(k)$ does not follow the power law distribution $\mathcal{F}_\alpha(k|\hat{\alpha})$.

Then, we apply the Kolmogorov-Smirnov (K-S) test (Appendix II). Given the empirical cumulative distribution function (CDF) obtained from the degree data $\mathcal{F}_0(k)$, and a target CDF $\mathcal{F}_1(k)$, with parameters estimated from the degree data, the null hypothesis is that the data has the target CDF and the alternative hypothesis is that the data does not have the target CDF. The results of the K-S test on hypothesis H_a and H_b are:

- For Small World model, there is no evidence to reject null hypothesis with p-value 0.0635 and MLE $\hat{\lambda} = 6.273$. Both CDFs are plotted in Fig. 14 to show the similarity.
- For Scale Free model, the null hypothesis is rejected.

This numerical result shows that the hypothesis H_{a0} is accepted; i.e., *the degree of the Bayesian network topology satisfies a Poisson distribution, or equivalently, the Bayesian structure can be approximated by a Small World network.*

Note that we need to verify that the conclusion of the K-S test is valid. The K-S test is proved to be effective for a large sample number M. Here we have $M = 207$. To decide whether this value is large enough for a significant K-S test result, we conduct another hypothesis testing to validate the K-S test with sample size $M = 207$. To that end, we have used the discrete Pearson system [28], [29], in which a CDF having the same first

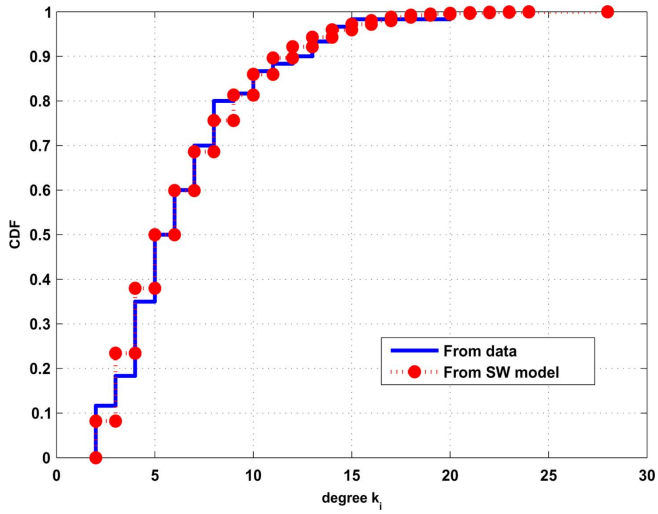


Fig. 14. Degree distribution of Small World model fits to the degree distribution of structure learnt from data.

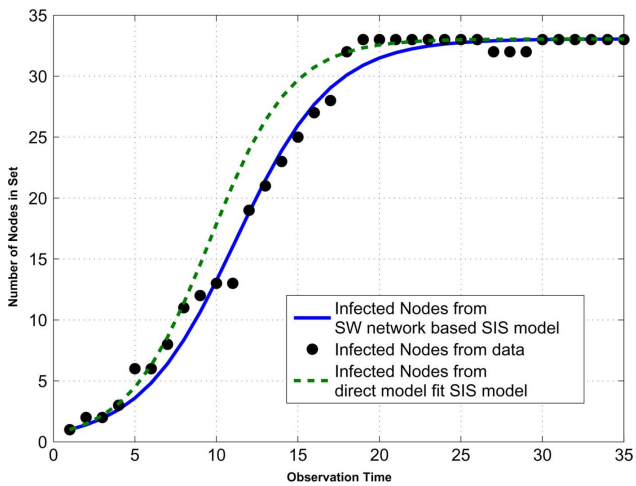


Fig. 15. Time evolution of i as a function of time: Direct model fit SIS model, Small World network based SIS model and real measurements.

four moments as the empirical distribution is generated and then compared with the original empirical distribution. The mathematical details could be found in Appendix III. Eventually, the Pearson system demonstrated that the sample number 207 is statistically valid.

3) Small World Model Fitting: Under the Small World model's Poisson degree distribution, we have $\lambda = (\gamma)/(\beta)$ [3], [31] (recall that γ and β are the parameters of the SIS model). Fitting the Small World model into the skeleton of the logical structure learned from the frequency data, we have $\hat{\lambda} = (1)/(M)\sum_{i=1}^{M} k_i = 6.273 \approx (\gamma)/(\beta)$. Note that in the direct model fitting method we have $(\beta = 0.042, \gamma = 0.252)$ and $(\gamma)/(\beta) = 5.97$, which is consistent. Thus, based on the Small World model parameter, we use $(\beta = 0.042, \gamma = \beta\hat{\lambda} = 0.264)$ (we use the same β obtained from the direct model fitting), obtain the corresponding SIS model, plot the number of infected nodes $I(t)$ in Fig. 15. We can see that Small World network based SIS model can approximate the time evolution of the frequency oscillation very well. It is similar to but better than the result obtained from the direct model fitting during the period when frequencies changes among nodes.

The reason for this observation might be that, for the direct model fitting, the majority of data points to fit the number of infected nodes are those corresponding to the stationary state of the observed event when all nodes are infected, as well as the stationary state before the spread of epidemic when all nodes are susceptible. Meanwhile, the Small World structure model fitting focuses more on the transient behavior or the state transition dynamics of the change in the number of infected nodes, which is the time interval of more interests. As shown in Fig. 15, the Small World network structure oriented model fitting fits better in that time interval. This demonstrates that the learned logic structure represents the structure of frequency dynamics very well.

VIII. Conclusion

In this paper we have studied the frequency dynamic in power networks, using the proposed iterative Bayesian structure learning algorithm with L-1 regulation to learn the logical structure. It has achieved a high frequency prediction accuracy and a good performance for detecting false data. By considering the frequency oscillation as epidemic propagation and the power network as a social network, we have numerically proved that the learned logic network can be well approximated by the Small World network structure model to describe the frequency dynamic. The logical network topology, instead of the physical one, has been obtained from the structure learning using the FNET frequency measurement data. This proposed approach has good performance in accuracy and sensitivity, and does not rely on topology information. Therefore our approach is practically useful in frequency dynamic analysis with good rate of prediction a few hundred milliseconds ahead of time when the grid is working normally or suffering frequency oscillation. However, our approach requires centralized data collection and the structure learning can not support real-time given limited computing resource.

Our future work includes:

- We will test our framework using more measurements and heterogeneous epidemic models as stated in previous footnote.
- Structure updating and distributed structure learning will be studied to make our approach support real-time computing to better fit practical use requirements.
- The stationary distribution of the epidemic model will be studied for more insights on frequency dynamics.

Appendix I
Introduction to K-2 Structure Learning Algorithm

A. K-2 Algorithm and Cooper-Herskovits Function

K2 algorithm [7] is widely used in structure learning. It greedily learns the Bayesian newtork G with edge matrix dag, by searching for the maximum possible number of parents $\|Pa_i^G\|$ for each node in $\{F_i\}_{i=1}^N$. The selection of the parent nodes should maximize the score function $g(F_i, Pa_i^G)$ within some searching space. Therefore such algorithms are also called search-and-score algorithms [18]. The score is the Cooper-Herskovits (CH) function [7]. Under the assumption

that searching on each node is independent, the resulting structure is I-equivalent [5] to the most probable Bayesian network given training data D:

$$Pr(D|G) = \prod_{i=1}^{N} g\left(F_i, Pa_i^G\right) = \prod_{i=1}^{N}\prod_{j=1}^{M_p}\left(\frac{(M_i-1)!}{(N_{ij}+M_i-1)!}\prod_{k=1}^{M_i}(N_{ijk})!\right), \tag{13}$$

where $F_j \in Pa_i$ has M_j distinct values; N_{ijk} is the frequency count of (d_{ij}, d_{jk}) with $\{d_{ik}\}_{k=1}^{M_i}$ being F_i's number of unique instantiation and $\{d_{ij}\}_{j=1}^{M_p}$ for each $F_j \in Pa_i$. $N_{ij} = \sum_{k=1}^{M_i} N_{ijk}$ is the number of all possible values F_i can take due to parents Pa_i. All statistics are counted in D.

B. Training Data and Cross Validation

Cross validation of the learning result helps to prevent over learning caused by the lack of sufficient samples. Each node F_i comes with frequency measurements of equal length $M = N_{\text{train}} + N_{\text{test}}$ as $\{F_i^j\}_{j=1}^{M}$. The training data should be sufficient long N_{train} and exclusive with testing data. Details of a K-fold cross validation can be found in [5]. We not only need to learn the structure due to the spatial correlation, but also the structure due to the correlation in the time domain, or the cause-effect relationship between two different sample slots. These two structures are learned separately using different sets of training and testing data. The Bayesian network in Fig. 2 is learned using training data $\{F_i^j\}_{j=1}^{N_{\text{train}}-1}$ for node F_i^t and $\{F_i^j\}_{j=2}^{N_{\text{train}}}$ for node F_i^{t+1}. The remaining testing data is used for validation by checking the inference accuracy in frequency prediction. The structure corresponding to time t in Fig. 3 is learned using the training data $\{F_i^j\}_{j=1}^{N_{\text{train}}}$ for each node to obtain the spatial causality between nodes. The structure corresponding to node F_1 at time $t+1$ in Fig. 3 is learned using the time shifted training data $\{F_i^j\}_{j=1}^{N_{\text{train}}-1} \cup \{F_1^j\}_{j=2}^{N_{\text{train}}}$ to see whose frequency at time t could influence F_1's frequency in the next time slot.

C. Ordering Input in K2 Algorithm

The K2 algorithm requires an ordered input with a set of potential parents $\hat{P}a_i = \{F_j\}_{j=1}^{i-1}$ for node F_i. Thus, $F_j \in Pa_i$ implies $j < i$ or F_j must proceed F_i. This *ordering* ensures the graph to be acyclic. However, it requires prior information or other techniques such as Conditional Independence (CI) test [19] to decide an order. Suboptimal ordering incurs inaccuracy in structure learning [21]. As an example in Fig. 16, which is the simplified case for the example given in Fig. 1 of [22], we suppose the best structure is Fig. 16(2), where A and B are conditionally independent given node C. When $\{A, C, B\}$ is the input order, this structure could be recovered. That is to say, when searching the parents for node B, both nodes A and C are considered. However, with input order $\{A, B, C\}$, when searching parents for node B, only node A is considered; then it is possible to result in the structure in Fig. 16(1), which is not I-equivalent to Fig. 16(2). This motivates us to propose the algorithm that handles random orders.

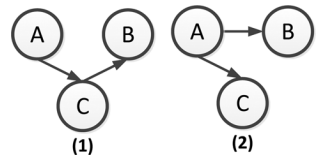


Fig. 16. Illustration of inaccuracy due to random sequences of input nodes.

Appendix II
Introduction to Kolmogorov-Smirnov Test

K-S statistic is widely used in K-S test as a measure of distance between empirical distribution functions (EDFs) and target CDF. EDF $\mathcal{F}_0(k)$ for M i.i.d. observations $\{k_i\}_{i=1}^{M}$ is:

$$\mathcal{F}_0(k) = \frac{1}{M}\sum_{i=1}^{M} I_{k_i \le K}. \tag{14}$$

Then, the K-S statistic for target CDF $\mathcal{F}_1(k)$ and EDF $\mathcal{F}_0(k)$ is

$$D_M = \sup_k |\mathcal{F}_0(k) - \mathcal{F}_1(k)|. \tag{15}$$

The reason we use K-S statistic as a measure of distance is due to the Glivenko-antelli theorem [29] which shows that D_M converges to 0 almost surely as $M \to \infty$ if the samples k_i comes from distribution $\mathcal{F}_1(k)$.

The Kolmogorov Theorem [29] shows that as $M \to \infty$, $\sqrt{M}D_M$ converges to the Kolmogorov distribution [29]. For p-value α, the null hypothesis that empirical distribution $\mathcal{F}_0(k)$ follows the target distribution $\mathcal{F}_1(k)$ is rejected if $\sqrt{M}D_M > K_\alpha$ with K_α being the critical value of the Kolmogorov distribution $Pr(k \le K_\alpha) = 1 - \alpha$, and vice versa. Thus, we provide both the hypothesis testing results and the corresponding p-value as K-S test results,.

Appendix III
Pearson System

The Pearson system [29] is a family of continuous distributions built in a systematical way. [28] extended the Pearson system to the discrete case. Concisely speaking, essentially the Pearson system uses a series of classical statistical methods (such as Gaussian distribution and beta distribution). It adjusts their locations and scales their distribution shapes to generate a family of distributions that fit the first four standardized moments of the given target distribution. The resulting PDF is described by a difference equation.

Given M instances sampled from the empirical distribution of node degrees with CDF $\mathcal{F}_0$, the Pearson system computes another M samples that are drawn from a discrete distribution with CDF $\mathcal{F}_2$, which has the same first four moments as those of $\mathcal{F}_0$. Using tools in Matlab, we can obtain that $\mathcal{F}_2$ satisfies the following equation:

$$\log\left(\frac{\mathcal{F}_2(k+1)}{\mathcal{F}_2(k)}\right) = \frac{-(k+1.4739)}{1.0419k - 0.1580}. \tag{16}$$

Therefore, the target CDF $\mathcal{F}_2$ and empirical CDF $\mathcal{F}_0$ should pass the K-S test. This provides us a way to validate the K-S test by using hypothesis testing c:

- H_{0c}: $\mathcal{F}_0$ follows the Pearson system CDF $\mathcal{F}_2$;
- H_{1c}: $\mathcal{F}_0$ does not follow the Pearson system CDF $\mathcal{F}_2$.

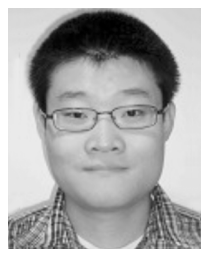
The K-S test on the above hypotheses shows that the null hypothesis is accepted with p-value 0.0107. This verifies that $N = 207$ is sufficiently large to be statistically valid.

ACKNOWLEDGMENT

The authors sincerely thank Prof. Yilu Liu for providing the frequency measurement data and the insightful discussion.

REFERENCES

[1] H. Li, G. W. RosenWald, J. Jung, and C. C. Liu, "The strategic power infrastructure defense," *Proc. IEEE*, vol. 93, pp. 918–933, May 2005.
[2] J. Machowski, J. W. Bialek, and J. R. Bumby, *Power System Dynamics: Stability and Control*, 2nd ed. Hoboken, NJ: Wiley, 2008.
[3] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamic Processes on Complex Networks*. Cambridge, U.K.: Cambridge Univ. Press, 2008.
[4] F. Vega-Redondo, *Complex Social Networks*. Cambridge, U.K.: Cambridge Univ. Press, 2007.
[5] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*. Cambridge, MA: MIT Press, 2009.
[6] Z. Zhong *et al.*, "Power system frequency monitoring network (FNET) implementation," *IEEE Trans. Power Syst.* vol. 20, p. 1914, 2005 [Online]. Available: http://powerit.utk.edu/fnet.html
[7] G. Cooper and E. Herskovits, "A Bayesian method for the induction of probabilistic networks from data," *Mach. Learn.*, vol. 9, pp. 309–347, 1992.
[8] H. Bevrani, *Robust Power System Frequency Control*. New York: Springer, 2008.
[9] A. Girgis, J. W. Stephens, and E. B. Makram, "Measurement and prediction of voltage flicker magnitude and frequency," *IEEE Trans. Power Del.*, vol. 10, no. 3, pp. 1600–1605, 1995.
[10] H. Karimi *et al.*, "Estimation of frequency and its rate of change for applications in power systems," *IEEE Trans. Power Del.*, vol. 19, no. 2, pp. 472–480, 2004.
[11] A. Adly, G. William, and L. Peterson, "Adaptive estimation of power system frequency deviation and its rate of change for calculation sudden power system overload," *IEEE Trans. Power Del.*, vol. 5, no. 2, pp. 585–594, 1990.
[12] Y. Zhu, L. Huo, and J. Lu, "Bayesian networks-based approach for power systems fault diagnosis," *IEEE Trans. Power Del.*, vol. 21, no. 2, pp. 634–639, 2006.
[13] D. C. Yu, T. C. Nguyen, and P. Haddawy, "Bayesian network model for reliability assessment of power systems," *IEEE Trans. Power Syst.*, vol. 14, no. 2, pp. 426–432, 1999.
[14] S. Arianos, E. Bompard, and A. Carbone, "Power grid vulnerability: A complex network approach," *Chaos*, vol. 19, no. 1, 2009.
[15] G. Chen, Z. Dong, and D. J. Hill, "Attack structural vulnerability of power grids," *Physica A: Stat. Mech. Appl.*, vol. 383, pp. 595–603, 2010.
[16] H. Xiao and E. M. Yeh, "Cascading link failure in the power grid: A percolation based analysis," in *Proc. IEEE Int. Workshop Smart Grid*, 2011.
[17] R. Chen and E. H. Herskovits, "Network analysis of mild cognitive impairment," *NeuroImage*, vol. 29, pp. 1252–1259, 2006.
[18] P. Spirtes, C. Glymour, and R. Scheines, *Causation, Prediction, and Search*, 2nd ed. Cambridge, MA: MIT Press, 2000.
[19] R. Chen, K. Sivakumar, and H. Kargupta, "Learning bayesian network structure from distributed data," in *Proc. 3rd SIAM Int. Data Mining Conf.*, 2003, pp. 284–288.
[20] H. Chan and A. Darwiche, "Sensitivity analysis in Bayesian networks," in *Proc. 20th Conf. Uncertainty AI*, 2004, pp. 67–75.
[21] X. W. Chen, G. Anantha, and X. T. Lin, "Improving Bayesian network structure learning with mutual information-based node ordering in the K2 algorithm," *IEEE Trans. Knowl. Data Eng.*, vol. 20, pp. 628–640, 2008.
[22] J. Cheng, A. David, and W. Liu, "An algorithm for Bayesian belief network construction from data," in *Proc. AI STAT*, 1997, pp. 83–90.
[23] P. Ravikumar and J. Lafferty, "High-dimensional graphical model selection using L1-regularized logistic regression," in *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press, 2007, vol. 19, pp. 1465–1472.
[24] M. Schmidt, R. Niculescu-Mizil, and K. Murphy, "Learning graphical model structure using L1-regularization paths," in *Proc. 22th AAAI Conf. Artif. Intell.*, 2007, pp. 1278–1283.
[25] M. Park and T. Hastie, "L1 regularization path algorithm for generalized linear models," Dept. Statistics, Stanford Univ., Tech. Rep., 2006, .
[26] M. Barthelemy; and L. Amaral, "Small-world networks: Evidence for a crossover picture," *Phys. Rev. Lett.*, vol. 82, no. 15, p. 3180, 1999.
[27] D. J. de Solla Price, "Networks of scientific papers," *Science*, vol. 149, pp. 510–515, 1965.
[28] J. K. Ord, "On a system of discrete distributions," *Biometrika*, vol. 54, no. 3/4, pp. 649–656, 1967.
[29] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*. New York: Wiley-Interscience, 1994.
[30] V. Poor and O. Hadjiliadis, *Quickest Detection*. Cambridge, U.K.: Cambridge Univ. Press, 2008.
[31] S. Peng *et al.*, "Steady states and critical behavior of epidemic spreading on complex networks," in *Proc. 7th World Congr. Intell. Control Autom.*, 2008, pp. 3481–3486.
[32] M. Small and C. K. Tse, "Modeling the SARS outbreak in Hong Kong with small world or scale free networks," in *Proc. Int. Symp. Nonlinear Theory Appl.*, 2004, pp. 147–152.



Hannan Ma (S'05-M'12) received the B.S. degree in electrical engineering from Tsinghua University, Beijing, China, in 2008, and the M.S. degree in electrical engineering from Washington State University, Pullman, WA, in 2010. He is currently working toward the Ph.D. degree in electrical engineering at the University of Tennessee, Knoxville. His Ph.D. advisor is Dr. Husheng Li.

He is the recipient of the Best Bachelor thesis in electrical engineering from Tsinghua University in 2008. In 2012, he did an internship with Broadcom for nine months.

Husheng Li (S'00–M'05) received the B.S. and M.S. degrees in electronic engineering from Tsinghua University, Beijing, China, in 1998 and 2000, respectively, and the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ, in 2005.

From 2005 to 2007, he worked as a senior engineer at Qualcomm Inc., San Diego, CA. In 2007, he joined the EECS department of the University of Tennessee, Knoxville, as an Assistant Professor. He is also an International Scholar of Kyung Hee University, Korea. His research is mainly focused on wireless communications and smart grid.

Dr. Li is the recipient of the Best Paper Award of the *EURASIP Journal of Wireless Communications and Networks*, 2005 (together with his Ph.D. advisor: Prof. H. V. Poor), the best demo award of Globecom 2010 and the Best Paper Award of ICC 2011.